Enhancing Physician Quality Performance Measurement and Reporting Through Data Aggregation

The Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project

FINAL REPORT
October 31, 2008

Prepared for: Centers for Medicare & Medicaid Services

Contract No. HHSM-500-2006-MD002C

Prepared by:
Delmarva Foundation for Medical Care
Easton, Maryland

Delmarva Foundation
Improving Health in the Communities We Serve

Enhancing Physician Quality Performance Measurement and Reporting Through Data Aggregation

The Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project

FINAL REPORT
October 31, 2008

Prepared for: Centers for Medicare & Medicaid Services

Contract No. HHSM-500-2006-MD002C

Prepared by:
Delmarva Foundation for Medical Care, Inc.
Easton, Maryland

The analyses upon which this publication is based were performed under Contract Number HHSM-500-2006-MD002C , entitled "Quality Improvement Organization (QIO) for the State of Maryland," sponsored by the Centers for Medicare & Medicaid Services, Department of Health & Human Services. The content of this publication does not necessarily reflect the views or policies of the Department of Health & Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

The author assumes full responsibility for the accuracy and completeness of the ideas presented. This article is a direct result of the Health Care Quality Improvement Program initiated by the Centers for Medicare & Medicaid Services, which has encouraged identification of quality improvement projects derived from analysis of patterns of care, and therefore required no special funding on the part of this contractor. Feedback to the author concerning the issues presented is welcomed.

Contents

Executive Summary 1
Introduction and Overview of the BQI Project 1
Measure Selection 1
Combining Medicare FFS and Other Data 1
Analyses Related to Methodological Questions 2
1. Attributing Beneficiaries to Physicians 2
2. Attributing Physicians to Physician Groups 2
3. Reliability of Physician Performance Rankings 3
4. Correlations of Performance Rates Across Payers 3
Reporting Performance Measure to Physicians and Consumers: Lessons Learned 3
Issues Related to Individual Physician-Level Calculation and Reporting 3
Project Evaluation and Conclusions 4
Goal 1: Study the issues and benefits of aggregating Medicare FFS data with other regionally-based quality data, including commercial payer administrative data, provider-submitted data, etc. to calculate quality measures for ambulatory care 4
Goal 2: Study the benefits of reporting quality measures to physicians and other providers of care. 4
Goal 3: Study the benefits of reporting quality measures to Medicare beneficiaries. 5

1. Moving Toward Transparency in Health Care 1-1
The Evolution of the BQI Project 1-2
Achieving National and Local Collaboration 1-3
BQI's Impact on the Performance Measurement and Public Reporting Environment 1-5

2. Selecting Appropriate Measures of Physician Performance for Medicare Beneficiaries 2-1
Applicability to Medicare Beneficiaries 2-1
Available Measure Specifications 2-3
Importance of Precise and Consistent Measure Defintions 2-4
Consensus Process Used to Arrive at Selected Group of Measures 2-6

3. A Comprehensive Picture of Physician Performance: Combining Medicare Fee-for-Service (FFS) and Other Data 3-1
1. Is it technically and statistically possible to combine the Medicare FFS data with other datasets? 3-2
2. If technical issues are overcome, is the added information from Medicare administrative data useful in reporting to beneficiaries and providers? 3-3
3. Does including Medicare FFS data affect the measures? 3-5
4. Does reporting performance measure results in an aggregated fashion produce more meaningful information? 3-6
5. Does the addition of Medicare FFS data to other payer data increase meaningful results for individual providers who have small patient numbers? 3-7

4. Analyses Addressing Methodological Questions 4-1
Analysis of Attribution of Beneficiaries to Physicians 4-2
1. Are physicians' process measure rates substantively affected by the selection of the attribution rule? 4-4
2. Are the average reliabilities of the process measure rates substantively affected by the choice of attribution rules? 4-5
3. Are percentages of included physicians, or the percentages of beneficiaries attributed to physicians, substantively affected by the choice of attribution rules? 4-5
Summary 4-6
Attributing Physicians to Physician Groups 4-7
Defining Physician Groups to Report Quality Measures 4-7
Physician Grouping Methods 4-7
Options of Physician Grouping Methods 4-8
Results 4-9
Conclusions Regarding Assigning Physicians to Groups Using Administrative Data 4-10
Reliability of Physician Performance Results 4-11
Reliability Results 4-12
Discussion 4-13
Correlations of Performance Rates Across Payers 4-14
An Example of the Effects of Managed Care Systems on Provider Rankings 4-15
Summary 4-16

5. Reporting of Results to Physicians and Consumers 5-1
Customers for Reporting, Uses for Information 5-1
Experience with Use of Quality Data Reports by Providers 5-3
Experience with Use of Quality Data Reports by Consumers 5-4
Summary of the Value of Public Reporting 5-5
Unintended Adverse Consequences of Public Reporting 5-6
Summary of Reporting Preferences 5-6
Data Feedback to Providers—Data Sharing Issues 5-7

6. Individual Physician-Level Measurement and Reporting 6-1
Experiences with Producing Individual Physician-Level Results 6-1
Barriers to Individual-Level Reporting and Mitigation Strategies 6-3
An Example of Successful Individual Physician-Level Reporting Through Patient Experience of Care Measures 6-4

7. Project Evaluation and Conclusions 7-1
Pilot Sites' Assessment of the Effect of the BQI Project on Their Communities 7-1
California Cooperative Health Care Reporting Initiative (CCHRI) 7-1
Massachusetts Health Quality Partners (MHQP) 7-2
Minnesota Community Measurement (MNCM) 7-3
Wisconsin Collaborative for Healthcare Quality (WCHQ) 7-3
Indiana Health Information Exchange (IHIE) 7-4
Center for Health Information and Research (CHIR) 7-4
Conclusions from the BQI Project 7-5
Effective Methods of Aggregating Medicare Data and Other Regional Data 7-5
Providing Information to Medicare Providers for Quality Improvement 7-6
Provide Medicare Beneficiaries with Physician Performance Information 7-6
Conclusions Regarding Performance Measure Selection and Calculation 7-7
Implications of the BQI Project 7-7

Acknowledgements 1

Appendix 1
AQA Questions and Responses

Appendix 2
Pilot Site Descriptions
- CCHRI
- CHIR
- IHIE
- MHQP
- MNCM
- WCHQ

Appendix 3
All Measures Considered

Appendix 4
Pilot Medicare Data Processes
Pilot Attribution Methods

Appendix 5
The Effects of Differing Attribution Methods—Beneficiaries to Physicians

Appendix 6
Defining Physician Groups to Report Quality Measures

Appendix 7
Reliability of Physician Performance Measures

Appendix 8
Final Cycle 1 Data Aggregation Report

Appendix 9
Final Cycle 2 Data Aggregation Report
Report Appendix
- Appendix A: BQI Project Measures
- Appendix B: BQI Cycle 2 Measures, by Pilot
- Appendix C: Pilot Physician-Physician Group Attribution Methods
- Appendix D: Physician Group Attribution Results
- Appendix E: List of Evaluation and Management Codes used to identify physicians eligible to be attributed to beneficiaries
- Appendix F: Definition of PCP by Provider Specialty and Measure

Appendix G: Attribution Analysis Results
Appendix H: Attribution & Reliability Analysis Results
Appendix I: Reliability Scores for Selected Denominators

Appendix 10

Final Cycle 3 Report
Report Appendix
Appendix A: BQI Project Measures
Appendix B: BQI Cycle 3 Measures, by Pilot

Appendix 11

Data Feedback to Providers: Experience and Lessons Learned

Appendix 12

Glossary of Terms

Appendix 13

References and Resources

Executive Summary

Introduction and Overview of the BQI Project

The Centers for Medicare & Medicaid Services (CMS) launched the Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project in 2006. As part of the Department of Health and Human Services' (HHS) Value-Driven Health Care Initiative, the three overarching goals of the Project were to study:

1. The issues and benefits of aggregating Medicare fee-for-service (FFS) data with other regionally-based quality data to calculate quality measures for ambulatory care.
2. The benefits of reporting quality measures to physicians and providers.
3. The benefits of reporting quality measures to Medicare beneficiaries.

For this two-year project, CMS contracted with the Delmarva Foundation for Medical Care (Delmarva), the Quality Improvement Organization (QIO) for the State of Maryland. Delmarva in turn contracted with the following six regional collaboratives (pilot sites) to accomplish these goals:

1. California Cooperative Healthcare Reporting Initiative (CCHRI)
2. Center for Health Information and Research-Arizona State University (CHIR)
3. Indiana Health Information Exchange (IHIE)
4. Massachusetts Health Quality Partners (MHQP)
5. Minnesota Community Measurement (MNCM)
6. Wisconsin Collaborative for Healthcare Quality (WCHQ)

Delmarva Foundation
Improving Health in the Communities We Serve

Under Delmarva's coordination, the pilot sites tested effective ways to select measures and aggregate data for public reporting

Under Delmarva's coordination, the pilot sites tested effective ways to select measures and aggregate data for public reporting. Key outcomes in these areas include:

Measure Selection

- The Project found that calculating measures from administrative claims data only limited the number and types of quality measures that were able to be calculated.
- Small differences in measure specifications for numerators and denominators can cause significant differences in results, thereby complicating comparisons across different regions.
- There is a need for a national consensus on measure specifications defined for each different type of data source (commercial managed care, Medicare FFS, clinical records).

Combining Medicare FFS and Other Data

- Including information about Medicare beneficiaries is essential for establishing a complete picture of the quality of community medical care.

- Technical issues related to aggregation are complex but can be overcome.
- Lack of uniform standards means combining data from differing sources and requires constant, detailed attention, time, and resources.
- Data more than 18 months old becomes increasingly irrelevant to customers.
- Administrative data alone cannot accurately identify a patient's physician or physician membership in medical groups. Feedback from providers is also necessary.
- For most measures, Medicare FFS results were different from commercial and Medicare managed care results. Combining data changed results and also rankings in some cases.
- Uncertainties introduced through statistical methodologies may offset the advantages of having larger numbers of patients from aggregating data.
- Merging data sets is most useful when the measure results are similar and the performance rankings are correlated across the sets. In other cases results may be more meaningfully reported separately by payer.

Analyses Related to Methodological Questions

To clarify the impact differing methodological choices have on performance measure calculations, Delmarva studied four methodological questions using Medicare FFS data only for the six states from Calendar Year 2006

To clarify the impact differing methodological choices have on performance measure calculations, Delmarva studied four methodological questions using Medicare FFS data only for the six states from Calendar Year (CY) 2006.

1. Attributing Beneficiaries to Physicians

After comparing five methods for assigning patients to responsible physicians using only administrative claims data, findings include:

- The primary difference among them was whether the method assigned the patient's outcome to one physician or to multiple treating physicians.
- Calculated measure results were not strongly affected by the choice of method.
- Methods that assigned patients to only one doctor resulted in a smaller number of doctors having a large enough patient population to be meaningfully reported.

2. Attributing Physicians to Physician Groups

Delmarva also studied and compared three methods of assigning physicians to physician groups by using administrative claims data. Findings include:

- Although UPIN numbers for physician groups can give clinic-site level assignments, the group UPIN numbers are not available in all administrative data.
- Tax ID numbers (TINs) are an attractive option because they are consistently available. However, sometimes TINs identify only very large, corporate groups.
- All methods that create physician group rosters from claims data have an inherent error of omitting actual members and of including incorrect members. In the example validated by WCHQ, the error rate ranged from 6 to 14 percent.
- Techniques that allow doctors to edit and correct rosters are recommended.

3. Reliability of Physician Performance Rankings

Delmarva performed a multi-variable analysis of the reliability of distinguishing between physicians' performance by using their calculated ranking in measure results. Findings include:

- Patient sample sizes for each doctor of at least 30-50 patients were necessary in most cases for making reliable distinctions. The number of patients necessary depended on the measure, on the difference between the rankings, on the actual result (low-medium-high performance), and on the variability of a doctor's measurement results around their average.
- For some measure results, patient sample sizes of much more than 50 would be required to make a reliable distinction based on rankings. Because multiple independent factors affect reliability, calculations of a result's reliability are best done for every specific case.
- Excluding single providers with eligible patient populations less than 50 for certain measures may result in many if not most doctors not having any reportable result.

4. Correlations of Performance Rates Across Payers

Delmarva examined issues related to conducting and interpreting correlation of performance rates across payers. Findings include:

- More physicians will have more reportable results if multiple populations' data are combined to give a larger patient population. In combining these data, a provider's performance results for one payer population needs to be well correlated with the same provider's results for another payer's patient population. Additionally, the magnitude of the results needs to be similar as well.
- Rankings of provider performance across payers were poorly correlated for many examples. Analysis showed that changes in rankings can be caused by differing provider behavior, differences in patient demographics, and by small changes in system support effects (such as managed care and coverage options).
- When low correlation of performance mesures across differing payers is found, an analysis that considers local factors is necessary to investigate the causes.

Reporting Performance Measure to Physicians and Consumers: Lessons Learned

- Varying stakeholders have differing needs and preferences for reported measures.
- For most customers and most purposes, annual reports are an adequate frequency.
- Providers of care prefer the most detailed measures. Current data is sometimes needed.
- Patients will accept clinic-site level data if it is more reliable and accurate than individual doctor-level data. Patients prefer broader measures, placed in a context relevant to them.
- In addition to HIPPA rules, QIO-specific regulations limited the ability of BQI pilot sites to share results from Medicare FFS data with physicians.

Issues Related to Individual Physician-Level Calculation and Reporting

- Most of the pilot sites preferred publicly reporting data at the level of physician groups.
- Individual provider-level data is desired by patients and payers for accountability, but

only if the results are understandable, valid and reliable; otherwise trust is lost.
- As medical care is increasingly team-based, group-level reporting may be most appropriate for many measures.
- Other reasons for group-level reporting include: lack of a clear, shared understanding among individual providers of who is responsible for specific patients; data submission issues and small sample sizes; and comfort and trust of providers.
- Barriers may be appropriately overcome in the future by identified strategies.

Project Evaluation and Conclusions

Along with pilot site assessments of how the BQI Project affected their communities, the report's concluding section summarizes the main findings related to the Project's goals.

Goal 1: Study the issues and benefits of aggregating Medicare FFS data with other regionally-based quality data, including commercial payer administrative data, provider-submitted data, etc. to calculate quality measures for ambulatory care.

- All six BQI pilot sites found that Medicare data was essential to creating a complete picture of the quality of medical care within their communities.
- Technical issues of combining Medicare with other data sets were complex but not insurmountable. Uniform data standards would simplify the task.
- Data sets are best aggregated at least at the level of individual providers, to minimize issues with medical group definition. Measure results can then be combined using consistent rosters for medical group-level result reporting.
- Data at the individual patient level is most useful for validation of patient responsibility by providers and for process improvement.

Goal 2: Study the benefits of reporting quality measures to physicians and other providers of care.

- The pilot sites demonstrated that engaged providers do improve their care processes in response to performance measure results. Sites that have been reporting for several years have seen steady increase in measure results.
- The consensus of providers and payers was that the speed of process improvement is accelerated by both Pay-For-Performance (P4P) and public reporting.
- Providers and patients want the most comprehensive data possible.
- When Medicare FFS outcomes are dissimilar to outcomes from other payers, reports separated by payer are useful to determine the reasons.
- Providers prefer the most detailed and current measures possible.
- Importantly, providers at several of the pilot sites use individual patient-level results from administrative data for identifying patients who need medical care and protection from harm.

Goal 3: Study the benefits of reporting quality measures to Medicare beneficiaries.

- Patients prefer broader quality measures placed into a context that shows how the measures are relevant to their experience. Beneficiaries are also interested in other aspects of medical quality, including access and patient experience of care.

The BQI Project has been the only initiative to combine multiple differing "raw" datasets to derive performance measures for ambulatory care. Issues surrounding data combination, measure selection, and performance measure calculation have been analyzed. Experience has been reported regarding preferred methods of reporting quality measure results to providers, beneficiaries, and other stakeholders.

The BQI Project has been the only initiative to combine multiple differing "raw" datasets to derive performance measures for ambulatory care

Additionally, the BQI pilot sites have demonstrated the efficacy of using public-private regional coalitions to improve ambulatory care. The lessons and insights will be useful to other regional collaboratives committed to community improvement, including Chartered Value Exchanges (CVEs) and the Robert Wood Johnson Foundation's Aligning Forces for Quality.

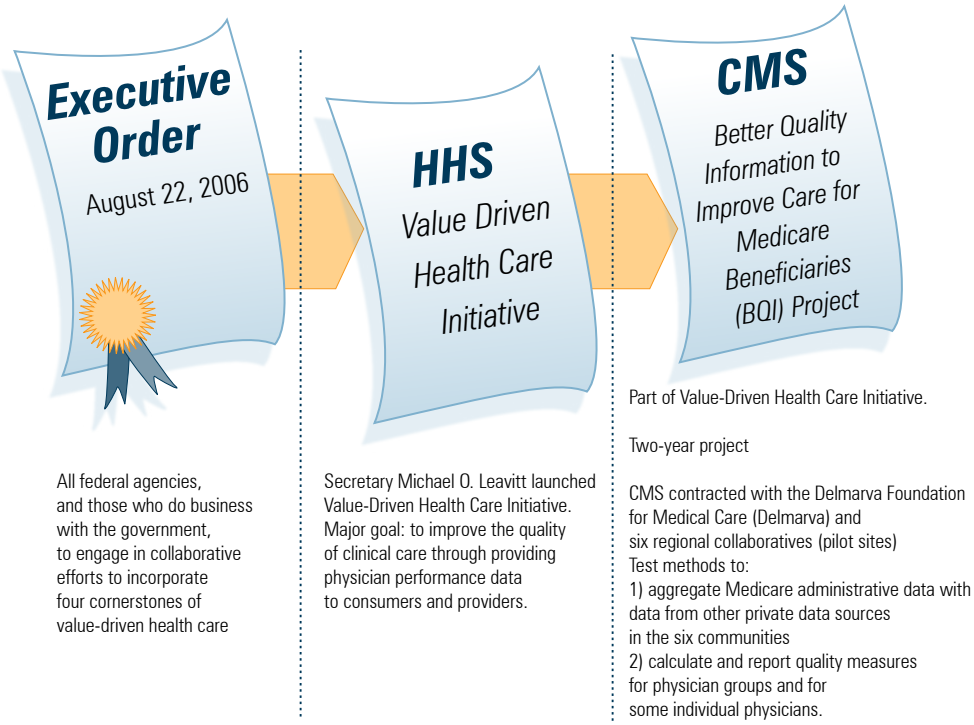
1

Moving Toward Transparency in Health Care

On August 22, 2006, President George W. Bush issued an Executive Order calling on all federal agencies, and those who do business with the government, to engage in collaborative efforts to incorporate four cornerstones of value-driven health care[1]:

1. Connecting the system through interoperable health information technologies.
2. Measuring and publishing quality data based on agreed-upon standards.
3. Measuring and publishing price information for specific services to patients.
4. Creating positive incentives that reward high-quality, cost-effective care and encourage consumers to actively choose the care that meets their needs.

Figure 1-1. Better Quality Information to Improve Care for Medicare Beneficiaries (BQI) Project



In response to the President's directive, Department of Health and Human Services (HHS) Secretary Michael O. Leavitt launched the Value-Driven Health Care Initiative[2]. A major goal of the initiative is to improve the quality of clinical care through providing physician performance data to consumers and providers.

A key component of value-driven health care is local and regional collaboration among providers, employers, health plans, and consumers. Because health care is provided to patients locally, solutions for delivering the best health care must be created locally as well. Delivery of care also differs geographically by demographic and market characteristics. To ensure that patients receive the same quality medical care regardless of where they live, nationally agreed-upon standards of care must be implemented. The regional collaboratives participating in this initiative agree with the importance of regional and national coordination to improve quality of care.

1 Value-Driven Healthcare: Four Cornerstones. Available at: http://www.hhs.gov/valuedriven/fourcornerstones/index.html.

2 Value-Driven Healthcare: Value Driven Healthcare Home. Available at: http://www.hhs.gov/valuedriven/index.html.

The Centers for Medicare & Medicaid Services (CMS) launched the Better Quality Information to Improve Care for Medicare Beneficiaries (BQI) Project in 2006 as part of the Value-Driven Health Care Initiative. For this two-year project, CMS contracted with the Delmarva Foundation for Medical Care (Delmarva) and six regional collaboratives (pilot sites) to test methods to: 1) aggregate Medicare administrative data with data from other private data sources in the six communities; and 2) calculate and report quality measures for physician groups and for some individual physicians.

In this Final Report, Delmarva presents our analysis of the results, issues, and lessons learned during the project. We have developed a high-level narrative that features important insights captured throughout the project. For further reference, we also include as appendices more technical, in-depth analyses of particular aspects discussed in the report.

The Evolution of the BQI Project

In 2005, the multi-stakeholder Ambulatory Quality Alliance[3], now known as the AQA, began to explore ways to measure and report physician-level clinical quality. The AQA determined that physicians would first need a comprehensive picture of their patient population in order to create relevant reports for providers and ultimately consumers.

Physicians typically receive patient population information in a piecemeal manner.[4] The reasons for this variability are twofold. First, physicians care for patients with a variety of public and private health insurers. Second, quality reports from single insurers can result in conflicting messages because they often use various measures applied to different patient populations. Consequently, neither providers nor consumers can easily obtain a complete picture of physician or medical group performance.

The AQA devised a pilot project to study effective ways of providing the public and providers with consistent measures of quality care. A significant focus of the project would address the feasibility of aggregating public and private data for the purposes of performance measurement and public reporting.

Collaborating with the Agency for Healthcare Research and Quality[5] (AHRQ), AQA developed study questions to test the feasibility of using multi-payer data in measurement and reporting. The topic areas for study that align with the BQI Project's framework are provided below. In Appendix 1, we include BQI responses to the complete list of AQA questions.

Measurement: Selecting suitable performance measures must take into consideration the measures' clinical importance and scientific validity. The process should also consider the feasibility of using available data sources and the measures' relevance to physician performance, purchasers and consumers. The BQI Project studied the effect that combining multiple datasets had upon the selection and calculation of measures.

3 AQA.Avaiable at: http://www.aqaalliance.org/.

4 US Department of Health & Human Services, Centers for Medicare & Medicaid Services. *2007 CMS Statistics*. June 2007. Avaiable at: http://www.cms.hhs.gov/CapMarketUpdates/Downloads/2007CMSstat.pdf.

5 Agency for Healthcare Research and Quality. Available at: http://www.ahrq.gov/.

Data Collection, Aggregation, and Validation: Determining an effective method for collecting and combining valid, reliable data from different sources (such as administrative data and medical records data) is key to publicly reporting performance data.

Reporting: Determining effective ways to report performance measurement information is integral to promoting quality improvement among providers and informing other audiences (e.g., consumers, employers, and other stakeholders).

The original AQA project design called for recruiting successful regional collaboratives from around the country to address some of the key study questions. These collaboratives, some of which were formed in the 1990s, have practical experience with engaging multi-stakeholder coalitions in measuring and reporting on performance measurement within their respective communities.

Achieving National and Local Collaboration

CMS contracted with Delmarva, the Quality Improvement Organization (QIO) for Maryland, to manage the BQI Project, conduct analyses on the Medicare administrative data, and to synthesize and report on the project's results. Delmarva was also responsible for providing the six BQI pilot sites with Medicare fee-for-service (FFS) claims data for their region to aggregate with local payer information. The following criteria were used to select the six BQI pilot sites:

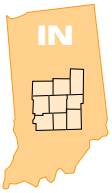
- Strong physician leadership engaged in creating the coalition
- Multiple employer participation
- Experience in measuring and aggregating physician-level data
- Presence of a public website for consumers to access information
- Demonstrated capacity and interest to accept additional tasks
- Willingness to work with a viable health information network if available
- Capability and infrastructure to begin data collection with a short time frame

The pilot sites selected for the project were:

- California Cooperative Healthcare Reporting Initiative (CCHRI)
- Center for Health Information and Research-Arizona State University (CHIR)
- Indiana Health Information Exchange (IHIE)
- Massachusetts Health Quality Partners (MHQP)
- Minnesota Community Measurement (MNCM)
- Wisconsin Collaborative for Healthcare Quality (WCHQ)

This diverse group of existing regional coalitions consists of providers, employers, and health plans working together to improve the quality of health care provided in their geographical areas. They represent long-standing collaboratives as well as newly organized ones (Figure 1-2). Each brought a unique experience, varying resources, and differing approaches to aggregating, calculating, and reporting performance measures. A more complete description of each pilot site may be found in Appendix 2.

Figure 1-2. BQI Pilot Sites

Pilot	CCHRI	CHIR	MHQP	MNCM	WCHQ	IHIE
Founded	1993	1995	1995	2002	2002	2004
State						
Composition	Health plans, employers, providers	Physicians, employers, health plans, Arizona State University, health systems	Health systems, health plans, providers, patients, government agencies, medical society	Health plans, medical groups, patients, employers	Medical groups, health systems, employers	Health systems, health plans, providers,employers, researchers, public health groups, economic development groups

The following examples demonstrate the diversity of the six pilot sites. WCHQ receives all-patient data (including Medicare, Medicaid, commercially insured and un-insured) directly from its member physician groups, while other pilot sites use administrative data from health plans. Some pilots (IHIE, MNCM, WCHQ) have access to clinical data from medical records; some have access to lab or pharmacy results (CHIR, IHIE, MHQP). Some pilots have a wealth of experience in publicly reporting health care performance information (CCHRI, MHQP, MNCM, WCHQ), while others (CHIR, IHIE) had not yet publicly reported prior to joining BQI.

CMS worked through Delmarva over the next two years to provide each site with Medicare FFS claims data for their region. During four aggregation cycles (see Table 1-1), pilot sites reported on selected quality measures for aggregated data based on a calendar year. Where possible the sites used various techniques to aggregate these data with local data provided by coalition partners (such as commercial payer administrative data, provider-based quality data, etc.). The sites studied the effects of combining the datasets, calculated measures of ambulatory care quality in their regions, and provided performance information through various methods of reporting.

Table 1-1. BQI Data Aggregation Cycles

	CHIR	CCHRI	IHIE	MHQP	MNCM[1]	WCHQ
Cycle 1	First 5 measures using calendar year (CY) 2005 data					
Cycle 2	Next 7 measures using CY 2006 data First 5 measures using CY 2006 data					
Cycle 3	12 measures July 2006–June 2007	N/A	12 measures July 2006–June 2007	N/A	6 measures, CY2007 and 7/1/07-6/30/08	Subset of 12 measures July 2006–June 2007
Cycle 4	All 12 measures using CY2007 data					

[1] *MNCM performed semi-annual reporting through its direct data submission.*

BQI's Impact on the Performance Measurement and Public Reporting Environment

The experience and lessons learned from the BQI Project are relevant to many current and future physician-level measurement and reporting initiatives. For example, new regional collaboratives (or Chartered Value Exchanges[6] [CVEs]) that are emerging as part of the Value-Driven Health Care Initiative can use BQI insights to develop methods for aggregating data from multiple sources.

Similarly, the Physician Quality Reporting Initiative[7] (PQRI) can also use information from BQI to develop effective methods for reporting on quality measures data for services covered by Medicare. Initiatives helping communities build new health care coalitions and systems, such as Bridges to Excellence[8] (BTE) and Aligning Forces for Quality[9] (AF4Q), will find lessons relevant to adapting national measures to regional issues of interest.

We are confident that these and other lessons derived from the BQI Project have much to contribute to the ongoing effort of transforming American health care into a truly value-driven system.

6 Value-Driven Healthcare: Chartering Value Exchanges. Available at: http://www.hhs.gov/valuedriven/communities/exchanges.html.

7 Physician Quality Reporting Initiative: Overview. Available at: http://www.cms.hhs.gov/PQRI/01_Overview.asp.

8 Bridges to Excellence. Available at: http://www.bridgestoexcellence.org/.

9 Aligning Forces for Quality: The Regional Market Place. Available at: http://www.forces4quality.org/.

2

Selecting Appropriate Measures of Physician Performance for Medicare Beneficiaries

The first task of the BQI Project was to determine a set of measures for calculation and reporting. The first set of measures we considered was the AQA Performance Measurement Workgroup's proposed set of 26 ambulatory care measures. Originating in 2005, these measures were selected for their ability to meet the following five criteria:

1. Clinical importance and scientific validity
2. Feasibility
3. Relevance to physician performance
4. Consumer relevance
5. Purchaser relevance

Recognizing that this set was merely a starting point, the AQA has since expanded its approved list[1] of measures to include Cardiology, Emergency Medicine, and other specialty measures. The AQA has also approved measures that follow certain parameters,[2] which, while broad, represent the most appropriate and applicable list available.

Measure Selection: Lessons Learned

- Use of only administrative claims data limits the number and types of measures that may be calculated.
- Because the specifications for selected national measures were designed for managed care datasets, decisions had to be made by pilot sites to apply them to data from other sources, such as Medicare.
- Small changes in specifications can cause significant changes in measure results.
- Measures must be precisely and consistently defined to be used across payers and regions.

The BQI Project followed these same criteria and parameters when selecting the measures used for the project. In this section we summarize the sources of selected measures and the measures' applicability to Medicare beneficiaries. We also discuss the importance of precisely and consistently defining measures and the consensus process followed to select measures.

Applicability to Medicare Beneficiaries

We initially envisioned the AQA's starter set of 26 measures providing enough applicable specifications for each pilot to aggregate data and calculate 12 measures. However, we soon discovered that although these measures can be used to measure ambulatory care performance in general, not all were applicable to the Medicare FFS beneficiary population.

Following the same five criteria to assess each measure in the starter set, Delmarva and the six pilot sites evaluated the list of measures. We discovered two factors in particular that weighed heavily on the

1 AQA. Available at: http://www.aqaalliance.org/performancewg.htm.

2 AQA Parameters for Selecting Measures for Physician Performance. Available at: http://www.aqaalliance.org/files/AQAParametersforSelectingAmbulatoryCare.doc.

We discovered two factors in particular weighed heavily on the BQI project's ability to use the measures: feasibility and consumer relevance

BQI project's ability to use the measures: feasibility and consumer relevance. Regional differences such as prior work, how the pilot sites were organized, and even the size of the communities, also influenced measure selection, as each pilot site had varying capabilities and resources available to select measures. Because the Medicare population was the project's target group, we were limited to measures that could be calculated using data elements available in the Medicare FFS claims data. This limitation meant that we were not able to select any measures that used a hybrid method of data collection, survey measures, and measures that required lab results. In addition, for Cycle 1 data aggregation, which was based on a calendar year 2005 measurement period, we could not select measures that required pharmacy claims, since the Medicare Part D benefit did not take effect until 2006. Lastly, because selected measures needed to be relevant to the Medicare population, measures with an upper age limit less than 65 years of age were eliminated.

Table 2-1 presents our analysis of the AQA starter set of measures and their use in the BQI project.

Table 2-1. AQA Starter Set Measures Evaluation

Measures Included in Cycle 1 (Measurement Year 2005)	
Measure	**Considerations**
Breast Cancer Screening	This measure was selected by all six pilots. Differences in the denominator age range were discussed and accepted.
Lipid Measurement (LDL testing) – Diabetes	This measure was selected by all six pilots. Differences in the method for identifying diabetics were discussed and accepted.
HbA1C Testing – Diabetes	This measure was selected by all six pilots. Differences in the method for identifying diabetics were discussed and accepted.
Colorectal Cancer Screening	This measure was selected by five out of six pilots. The sixth pilot added the measure in Cycle 2. Limitations due to lack of 10 years of historical Medicare and other payer data were accepted.
Eye Exam – Diabetes	This measure was selected by three of the six pilots (CHIR, MHQP, MNCM). One pilot added the measure in Cycle 2 (IHIE). Differences in the method for identifying diabetics were discussed and accepted. Potential lack of available vision data was accepted as a limitation.
HbA1C Management Control – Diabetes	This measure was selected by pilots with a provider direct data submission process (WCHQ, Cycle 1) or an established chart review process (MNCM, Cycle 2). Two pilots tested the feasibility of the measure using available laboratory data (IHIE, MHQP) in later Cycles.
Measures Included in Cycle 2 (Measurement Year 2006)	
Measure	**Considerations**
LDL Cholesterol Level (<130mg/dL) – Diabetes	This measure was selected by pilots with a provider direct data submission process (WCHQ) or an established chart review process (MNCM). Two pilots tested the feasibility of the measure using available laboratory data (IHIE, MHQP).
Antidepressant Medication Management – Acute Phase	One pilot (CHIR) selected this measure.
Beta-Blocker Treatment after Heart Attack	Three pilots (CCHRI, IHIE, MHQP) selected this measure.
Persistence of Beta-Blocker Therapy – Post MI	Four pilots (CHIR, CCHRI, IHIE, MHQP) selected this measure.
ACE Inhibitor /ARB Therapy	One pilot (CHIR) selected this measure.
Drug Therapy for Lowering LDL Cholesterol	Three pilots (CHIR, CCHRI, MHQP) selected this measure.
Blood Pressure Management – Diabetes	Two pilots with provider-submitted data or established chart review process (WCHQ, MNCM) selected this measure.
Pneumonia Vaccination	One pilot (WCHQ) selected this measure, using administrative and electronic medical record (EMR) data only.

Measures Not Included	
Measure	**Reasons for Exclusion**
Antidepressant Medication Management – Continuation Phase	Sufficient historical pharmacy data not available for Medicare beneficiaries.
LVF Assessment	Measure requires chart review.
Cervical Cancer Screening	Measure not relevant to the Medicare population.
Tobacco Use	Measure is collected via a survey and is not available in administrative claims.
Advising Smokers to Quit	Measure is collected via a survey and is not available in administrative claims.
Influenza Vaccination	Measure as specified by AQA is collected via a survey and is not available in administrative claims.
Use of Appropriate Medications for People w/ Asthma	Measure not relevant to the Medicare population.
Asthma: Pharmacologic Therapy	Measure not relevant to the Medicare population.
Screening for Human Immuno-deficiency Virus	Measure not relevant to the Medicare population.
Anti-D Immune Globulin	Measure not relevant to the Medicare population.
Appropriate Treatment for Children with Upper Respiratory Infection	Measure not relevant to the Medicare population.
Appropriate Testing for Children with Pharyngitis	Measure not relevant to the Medicare population.

Available Measure Specifications

Once it was apparent that the AQA starter set would not provide sufficient measures for the project, we reviewed other measures. There are several performance-measure developers whose measures were considered for the BQI Project. We reviewed the **National Committee for Quality Assurance**[3] (NCQA), whose Healthcare Effectiveness Data and Information Set (HEDIS®) measurement specifications for managed care organizations have been in place since 2003.[4] In addition, we reviewed the American Medical Association's **Physician Consortium for Performance Improvement**[5] (AMA PCPI), whose measure specifications focus on physician-level measurement.

For each measure considered, we assessed whether it had been approved by the AQA and/or whether it had been endorsed by the **National Quality Forum**[6] (NQF), which seeks consensus on precise definitions of measures and the populations that are appropriate for them.

We also reviewed whether a candidate measure was under consideration or used by other measurement efforts, such as CMS' **Physician Voluntary Reporting Program**[7] (PVRP) and its successor, the **Physician Quality Reporting Initiative** [8](PQRI). Unlike PVRP, PQRI required

3 NCQA. Available at: http://www.ncqa.org/.

4 NCQA's physician-level HEDIS® specifications were in draft form at the time measures were selected for the BQI Project and not due for release until measurement year 2007.

5 Physician Consortium for Performance Improvement. Available at: http://www.ama-assn.org/ama/pub/category/2946.html.

6 NQF: National Quality Forum. Available at: http://www.qualityforum.org/.

7 Press Releases: Physician Voluntary Reporting Program. Available at: http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=1701.

8 Physician Quality Reporting Initiative: Overview. Available at: http://www.cms.hhs.gov/pqri/.

establishing a physician quality reporting system that offers incentive payment to eligible professionals who satisfactorily report data on quality measures for covered Medicare services.

Along with the specified measures established by these measure developers, the total number of actual specified measures is in the hundreds. However, although exact specifications may vary by developer, we found there are consistent disease and clinical topics across all developers, such as Diabetes and Heart Failure.

Three pilot sites (IHIE, MNCM, and WCHQ) also opted to include their pre-existing measure specifications, which were largely based on other measure sets such as NCQA HEDIS®. Appendix 3 provides a listing of all the measures considered for use in the BQI Project. All measures selected for the BQI Project were listed in Appendix A of the original Cycle 2 report. The complete Cycle 2 report is included for review in the present volume as Appendix 9.

While it was challenging to select measures that were applicable and appropriate for this project, we were encouraged by both the consistency and the variation. Having consistent measures allowed us to test the effects of minor specifications differences on the measures. At the same time, having variation in the measures allowed us to provide CMS with a clearer picture of which measures are most appropriate for future work.

Importance of Precise and Consistent Measure Defintions

The six BQI Pilot sites each brought a unique experience and perspective to the project. The variability of their experiences also resulted in differing decisions regarding the precise definitions of some of the measures calculated for care quality.

Variations existed in the data sources and methods of data collection; the measure specifications employed; the populations included; the age ranges for the measures; the measure denominator and numerator timeframes; and the medical professionals and specialties included in the analyses. We determined that these variations, detailed in the Cycle 1 and Cycle 2 report appendices (Appendix B of each report), must be taken into consideration when conducting any analyses or interpreting results based on the aggregated data, as they severely limit, if not preclude, any measure rate comparisons across the pilot sites. The variations are summarized in Table 2-2.

Table 2-2. Variations Across Pilots That Affect Measures

Variation	Description
Data Sources	Four out of the six pilot sites (CHIR, CCHRI, IHIE, and MHQP) used expanded administrative data (claims as well as some additional clinical data). The two remaining sites (MNCM, WCHQ) used a combination of expanded administrative data, hybrid data collection (e.g., chart review), and provider-submitted reporting.
Performance Measure Specifications	Four out of the six pilot sites (CCHRI, IHIE, MHQP, MNCM) used HEDIS®-based specifications for measure calculations. One site (CHIR) used a specific software package (EBM Connect, from Symmetry) as well as its own primary care physician imputation method. The sixth pilot, WCHQ, has its own specifications which are similar to HEDIS®.

Variation	Description
Populations Covered	All pilot sites incorporated the Medicare FFS data into their results. Four out of six pilots (CHIR, IHIE, MNCM, and WCHQ) included some Medicaid data (FFS and/or Managed Care) in their results. Three pilots (CCHRI, MNCM, MHQP) included commercial data, such as Health Maintenance Organization (HMO), Point of Service (POS), and Preferred Provider Organization (PPO) data. One site (IHIE) had commercial fee-for-service and managed care data, and the remaining site (WCHQ) included all payers.
Population Age Ranges	The age ranges for the measures across the pilots varied, depending on the measure. For example, for breast cancer screening, 5 out of 6 pilots used an age range of 52-69, while the 6th (WCHQ) used an age range of 40-68.
Measurement Periods	The measurement periods were mostly consistent across the pilot sites for all measures. One pilot (CHIR) used a 27-month time period for identifying diabetics (versus 2 years in the other pilots). For Breast Cancer Screening, one pilot (WCHQ) used 3 years' worth of data to define the denominator, whereas the others used a 2-year period.
Medical Professionals and Specialties	There were variations in the medical professionals to whom care could be attributed. In addition, the specialties included in the measure specifications vary. For the most part, pilots included Doctors of Medicine and Osteopathy in their attribution methodology. Some also included Physician Assistants and Nurse Practitioners. Others only attributed care to Doctors of Medicine and Osteopathy, but gave these medical professionals "credit" for care provided by other medical professionals.

The pilot sites felt strongly that the project would benefit from learning about variations that were instrumental to the success of their regional collaboratives

The pilot sites felt strongly that the project would benefit from learning about variations that were instrumental to the success of their regional collaboratives and the trusting relationships each site had built within their community. In the absence of a specific national standard (other than approval by the AQA and/or endorsement by the NQF) each pilot site made slightly different decisions regarding the definitions of measures. The pilots discovered that measure specifications had to be modified based on the source data available. For example, the HEDIS® specifications were built to be used with managed care (MC) encounter datasets, while other datasets may not contain the exact data elements or may not be defined the same way.

All of the pilot sites agreed that an important lesson from the BQI experience is that each measure's definition should be detailed and precise to allow for using the same specifications across all data sources. Harmonization of the definitions across regions would be necessary to make meaningful comparisons of results among regions.

The BQI pilot sites found that some nationally-recognized measures had specifications that were based on data elements commonly present in managed-care data, but not in other data sources such as Medicare FFS data or clinical records. To avoid having regional collaboratives make local decisions about measure specifications, a national consensus on measure specifications that are defined for varying types of data sources is needed.

Consensus Process Used to Arrive at Selected Group of Measures

The BQI Data Conquerors workgroup, comprising the technical and analytical representatives from the pilot sites, CMS, NCQA, and Delmarva, reviewed each measure in detail for applicability to the project. The pilot sites selected their measures based not only on data availability, but also relevance to and acceptance by the local community.

All participants understood and accepted that there would be methodological differences across the pilots. Each pilot's selection of measures and specifications was discussed on Data Conqueror calls. The group accepted these differences and acknowledged that it was important to address local community issues throughout the project.

Each BQI pilot site emphasized the importance of convening local stakeholders, including medical providers, at an early stage in establishing local collaboration. They also agreed that throughout the collaborative process, measure selection must be transparent and inclusive. Also, every effort must be made to achieve a consensus so that the results are considered valid.

3

A Comprehensive Picture of Physician Performance: Combining Medicare Fee-for-Service (FFS) and Other Data

One of the overarching goals of the BQI Project was to assess the effects of combining Medicare FFS data with other payer data for purposes of calculating and reporting measures of ambulatory quality. Each pilot combined Medicare FFS Inpatient, Outpatient, Skilled Nursing Facility, Drug Prescription (Part D), and Physician Office (Carrier) claims with other payer data in this project. In addition to claims data, each pilot used Medicare beneficiary enrollment and demographic files, physician and group files, and other health care provider of services files.

We sought to answer several specific questions:

1. Is it possible, from a technical and statistical point of view, to combine the Medicare FFS data with other datasets?
2. If technical issues are overcome, is the added information from Medicare administrative data useful in reporting to beneficiaries and providers?
3. Does the inclusion of Medicare FFS data affect the measures?
4. Does aggregating the Medicare FFS data, and reporting the results in an aggregated fashion, produce more meaningful results?
5. Does the addition of Medicare FFS data to other payer data increase meaningful results for individual providers who have small patient numbers?

Combining Medicare and Other Data: Lessons Learned

- Including information about Medicare beneficiaries is essential to creating a complete picture of quality of health care provided in a community.
- Lack of uniform dataset standards means aggregation requires constant, detailed attention, time and resources.
- Addition of new datasets as well as modification of existing datasets requires learning and rerun time.
- One cannot rely on administrative data to accurately identify physicians and physician groups; significant local effort is necessary.

In general, the inclusion of the Medicare FFS data in aggregation efforts has been useful and rewarding, especially for those pilot sites where previous results did not include the Medicare FFS population (CHIR, CCHRI, IHIE, MHQP, and MNCM). For WCHQ, the addition of the Medicare FFS data allowed them to test their all-patient measurement methodology against administrative data. All pilot sites agreed that including Medicare data gave a more complete picture of the quality of health care in their communities.

Before answering our key questions about combining Medicare FFS data with other data, which for this project is termed "data aggregation," we must first discuss the various approaches to data aggregation used by the pilot sites. These approaches are presented in Table 3-1, and are further detailed in Appendix 4.

Table 3-1. Data Aggregation Approaches

Aggregation Method	Advantages	Disadvantages	Pilot Experience
1. Obtain measure results (numerators and denominators) from payers by physician or physician group 2. Add measure numerators and denominators to obtain aggregated result by physician or physician group	• Straightforward method to obtaining aggregated results • Measure results can be generated for individual physicians which can then be rolled up to a group	• Definition and composition of physician or physician group across payers may not be consistent • Patient level data not available for measure validation or quality improvement efforts	CCHRI has used this approach in its existing measurement efforts. MNCM uses this aggregation method for its ongoing measurement efforts and also used it in BQI. MHQP uses this aggregation method for its ongoing commercial measurement efforts.
1. Receive and analyze de-identified patient level data for all payers 2. Produce measure results by physician or physician group	• Definition and composition of physicians or physician groups across payers is consistent • Measure results can be generated for individual physicians which can then be rolled up to a group	• Patients cannot be linked across payers, therefore patient level data for measure validation or quality improvement efforts involves separate patient listings for each data source	CCHRI used this approach for BQI. MHQP used this approach for BQI for the Medicare FFS and other payer data. MNCM uses this method for its hybrid measures.
1. Receive and analyze identified patient level data for all payers 2. Produce measure results by physician or physician group	• Definition and composition of physicians or physician groups across payers is consistent • Patients can be linked across payers which simplifies patient-level listings used for measure validation and quality improvement • Measure results can be generated for individual physicians which can then be rolled up to a group	• Can be very resource intensive	CHIR developed a Medicare-only database for the BQI project, modeled after its existing multi-payer, patient level database. IHIE added the Medicare patient-level data to an existing multi-payer data system.
1. Receive measure results from physicians or physician groups directly	• Measure results include all-payer results • Definition and composition of physicians or physician groups is consistent	• Validation of physician or physician group processes to extract and submit data can be very labor intensive	WCHQ receives all-payer results from the physician groups that comprise its membership. MNCM (For semi-annual Cycle 3 reporting only)

These varying approaches have allowed us to better understand the data aggregation process and the considerations that must be made when conducting such an effort. Our experience during the project led us to identify and answer five key questions.

1. Is it technically and statistically possible to combine the Medicare FFS data with other datasets?

Yes, the BQI Project discovered that it was possible to combine the Medicare data with other datasets. However, a lack of uniform dataset standards across sources means that the data aggregation process required constant attention to detail, a significant amount of time, and considerable human and computing resources. Working with new complex datasets (especially Medicare) requires time to fully understand how to apply them. Preparing any data for analysis, and in particular the Medicare dataset, involved defining a useful unit of analysis from the datasets (such as a physician visit); combining data elements from other databases (such as

beneficiary or provider files) with claims data; and identifying, removing, and modifying superfluous claims (such as duplicate transactions).

The initial data aggregation cycle (Cycle 1) was the longest. From the time the first pilot site received its files, it took approximately six months for lessons learned and challenges to be assimilated. The Data Conquerors shared lessons learned and acted as the final authority on data management and analytic issues for the project. Delmarva staff provided guidance based on their knowledge of the Medicare claims data; we also made extensive use of the documentation and technical support available from the Research Data Assistance Center (RESDAC)[1].

The BQI Project is the only project of its kind that aggregated various "raw" health care datasets from which to derive measures

The BQI Project is the only project of its kind that aggregated various "raw" health care datasets from which to derive measures. For three pilots (CHIR, CCHRI, and IHIE), the project has served as a catalyst for larger data aggregation and measurement initiatives. For the other three (MHQP, MNCM, and WCHQ), it provided yet another payer on which to measure physician group performance; it also supplemented already-existing efforts in physician group performance measurement and reporting.

In addition to the challenges with the Medicare data that we have already described, each pilot faced challenges with the other data it chose to include in the project. Issues related to other payer submission included missing data, incomplete data, and resource constraints (both human and technological). At CCHRI, for example, not one of its three other data contributors (health plans) successfully provided data in all three of CCHRI's data aggregation cycles. As the number of available sources of relevant medical data increases, the opportunity for delays due to at least one source being deficient becomes much greater. Consequently, the overall success of timely data aggregation becomes increasingly problematic.

The technical issues surrounding combining multiple source data are complex, detailed, and specific to each new dataset and change of format

CCHRI observed that for data aggregation projects such as BQI, far more time must be allowed for submitting and processing data. They also recommended that all interested stakeholders be involved from the inception of the project. CHIR, IHIE, and MHQP reported similar challenges with receiving and processing data from other payers. By the nature of their other/all-payer reporting, MNCM and WCHQ did not have these same data processing challenges. All pilots, however, agreed that additional time should be allotted for the data gathering process.

The technical issues surrounding combining multiple source data are complex, detailed, and specific to each new dataset and change of format. Data aggregation requires constant attention and involves a steep learning curve.

2. If technical issues are overcome, is the added information from Medicare administrative data useful in reporting to beneficiaries and providers?

Yes, the pilot sites found that having the Medicare FFS data during the BQI Project was essential to providing a complete picture of care in their communities. The Medicare patient population is a large part of every patient community, and beneficiaries generally consume a large per-capita

1 http://www.resdac.umn.edu/

proportion of health care resources. Performance measurement that does not include Medicare beneficiary information is only a fragment of community experience.

However, the currency of the Medicare FFS data, as compared to data from other sources, presents some challenges when reporting results to physicians, groups, or consumers. For example, IHIE's Quality Health First RM program[2] is able to create measure reports that are based on three-month-old administrative data and less than 30-day-old clinical and "point of care" data. When adding the Medicare FFS data, however, the IHIE reports become at least one year old, since the Medicare claims files for any given calendar year are not considered complete until roughly six months after the end of the year. In that community, providers feel that reporting of such dated information is not as relevant to the provider or to the patients. Additionally, physicians are reticent to spend valuable time reconciling or justifying old results, preferring instead to focus on doing what is necessary to improve care today.

The pilot sites concurred that for some measures with a long timeline, results calculated annually and current within nine to 18 months was sufficient. Examples of such measures could include the number of patients who had a colonoscopy within the last 10 years, or a mammogram within the last five years. For other outcome measures, providers may wish to make rapid improvements, and may need feedback results quarterly or even monthly. For rapid-cycle process improvement use, measures may need to be calculated by the providers themselves, or in combination with real-time administrative and lab data. Two pilots, CHIR and IHIE, do in fact receive administrative and other datasets on a monthly and quarterly basis, allowing them to generate reports that are more current. WCHQ will be enhancing its direct data submission program to allow up to monthly submission in 2009.

The addition of the Part D (pharmacy) claims data from Medicare allowed the BQI project to expand its list of available measures. As a "new" dataset, however, these data introduced additional questions regarding the completeness and quality of the data. CCHRI performed extensive analysis on its Part D measure results, which, when reviewed in conjunction with the results at other pilots (CHIR, IHIE, MHQP), led us to conclude that the Part D claims data may not be sufficiently mature at this time for measure calculation (See Cycle 2 Report). Medicare rates for measures that are pharmacy dependent may be understated due to apparently incomplete pharmacy claims data. The level of understatement probably varies from measure to measure.

The inclusion of Skilled Nursing Facility (SNF) claims in Cycles 2, 3, and 4 allowed us to test a claims-based method of excluding institutionalized beneficiaries. We tested this approach in response to providers' concerns that their performance measures be based on care of only those patients that they can affect. There was concern that primary care providers who no longer provide care for beneficiaries in institutionalized settings would still be held accountable. One pilot (MHQP), using the Medicare SNF claims to exclude institutionalized beneficiaries, found that the differences in the measure results were very small.

Regardless of the aggregation methodology chosen, it is critical that physicians and physician groups be accurately identified across all data sources. Doing so will ensure that measure results are accurate and credible. In the current state of data completeness it is often not possible to

2 http://www.ihie.com/about_us.htm

confirm whether members of a physician group in one database are the same as those defined as members of the same physician group by another dataset. In fact, spot checks by several pilot sites have demonstrated that physicians assigned to a group by one payer's data are slightly different at times than members assigned by another payer in the same market.

Such challenges encountered by pilot sites suggest that a national implementation using just Medicare data to assign physicians to groups may generate results that are not credible in local health care communities. Significant time and effort must be expended in verifying physician and physician group information. A feedback mechanism allowing providers to edit their group rosters may improve accuracy and the providers' sense of ownership. Implementation of the National Provider Identifier (NPI) may ease some concerns. However, full and exclusive use of this identifier did not occur within the timeframe of the BQI Project.

3. Does including Medicare FFS data affect the measures?

Yes, including Medicare FFS data does affect the measures.

MNCM performed specific analyses to assess the effect of the Medicare FFS data on its measure results. The site discovered first that the composition of the population being measured changed. Figures 3-1 and 3-2 demonstrate how adding Medicare FFS data altered the composition of the diabetes population being measured for MNCM during the 2007 reporting year (2006 data). After including the Medicare FFS data, the managed care populations (State Public Program, Medicare and Commercially Insured Managed care) comprised less than 50 percent of the diabetes measure population.

Figure 3-1. MNCM Diabetes Measure Population Prior to Medicare FFS Inclusion

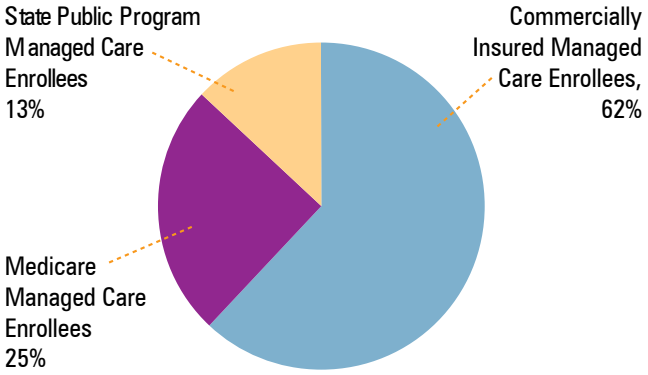
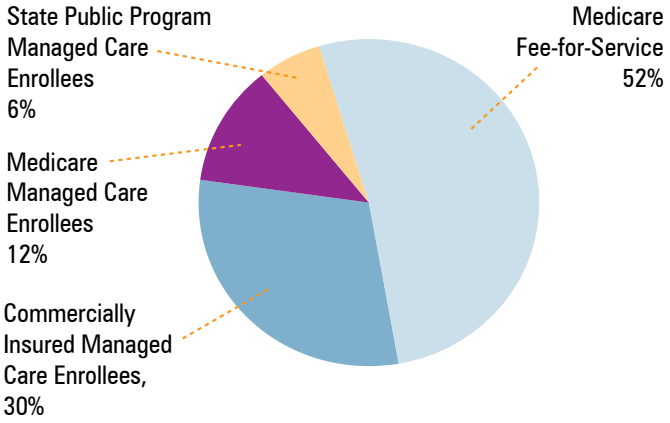
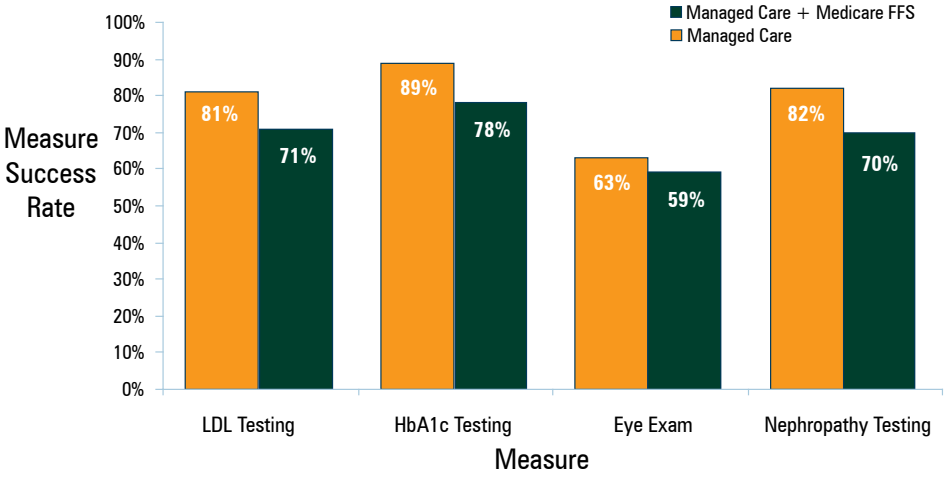


Figure 3-2. MNCM Diabetes Measure Population After Medicare FFS Inclusion



Given the dramatic change in the measure population (from solely a managed care population to less than 50 percent managed care), MNCM discovered that the measure results changed as well. As shown in Figure 3-3, MNCM diabetes measure rates dropped when the Medicare FFS data was added to the measure population.

Figure 3-3. Diabetes Measure Rates, Minnesota, 2006



In a subsequent analysis, MNCM discovered that most medical groups generally maintained their relative rank ordering for most measures, although there were some changes in the rank order for medical groups. This positive correlation suggests that providers are generally attempting to provide equal care to all patients. Further, it shows that other factors, such as system support mechanisms and patient demographics, may affect performance scores.

MNCM also discovered that the measure success rates for Medicare FFS beneficiaries are generally lower than other payers. Medical groups with a larger proportion of Medicare FFS patients will be most affected by measure performance for that population. In at least one measure, the highest performing clinic in MNCM became the lowest performing clinic when Medicare FFS data were added. Such a change in rank would carry a strong message to examine health system factors within the clinic that resulted in lower performance among their most elderly patients.

4. Does reporting performance measure results in an aggregated fashion produce more meaningful information?

The answer depends on the performance measure being considered. It also depends on whether Medicare patients' measure results are similar to results for other payers. Merging datasets is most useful when the measure results are similar and the rankings are correlated across the datasets. In other cases, results may be more meaningfully reported separately by payer.

If the provider's measure results are similar across all payer groups, then aggregating the datasets will result in a larger measurement population with more medical providers receiving a result reliable enough for reporting. On the other hand, if the performance measure's results for Medicare FFS are significantly different than those for commercial patients, Medicaid patients, or other population subgroups, then combining all the data together may result in misleading averages that de-identify disparities. In this situation leaving performance measures stratified by payer group or patient demographics may be more meaningful to providers and consumers alike.

The way Medicare FFS beneficiaries receive health care may differ enough from other care delivery systems that aggregated measure results may not accurately reflect the care being provided. The lag time we experienced between when services were provided and when the data are available for

analysis and reporting may decrease the relevancy of the results for some measures.
Because of these issues, the decision to report measure results combined across payers or separately by payer is best made on a case-by-case basis depending on the customer and the use of the information.

5. Does the addition of Medicare FFS data to other payer data increase meaningful results for individual providers who have small patient numbers?

Yes. However, the advantages of gaining a larger number of observations by aggregating data may be offset by uncertainties introduced through statistical processing. If Medicare FFS results are similar in success and strongly correlated by rankings when compared to other patient populations, then combining the populations will be meaningful and will increase the number of physicians with a large enough population to have a reported result. If, however, the results for Medicare beneficiaries are significantly different from those of other insurance populations, then combining the two datasets may give a result that is misleading for both patient populations and masks apparent disparities in care that should be investigated. This is true if the results are similar in magnitude but the rankings of providers are poorly correlated.

If Medicare FFS results are similar in success and strongly correlated by rankings when compared to other patient populations, then combining the populations will be meaningful

CCHRI's analysis revealed that despite adding the Medicare FFS data, small sample sizes affected physician-level rates in Cycle 2 (2006 measurement timeframe). Depending on the prevalence of the condition or event, the impact varies by measure.

When CCHRI analyzed the statistical reliability of physician measure results, they also learned that a score's reliability is influenced by the overall variance in physician scores and by the total sample size of the patient population. Specifically, requiring a minimum patient denominator size of 20-30 was not sufficient for all measures. Therefore, determining how the application of a reliability threshold influenced the percentage of physicians achieving a reliable score varied by measure. Our analysis in the following section examines the question of reliability in greater detail.

CCHRI further suggested several ways to increase the total number of physicians achieving reliability. These approaches included applying composite measures, pooling data over multiple years or longer time frames, incorporating data from more health plans, and summarizing results at a practice-site level.

In summary, the BQI Project discovered that it was possible to combine the Medicare data with other datasets. A lack of uniform dataset standards across sources meant that the data aggregation process required constant attention to detail, a significant amount of time, and considerable human and computing resources.

The pilot sites also found that Medicare FFS data was essential to providing a complete picture of care in their communities. Furthermore, including Medicare FFS data affected measures by changing the composition of the population being measured, and the measure rates. They also found that reporting measures results in an aggregated fashion may or may not be more meaningful, depending on the performance measure and on whether Medicare patients' results are similar to other payers. Lastly, the sites determined that the advantages of a larger number of observations by aggregating data may be offset by uncertainties introduced through statistical processing.

4

Analyses Addressing Methodological Questions

Within the first year of the BQI Project, the six pilot sites worked with new data sources and applied their varying methodologies to answer key analytical questions. A strength and a challenge for the BQI Project was that each pilot site had independently made programmatic decisions within their communities regarding the applied methodologies prior to the project; no two sites' choices were exactly the same. While creating an opportunity to capture diverse experiences, variations in methods also hindered our efforts to conduct a comprehensive analysis to answer several key performance measurement and public reporting questions.

A strength and a challenge for the BQI Project was that each pilot site had independently made programmatic decisions within their communities

To address the lack of uniformity in methodologies and the differences in research capabilities, we performed additional analyses based solely on Medicare FFS data. Our analyses, which examined the issues, tradeoffs, and strength of effects for several methodological choices, depended on two factors: 1) using only the Medicare FFS data, and 2) completing it within the project's time constraints and resources. We selected our focus areas based on their relative importance to advancing physician-level performance measurement and public reporting.

Our Analyses Focused on Four Interrelated Methodological Questions.

1. ***Attributing Beneficiaries to Physicians***—What are the options to attribute beneficiaries to physicians using administrative data only and the associated tradeoffs?
2. ***Attributing Physicians to Physician Groups***—What are the options to attribute physicians to physician groups and the associated tradeoffs? How are results affected by methods of requesting the data (i.e., physician-centric vs. patient-centric, in/out-of-state issues, methods to risk-stratify patients)?
3. ***Reliability of Physician Performance Results***—What is the minimum number of patients needed for a physician to achieve a reliable measure result?
4. ***Correlations of Performance Measure Rates Across Payers***—What are the issues around conducting and interpreting correlation across payers?

Chief among these questions is beneficiary attribution to physicians. Attributing beneficiaries to a physician or physician group is an important first step in measuring performance. Although beneficiaries can be attributed and measures calculated at the individual physician or physician group level, there are numerous reasons why measuring and reporting performance results at the physician group level is more meaningful or even necessary.

Another key issue is the reliability of measure results. Stimulating competition through the use of free market forces by reporting physician quality of care performance depends on valid and reliable results. For this reason, reliability analysis is the backbone of publicly reporting physician performance.

Other key questions, including the applicability of composite measures as a method to increase the denominator size for individual physicians and the field validation of the methods and measures were not explored. The rationale behind this decision was related to the complexity of the analyses required to answer the four key questions discussed in this section.

Several of the BQI pilot sites, however, have explored the use of composite measures. MNCM currently uses two composite measures independent of the BQI Project. These composites include Optimal Diabetes Care Measures and Optimal Vascular Care.

The field validation of methods and measures required the pilot sites to engage their physicians in the process. In the reporting of results section of the report, the issues and challenges with validating this information with the physicians will be described.

Analysis of Attribution of Beneficiaries to Physicians

Five Methods of Assigning Patients to Physicians were Analyzed

- Different rules for assigning patients to one or several doctors may be appropriate for different quality measures.
- Although process measure rates are statistically significantly affected by attribution rule, the magnitude of the effect is relatively small.
- The number of doctors with reportable results is higher if a patient's care is assigned to multiple providers.

Holding physicians accountable for delivering patient care assumes that care can be 1) validly assessed, and 2) reliably compared across similarly practicing physicians. The inherent validity of the care's assessment is promoted through the use of measures that are under the control of the providing physicians. Nationally recognized process measures (e.g., periodic HbA1c testing) are particularly useful for this purpose because their results depend more on the actions of, and orders from, physicians than they rely on patients' lifestyles and other choices beyond physician control.

The reliability of the selected measures for comparing the results across physicians depends on the attribution rules used to assign beneficiaries to physicians, the minimum patient counts acceptable for reporting, empirical state differences, the specific measure examined, and the actual rates and their distributions for the specific measures, themselves. Yet, of all the issues and factors involved, beneficiaries must be attributed to physicians before any physician rates can be calculated.

Of all the issues and factors involved, beneficiaries must be attributed to physicians before any physician rates can be calculated

Attributing beneficiaries to physicians is anything but straightforward. Depending on the measure considered, responsibility or "ownership" of care may reside with the primary care physician, a medical team including allied providers, and certain medical specialists. For each of the measures studied in the BQI Project, the types of providers eligible for responsibility of care (primary care only, certain specialists) were determined by the AQA, NQF, or other authority recommending the measure.

There is no national consensus regarding the appropriateness of assigning each patient to only one responsible provider versus assigning responsibility for a patient's result to all qualified providers who participated in the patient's care. Decisions regarding assigning a patient to one or to multiple providers may be different for different quality measures. To test for differences in results that could arise from the methodologies associated with both perspectives, we examined five differing attribution rules (i.e., algorithms) that are in common use. Two rules attribute care to all physicians who have a minimum number of evaluation and management (E&M) visits with the eligible patients

(one-touch, two-touch). The other three rules attribute care to physicians based on whether a certain proportion of the patients E&M visits were with the physician(s) (30 percent, 50 percent) or whether the physician had the most number of E&M visits with the patient (maximum frequency). Table 4-1 lists these rules, with a brief description of each. For information about the attribution methods used by each pilot site, refer to Appendix 4.

Table 4-1. Attribution Rules

Attributing Beneficiaries to Physicians with Administrative Data	
Rule	**Description**
Potentially Attributes One Patient to Multiple Physicians	
One-touch rule	Beneficiary is attributed to every physician touched at least once (based on the number of E&M visits) by the beneficiary.
Two-touch rule	Beneficiary is attributed to every physician touched at least twice (based on the number of E&M visits) by the beneficiary.
30% rule	Beneficiary is attributed to each physician who submitted at least 30% of the total office visits (based on the number of E&M visits) for that beneficiary – to a maximum of three physicians.
Attributes One Patient to Only One Physician	
50% rule	Beneficiary is attributed to the physician who submitted at least 50% of the total office visits (based on the number of E&M visits). If tied, 1 physician was randomly selected.
Maximum frequency	The physician with the highest number of claims based on the number of E&M visits) for a beneficiary is attributed to them. If tied, the most recently performing physician is selected.

Our analysis used Medicare FFS administrative data from six states (i.e., Arizona, California, Indiana, Massachusetts, Minnesota, and Wisconsin) for the years 2003 through 2006. We used five HEDIS® process measures in this analysis: breast cancer screening, colorectal cancer screening, HbA1c testing, LDL testing, and nephropathy testing. The five measures were selected because they were all being calculated across all six states. Delmarva used for the purposes of our analysis the HEDIS® 2007 specifications for all five measures. Statewide measure results for these five measures are presented in Table 4-2. Appendix 5 contains the complete measure results from our analyses.

Table 4-2. Statewide Measure Rates, 2006

	AZ	CA	IN	MA	MN	WI
	Rate (%) (Denominator)	Rate (%) (Denominator)	Rate (%) (Denominator)	Rate (%) (Denominator)	Rate (%) (Denominator)	Rate (%) (Denominator)
Breast Cancer Screening	57.5% (52,659)	54.0% (299,436)	54.9% (99,112)	64.5% (86,210)	65.9% (52,686)	64.3% (73,122)
Colorectal Cancer Screening	38.6% (265,562)	37.1% (1,324,729)	38.3% (463,760)	40.8% (381,786)	41.6% (267,757)	41.7% (364,606)
HbA1c Testing	73.7% (48,155)	77.0% (281,941)	78.3% (107,790)	78.7% (74,319)	85.2% (44,742)	86.9% (70,320)
LDL Testing	69.7% (48,155)	76.0% (281,941)	72.2% (107,790)	71.7% (74,319)	77.5% (44,742)	79.0% (70,320)
Nephropathy Testing	62.2% (48,155)	62.8% (281,941)	55.7% (107,790)	64.6% (74,319)	66.2% (44,742)	65.0% (70,320)

To better understand the impact of patient volume on the reliability of physician rates, we calculated measures for each attribution method with minimum patient-count rules requiring physicians to have at least 10, 20, 30, or 50 eligible beneficiaries attributed to them for their health care. Multiple regression analyses were used to untangle the effects of the five attribution

rules (one-touch, two-touch, 30 percent, 50 percent, and maximum frequency) from the unique effects of the process measures, and from the application of minimum acceptable patient counts, while controlling for overall state effects. These techniques are described in detail in Appendix 5. Using multiple regression analysis allowed for comparison of the relative strength of change effects in each factor described above. These analyses addressed three overarching questions:

1. Are physicians' process measure rates substantively affected by the choice of attribution rule? If so, what are the trade-offs from those choices?
2. Are the average reliabilities of the process measure rates substantively affected by the choice of attribution rules? If so, what are the trade-offs from those choices?
3. Are percentages of physicians who have the minimum number of patients needed, or the percentages of beneficiaries attributed to physicians who have the minimum number of required patients, substantively affected by the choice of attribution rule? If so, what are the trade-offs from those choices?

Following a discussion of each of these overarching questions, we provide recommendations for future work needed to better understand the ramifications of these choices as they are manifest in practice.

1. Are physicians' process measure rates substantively affected by the selection of the attribution rule?

Statistically, process measure rates are significantly affected by attribution rule, but the substantive effect is small

Only slightly. Statistically, process measure rates are significantly affected by attribution rule, but the substantive effect is small. Physicians' process measure rates are only slightly affected by the choice of the attribution rule. The largest adjusted difference found between any of the five attribution rules was the 2-touch rule. In such instances, measure results calculated using the 2-touch rule were 3.1 percentage points higher than measure results calculated using the 50 percent rule.[1] This amount is small compared to the variation of results among measures. The effect of the choice of measures from just the five examined can be as high as a difference of 38.7 percentage points between the results of two different measures. In other words, HbA1c tests for diabetic beneficiaries were more frequently documented than colorectal screening for eligible beneficiaries.

Of the four categories examined (i.e., minimum patient count, attribution rule, process measure, and state), the state generating the data was, overall, a less important contributor to differences in physician process measure rates than the specific process measure examined. However, for any given measure, the state generating the data was a more important contributor to the result than the attribution rule followed or the minimum patient count required. On average, of the five measures studied, the choice of state can result in a change in adjusted physician process measure rates of as much as 7.6 percentage points. Not all states showed substantively large differences in physician process measure rates. However, state-by-state variation in general overwhelms con siderations of attribution rule and minimum patient count as important issues for estimating or explaining differences in physician process measure rates.

These state differences cannot be explained by differences in the measures (i.e., all are appropriately structured HEDIS® measures), attribution methods, or minimum denominator counts. All

1 Adjustment is accomplished through the compensating effect of the multiple regression analysis.

of these factors were controlled in the multiple regression analyses. The reasons for the geographic variations in results may be multi-factorial and could require local as well as comparative study. Consequently, the choice among possible attribution methods of patients should be based on issues of relevance to beneficiaries and providers, and not the possible effect on measure calculation.

2. Are the average reliabilities of the process measure rates substantively affected by the choice of attribution rules?

No, the largest difference found within the five attribution rules tested was between the 2-touch rule and the 50 percent rule. In that case, the difference in reliability was 0.025—in the present context reliability is the extent to which providers' scores can be distinguished from each other, that is, the extent to which higher performers can be distinguished from lower performers. For many measures in health care, 0.70 is often considered an acceptable reliability. On this scale, zero represents no reliability and 1.00 represents perfect reliability. For almost all situations, the attribution rule choice has a negligible effect on the reliability of the calculated measures.

The largest difference found within the five attribution rules tested was between the 2-touch rule and the 50 percent rule

For almost all situations, the attribution rule choice has a negligible effect on the reliability of the calculated measures

There are important trade-offs relating to the eventual reliability of the process measure rates. Some of these trade-offs are expected because of the mathematics used to calculate reliability. In general, larger numbers of cases or observations, wider differences in the performance rates of providers, and higher levels of internal consistency in the chosen measures will each increase the resultant level of or value for reliability.

For the present work, minimum eligible patient count is an obvious factor to consider. Over the range of minimum patient counts tested (i.e., 10 to 50 eligible patients), the improvement in the average reliability of physician rates was 0.128, which is over five times the potential effect from a choice in attribution rules. Counterproductively, however, increasing the minimum acceptable number of eligible patients decreases the percentages of physicians and of beneficiaries assigned to physicians with a calculated process measure rate. Maximizing reliability by only including large practices would also make the result non-representative of the large number of physicians with practices that are smaller than the cutoff value. Coincidental with the loss of physicians, their patients' clinical information would also not be captured. This report next turns to discussing the size of these effects.

3. Are percentages of included physicians, or the percentages of beneficiaries attributed to physicians, substantively affected by the choice of attribution rules?

Yes, and to a large extent. Both the percentage of included physicians and the percentage of beneficiaries attributed to physicians are substantively affected by the choice of attribution rules. The loss of physicians when using the 50 percent rule instead of the 1-touch rule was 18.4 percentage points, controlling for all other factors previously mentioned for the multiple regression analyses. Importantly, the percentage of physicians lost due to the rules that attribute a single beneficiary to only one physician (i.e., 50 percent and maximum frequency) were always larger than losses to rules attributing a single beneficiary to multiple physicians. On average when compared with the 1-touch rule (i.e., the greatest number of potential attributions), attribution

Both the percentage of included physicians and the percentage of beneficiaries attributed to physicians are substantively affected by the choice of attribution rules

rules that only allowed for a single physician per beneficiary showed a relative loss of about 55 to 60 percent more physicians, on average, than attribution rules allowing for multiple physicians per beneficiary.

Compared with the 1-touch rule, physician and beneficiary losses with different types of attribution rules are due to the requirement for minimum patient counts needed for reliable reporting in conjunction with the type of attribution rule used. Only the 50 percent rule and the maximum frequency rule necessarily assign a patient to a single physician. The other three rules can, and most often do, assign patients to multiple physicians. These assignments to multiple physicians may increase the size of the physicians' eligible population. Attribution rules that assign patients to multiple providers generally result in more providers having a reportable result.

The outcomes of attribution rules that assign one patient to multiple physicians identify more physicians and their attributed beneficiaries. However, the effects of these types of rules could include eroding the specificity of some individual physicians' rates. If a researcher is not careful, these rates could appear as a multiplying of the total number of eligible beneficiaries.

The decision will affect the result for every assigned doctor, as well as the number of doctors with a large enough denominator of patients to have a reported result

For example, if 10 patients each saw 3 eligible physicians once and no other physicians during the time period studied, all 3 physicians would have 10 patients in their eligible population under both the 1-touch and the 30 percent attribution rules. Assuming a required minimum patient count of 10 patients, all 3 doctors would have the necessary minimum patient count. Summing the number of patients assigned to these 3 physicians gives us 30 patients, rather than the 10 unique patients that actually exist. Whether one physician or multiple physicians should meaningfully share responsibility for a patient's outcome is a decision that may be reasonably made for each distinct measure. This analysis shows that the decision will affect the result for every assigned doctor, as well as the number of doctors with a large enough denominator of patients to have a reported result.

Summary

On balance, the choice of the attribution rule selected and followed is:

1. Only modestly related to the variation in physician process measure rates;
2. Strongly related to the percentages of physicians and beneficiaries attributed to physicians with a calculated process measure rate; and
3. Substantively unrelated to the reliability of physician rates.

Compared with the 1-touch rule, the 50 percent and maximum frequency rules have the largest overall effects. This is due to the differences between attributing a single beneficiary to a single physician versus attributing a single beneficiary to multiple physicians. Importantly, the policy ramifications of the differences in the potential attribution rules are likely to cluster in two camps: those who favor the attribution of a single patient to a single physician versus those who favor the ability to assess coordinated efforts across multiple physicians. At the stage of development for this issue, it seems that a methodologically acceptable middle ground has yet to be found.

Attributing Physicians to Physician Groups

Defining Physician Groups to Report Quality Measures

The structural, procedural, and reporting changes in the delivery of health care have created a concomitant interest in measuring and reporting physician performance at the physician group level, rather than for individual physicians. In preparing for this potential change in reporting, the BQI Project was designed to address some of the technical challenges to forming and using physician groups in reporting physician performance rates with Medicare FFS administrative data, as a starting point. The reasons that some physicians would prefer to see this shift to group reporting and accountability include:

- Greater comfort with group-level reporting of performance measures due to the manner in which an increasing number of physicians refer patients within their colleague groups.
- Group reporting more accurately reflects how health care is actually delivered in a team model of care.
- Grouping also generates more observations and thus more stable/reliable estimates of performance.

Three Methods of Assigning Doctors Medical Groups were Analyzed

- Group UPIN numbers can give information specific to practice sites, but are not available on all administrative data
- TaxID numbers (TINs) are consistently available, but sometimes only specify very large corporate billing groups.
- All methods of assigning doctors to groups using claims data have an error rate.
- Methods should be considered to allow doctors to edit and correct rosters.
- NPI numbers may help in the future.

All physician-grouping methods need to consider: the impact on the resulting groups from issues arising from the content of the Medicare claims data; the manner by which the data are requested to be extracted (e.g., based on physician residence or patient residence); the extent of the needed look-back period; and the medical specialties to be included. Further, the conceptual models that define a group must be clarified and brought to a consensus before they can be nationally implemented. These models define, at a minimum, whether a single physician can be defined as a group, the number of groups that a physician is allowed to be in, and state and border alignment issues (e.g., limiting to physicians and/or locations licensed within the state or contiguous states). Given these challenges, the pivotal policy question becomes: *Is physician grouping with Medicare FFS administrative data sufficiently accurate and robust to support its use in reporting physician-group performance measure rates?*

To address this question, we identified three sequential, intermediate activities:

1. Specifying feasible methods for attributing physicians to physician groups using Medicare FFS claims and related CMS data;
2. Comparing the methods for attributing physicians to physician groups by analyzing the impact of the selected options on the number of groups that can be created, the number and proportions of beneficiaries and physicians attributed to groups, and the average numbers of physicians and beneficiaries per group;
3. Comparing the methods identified above against the pilot communities' groups that are validated by provider feedback and interaction.

Physician Grouping Methods

Table 4-3 presents the three most promising methods that we identified for attributing physicians to physician groups using Medicare FFS claims and related CMS data. These methods utilize the

Unique Physician Identification Numbers (UPINs) and Tax IDs from FFS claims data and group UPINs from related CMS data. The granularity of the group UPINs (i.e., group UPIN6, identifying overarching groups) can be controlled through the use of the group UPIN4, which identifies the specific geographical locations within each group UPIN6.

Table 4-3. Identified Methods to Define Physician Groups Using Medicare FFS and Related CMS Data

Name	Description
1. UPIN to group UPIN6	Each individual performing and/or referring UPINs linked to group UPIN6.
2. UPIN to group UPIN6 + group UPIN4	Each individual performing and/or referring UPINs linked to group UPIN6 + group UPIN4.
3. UPIN to Tax ID	Each performing UPINs linked with the Tax ID on the same claim.

The table shows that there are two methods of grouping individual UPINs to group UPINs (i.e., methods 1 and 2 in the table). These two methods differ by whether the greater granularity of group UPIN4 (narrowed to practice setting) is considered appropriate to apply to the analysis.

It is important to recognize that there is an inherent lack of specificity for health care quality improvement activities in the CMS datasets because their original design was to support claims payment. Administrative data was not designed for use in grouping physicians. This lack of specificity erodes both the reliability and the validity of the results when creating physician groups. As noted earlier, both group UPIN6 and group UPIN4 do *not* appear in Medicare FFS claims data. For grouping methods using UPIN as the base frame of reference (i.e., methods 1 and 2 in the table) the claims data must be merged with the "Supplemental Provider (Member) file by physician UPIN."

If a physician belongs to multiple group UPIN6s in the Member file and also provides services in more than one location, there is no systematic way of determining which patients should be assigned to a particular group UPIN6 among those locations. For this reason, all patients are assigned to every group UPIN6 to which the physician belongs. This results in double assigning of both physicians and patients. If the claims submitted by physicians were also to include both the group UPIN6 and group UPIN4, this issue would be effectively resolved.

For the Tax ID method, we also started by creating a list of eligible physicians and their UPINs. We merged this list to related CMS data and created groups with selected minimum numbers of physicians (e.g., at least 1, at least 2…). This method had fewer ambiguities within the assignment structures than had been encountered when using methods 1 and 2 from the table but was still not wholly satisfactory when assessed at the pilot sites.

Options of Physician Grouping Methods

We considered two policy-relevant eligibility conditions for attributing physicians to groups within the three methods for forming the overall groups. These conditions address the minimum number of physicians that can form a group (i.e., 1 vs. 2 or more) and an in-state limitation for the group practice (i.e., yes or no). The first of these options is important to consider because any methodology that requires at least two physicians to form a group will systematically remove solo practitioners from the analysis. Depending on the particulars of the comparison, the loss of

physicians could exceed 50 percent. The second of these eligibility conditions (i.e., limited to in-state or not) functions in a similar manner by expanding or contracting the physician and patient numbers and groups, depending upon the geographic scope allowed.

These two conditions are combined to form the four options depicted in Table 4-4. Importantly, policy decisions on the relative merits of the choices in these options impact all three methods of group identification.

Table 4-4. Options for Physician Group Attribution

Options	Description
Option 1	Any business location, at least 1 physician per group
Option 2	Any business location, at least 2 physicians per group
Option 3	Business location restricted to state of interest, at least 1 physician per group
Option 4	Business location restricted to state of interest, at least 2 physicians per group

Results

Table 4-5 shows the number of physician groups across the six states when there are no restrictions on business location (any state) or group size. The combination of group UPIN6 with the inclusion of group UPIN4 creates more groups than the other two methods. For more information on the effect on four options refer to Appendix 6.

Table 4-5. Number of Groups for the three Grouping Methods (option 1)

	Tax ID Method	UPIN to Group UPIN6	UPIN to Group UPIN6+ UPIN4
# of Physician Groups Across 6 States	83,872	59,704	89,660

Regardless of business location granularity, requiring a minimum of two physicians per group reduces the number of physician groups at least by 40 percent. The change of business location specificity (i.e., from any state to state of interest) reduces the number of physician groups by approximately 50 percent. Simply, group UPIN6 combined with group UPIN4 method generates the largest number of physician groups because the group UPIN4 is a location-based subgroup of group UPIN6. If a physician were a member of a group based on the group UPIN6, practiced at multiple locations, but provided care at only one of these locations, this method would attribute all patients to all locations, although the service was provided in only one location.

The question then becomes one of finding a proper frame of reference to assess the correctness of the groups formed by using Medicare FFS claims and related CMS data. During the year, Delmarva requested that some pilots submit lists of their physicians groups to compare to groups created from administrative claim data. Although several pilots submitted a list, for different reasons, the submitted lists did not include physician-grouping information (Tax ID or group UPIN6). Hence, we were not able to conduct a thorough analysis of or comparison between the pilots' groups and the groups created from using only administrative data.

Nonetheless, WCHQ submitted a list which included Tax IDs representing 76 groups. We matched the Tax IDs of the 76 groups submitted by WCHQ (WCHQ groups) with 5,103 Tax

IDs for groups created from Medicare claims data (claim-created groups) and found that 40 of the Tax IDs matched. We then sent the list of the 40 Tax IDs and the physicians in each of the claims-created groups to WCHQ for validation. WCHQ's analysis found that only 78.7 percent of those physicians in claims-created groups actually belonged to the WCHQ groups. Eight of the nine providers submitted details regarding the physicians who were in WCHQ groups but not in the claims-created group or vice versa. According to WCHQ analysis, 343 physicians were in WCHQ groups but were not listed in claims-created groups, representing about 14 percent more physicians than had been identified in the claims-created groups. On the other hand, about 6 percent of physicians in claims-created groups were not members of the WCHQ groups.

Importantly, the 40 WCHQ groups that matched with claims data only represent about 1 percent of all the groups created from claims data, or about 27 percent of physicians in the claims-created groups for the state of Wisconsin. However, WCHQ groups are limited to only those provider groups with whom WCHQ works and do not represent all groups for the state of Wisconsin. This finding, therefore, should be interpreted with caution as it is not a representative sample of Tax ID groups in Wisconsin or of all WCHQ groups. Compared to all claims-created groups in Wisconsin, WCHQ groups are larger and have a higher number of physicians per group.

Wisconsin's analysis supported the earlier experience of MHQP in Massachusetts. Initially, MHQP had calculated group-level results. These results were based on physician grouping determined by each of several commercial payers in their market. However, MHQP noted that the physician memberships of the medical groups did not agree across varying payers. MHQP created an interactive process by which the medical groups review the assigned physician membership of their groups and make corrections as their memberships change. They found that all of the payer-created group rosters had significant inaccuracies. MHQP has found that allowing provider correction of group membership improved accuracy and garnered buy-in to the results. As a side benefit, it has also created a point of contact with the staff of the medical groups and built trusting relationships.

Conclusions Regarding Assigning Physicians to Groups Using Administrative Data

Is physician grouping with Medicare FFS administrative data alone sufficiently accurate and robust to support using it to report physician group performance measure rates? The results of the BQI Project suggest that the answer is: not quite

Is physician grouping with Medicare FFS administrative data alone sufficiently accurate and robust to support using it to report physician group performance measure rates? The results of the BQI Project suggest that the answer is: not quite.

Currently, the assignment of physicians to physician groups using Medicare administrative data, or any claims-based data alone, is not sufficiently accurate. To gain accuracy in reflecting actual medical group rosters, all the pilot sites of the BQI Project recommend validating the group rosters with the physician groups. The sites also recommended an interactive process that allows correction, additions, and deletions over time.

Nonetheless, with the right changes and quality improvement approaches to the challenges surrounding accuracy of data, combined with focused additional research, we believe that claims data could serve as an effective method of grouping physicians for public reporting of some

performance measure rates. For example, consistent use of the newly-implemented National Provider Identification (NPIs) numbers will likely improve accuracy of group attribution.

In markets dominated by very large medical groups, using the TaxID of the corporate group to calculate group-level measures has been acceptable to those providers. As the situation currently stands, the UPIN to Tax ID method is better positioned to group physicians compared with UPIN to group UPIN methods. The reason for this is both the Tax ID and UPIN are included in the claims submitted by physicians, while group UPIN6 and UPIN4 are not.

Another issue of great importance is the need for involving physicians in the process of defining methods for physician grouping. Unless the methods are transparent and acceptable to physicians, progress in this area and in measuring physician group performance will be significantly delayed.

With the right changes and quality improvement approaches... we believe that claims data could serve as an effective method of grouping physicians for public reporting of some performance measure rates

Reliability of Physician Performance Results

To spur quality improvement across broad areas of public health, both private and public entities are demanding provider accountability and rewarding those who meet certain stated quality goals. Public and private purchasers are also demanding more information about the quality of health care services from providers and hospitals.

Analysis of the Reliability of the Rankings of Providers by Measure Results

- Patient populations of 30-50 per doctor were most commonly needed for a reliable distinction to be made.For some measure results, patient population sizes of much more than 50 may be required.
- Excluding doctors with less than 50 patients for a certain measure may result in many doctors not reporting.

The purpose of this section is to describe the reliability of available information for making decisions about provider performance. Before reporting the results for the reliability of the provider performance data across the six states reviewed with Medicare FFS data, a brief discussion is in order on the interpretation of reliability in general but, especially, within the present context.

Reliability is consistency, and in the present context is the extent to which providers' scores can be distinguished from each other (e.g., higher from lower performers).[2] The scale for reliability forms a continuum from zero to one.[3] Higher reliability increases the likelihood that a provider or a provider group is correctly assigned, for example, to high, medium, or low performing categories of providers. The appropriate usage for the term, therefore, is one of "sufficient" reliability for a stated purpose, and not as an absolute construct that would suggest either a "yes" or "no" answer. Having sufficient reliability, though, is not the same as having appropriate validity for a given task. Performance measure results must first be ensured as valid before reliability calculations are begun. For the purposes of the current research on reliability, an appropriate level of validity for the measure is assumed.

Reliability is consistency, and in the present context is the extent to which providers' scores can be distinguished from each other

The level of reliability that is considered "sufficiently high" is quite subjective and can vary substantially depending upon the domain. The choice of precise cutoff value, then, raises questions about the trade-off between feasibility and scientific soundness. Researchers have generally found that a reliability of 0.70 or higher is typically acceptable for provider performance measurement purposes in public health. In general, larger numbers of cases or observations, wider differences

2 Beta-binomial reliability was used for the current research. Interested readers can see Appendix 7.

3. Zero represents a total lack of consistency, whereas 1.0 represents perfect consistency.

Larger numbers of cases or observations, wider differences in the performance rates of providers, and higher levels of internal consistency in the chosen measures will each increase...reliability

in the performance rates of providers, and higher levels of internal consistency in the chosen measures will each increase the resultant level of or value for reliability.

Reliability Results

To investigate the issues surrounding the reliability of performance results, we performed test analyses using complete Medicare FFS databases for CY 2006 for each of the six states represented by BQI pilot sites. For this investigation only Medicare databases were used, no commercial or pilot-site data was considered. Providers were categorized by eligible patient populations of 10 or more, 20 or more, 30 or more, and 50 or more for each of the five performance measures. We analyzed the effect of each of the five different types of provider attribution rules (1-touch, 2-touch, 30 percent, 50 percent, and maximum frequency). Reliability scores were calculated for the five measures that were used across all six BQI pilot sites, under each of the different attribution rules, and for each of the eligible population sizes for each of the six states.

The results of the analyses on the resulting reliabilities scores strongly suggest the need for very large patient populations to consistently produce sufficiently reliable measures using the current rate production methodology. The 1-touch attribution rule is used for the present discussion to illustrate this point. The reason for selecting the 1-touch rule is that it assigns patients to the largest number of providers. By doing so, this method results in the inclusion of more providers, when larger minimum patient counts are required. However, the choice of attribution rule does affect the reliability, and other attribution rules may have higher reliability for certain measures. See Appendix 7 for more detail.

Using the 1-touch rule, no state saw a single one of the five comparative measures reach an acceptable level of reliability (i.e., at least 0.70) for either 10 or more or for 20 or more eligible patients. This finding means that there were no state conditions where fewer than 30 eligible patients formed even a single rate with a reliability that reached 0.70 for the five measures. Between half and three quarters of the providers in each state did not have enough patients in their denominators to report a result. Further, these results indicate that between a fifth and half of all patients are also lost while using the most liberal attribution rule to retain as many providers and patients as possible.

When using at least 50 or more eligible patients in conjunction with the 1-touch rule, only the Nephropathy measure reached an acceptable level of reliability across all six states. This measure was followed by the Colorectal Cancer Screening measure, which reached an acceptable level of reliability in five of the six states. No other measure reached acceptable reliability in more than three states with any tested minimum number of eligible patients.

The HbA1c testing measure for diabetic patients only reached an acceptable level of reliability for one state and needed at least 50 eligible patients per provider to do so. Breast Cancer Screening only reached acceptable reliability values for two of the six states and only when at least 50 eligible patients were required. The LDL Screening measure only reached an acceptable reliability in three of the six states and, again, only when at least 50 eligible patients were required per provider. Of the five measures tested in six states (i.e., 5 x 6 = 30 tests), 13 measure-state combinations (i.e., 43%) never yielded a sufficiently reliable result with any tested minimum number of eligible patients.

Discussion

This analysis results in two important conclusions about the ability to reliably distinguish between the performance of providers. First, when a reliable rate was found, the number of patients in the denominator for each measure needs to be in the range of 30 to 50 patients, but there is wide variation depending on the magnitude of the measure results and the difference between providers. Conceivably, some measures may achieve a reliable distinction in ranking with fewer than 30 eligible patients. However, for many measures the necessary denominator may be much larger than 50 patients.[4] Larger numbers of cases or observations, wider differences in the performance rates of providers, and higher levels of internal consistency in the chosen measures will each increase the resultant level of or value for reliability.

Secondly, for many of the performance measures currently studied, requiring a single provider to have even 50 eligible patients excludes many, if not most, of the providers in each state studied. With these measures, requiring a large enough patient population to distinguish between physicians' performance would result in most patients not having any reported information about their doctor's performance.

Strategies to Improve the Reliability of Provider Rankings

Several strategies are suggested to address the issue of needing larger numbers of patients when using measures to distinguish between providers.

The primary strategy used by the BQI pilot sites is to report measures at the level of the medical group, resulting in larger numerators and denominators. The issues and trade-offs surrounding group-level versus individual doctor-level reporting are discussed in the sections that follow in this report.

Developing broader quality measures that are applicable to a larger segment of each provider's patient population would also increase the reliability of comparisons. Composite measures, disease-specific measures, or generalized care management measures would be examples.

Research from public health and other industries may suggest strategies for meaningfully distinguishing between outcomes with low prevalence. Specifically, methodologies that have been designed for, and tested with, small denominator situations in health care need to be assessed for their applicability in the current arena. For example, the nation's smallest QIO is the Virgin Islands Medical Institute (VIMI). VIMI has been a QIO for many years and has continuously experienced very small denominators for most of their performance measures. CMS has wrestled with appropriately reliable evaluation methodologies for VIMI for at least the current and previous two Scopes of Work.[5] By using small denominator-focused evaluation methodologies, the Virgin Islands QIO has been appropriately evaluated for the previous two Scopes of Work and is entering a third.[6] These methods in particular should be assessed for their transferability to, and for their applicability in, the current area.

4 Reliabilities reaching 0.70 for 30 eligible cases were attained in only 5 of 30 instances tested (17%) using conditions designed to yield the highest likelihoods of success. Further, the highest reliability for 30 eligible cases reached for any of the five measures for any of the six States was 0.72, with the 1-touch method, and were as low as 0.29 for the HbA1c measure in Minnesota.

5 Currently in the 9th Scope of Work, each Scope of Work lasts three years.

6 Specifically, these methodologies relied on repeated, appropriately weighted, rolling measures in conjunction with linear trending in sequential cross-sectional designs.

Finally, larger numbers of patients may be obtained for each provider by combining datasets across payers. In this way the identified number of patients per provider can be effectively increased. In some instances, the number can increase substantially, as when Medicare FFS data can be combined with commercial data. This strategy is most useful when the measure results are similar for the various combined datasets. This alternative is explored to the extent that the data were available in the analysis of correlation of performance results across payers discussed below.

Correlations of Performance Rates Across Payers

The previous section on the reliability of the current performance measure rates concluded that either a more statistically efficient performance measure rate production methodology is needed or, that larger numbers of patients per provider or aggregate or composite measures would be needed. One method to increase the effective number of patients per provider is by combining performance measure datasets across payers. A common method to determine the validity of combining the datasets is through an examination of the correlations of performance measure rates for the same providers, on the same measures, but across payers. If the correlations are sufficiently high, the datasets can be validly combined.[7]

Correlations of Performance Rates Across Payer

- High correlations between performances of differing payers can be used to justify combining data sets to achieve larger patient populations.
- Low correlations were often found by pilot sites. Factors can include provider behavior but also managed care effects and patient demographics.
- Analysis of reasons for low correlations must include many local factors.

The results show a wide range of correlations across the states for the different performance measures, attribution rules, and minimum patient counts. In exploring the potential reasons for several quite low correlations (i.e., approximately zero to ~0.3), "systems effects" emerged as a potential explanation.[8] For example, with MC, many providers have staff members who are responsible for calling and scheduling lab procedures and then arranging any needed transportation for patients. These providers, or provider groups, would likely have higher performance measure rates than providers who cannot make those arrangements for their patients.

Comparing performance rates for providers' FFS patients with the rates for their managed care patients is, therefore, not solely a measure of the difference in providers' behaviors

Structurally, managed care plans are designed to supplement provider effectiveness through the types of support that stand-alone providers could not as often offer. Comparing performance rates for providers' FFS patients with the rates for their managed care patients is, therefore, not solely a measure of the difference in providers' behaviors. Rather, it is a measure of potential system differences in effectiveness that include but is not limited to provider behaviors. This distinction means that correlations of provider performance measure rates across payers could represent the relationships between provider-associated health care delivery systems, making them essentially uninterpretable within the context of estimating provider-specific consistency across payers. If so, it would mean that the databases should not be combined in efforts to increase effective eligible patient counts.

7 The actual level of a sufficiently high correlation is not traditionally set, but 0.85 would likely be within 0.10 of an acceptable value to many biostatisticians and epidemiologists in public health.

8 System effects are used as a general term for all non-physician-related provider effects.

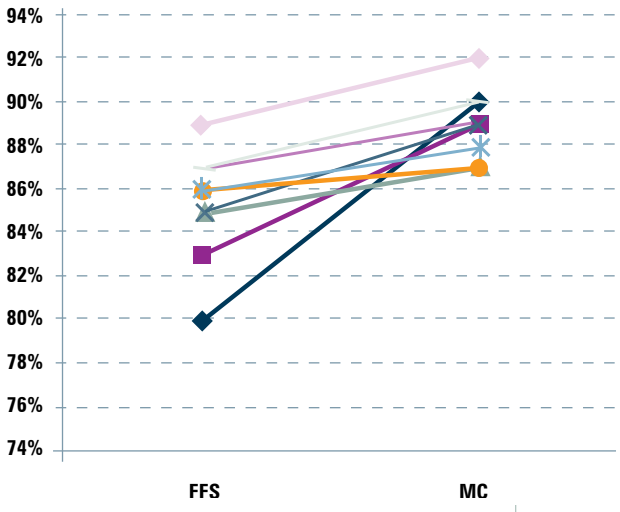
An Example of the Effects of Managed Care Systems on Provider Rankings

Table 4-6 and Figure 4-1 demonstrate a hypothetical system effect for ten different provider profiles.[9] We present the hypothetical performance measure rate for each provider, followed by the effect of the MC system for the providers' MC patients, and then the providers' MC rate (assumed to be additive). Importantly, system effects are assumed to have differential impacts across providers. The assumption made for a differential effect is consistent with the fact that various payers have different support systems in place. Further, various provider-patient combinations would likely differentially access system offerings in care due to differences in personal preference and patient compliance—in short, varying systems' effects.

Table 4-6. Hypothetical System Effects

Provider	FFS	System Effect	MC
#1	80%	10%	90%
#2	83%	6%	89%
#3	85%	2%	87%
#4	85%	4%	89%
#5	86%	2%	88%
#6	86%	1%	87%
#7	86%	1%	87%
#8	87%	2%	89%
#9	87%	3%	90%
#10	89%	3%	92%

Figure 4-1. Hypothetical System Effects



In this example, the MC system effect varied from 1% to 10%. The average effect size was 3.4 percentage points. Further, the system effect is somewhat larger for the lower FFS scores than the system effect for the higher FFS scores, as might be expected when the amount of room for improvement is considered. The net outcome of these system effects is to generate a correlation between the FFS rates and the MC rates of only 0.134, which indicates that the differences between the FFS rates and MC rates are substantively random, or not highly correlated.[10] Notably, this low correlation is not related to or influenced by there being only 10 cases in the table above. The number of cases is a statistical power issue and is related to statistical significance (i.e., the p-value), rather than to the magnitude of the correlation, itself.[11]

This hypothetical example highlights just one of the perils of correlating provider performance measures across payers. Provider behavior may not be correlated against itself but, rather, with the results of a quality-rearranging system. Realizing this potential shift in the construct being measured, the interpretation of the magnitude of the correlations also shifts. In this instance, one

9 Hypothetical data need be used because vetted data that distinguish system components do not yet exist.

10 The argument could be made that the performance measure rates' natural ratio scales, which support Pearson Product-Moment correlations, should be downgraded to ordinal levels of measurement due to the use of star-rating systems. Under this assumption, the correlation increases to 0.219, but is substantively still too small to suggest that the datasets could be combined.

11 Given that the presented cases are representative of the overall population of cases.

valid but not exclusive interpretation is the extent to which an MC system is proving itself to be effective, versus the FFS scenario. This interpretation suggests that low correlations could indicate effective MC systems. The lower the correlations, the better the MC system might be functioning to re-order and somewhat negate individual provider effects. That reordering and partial negation of provider effects might be considered a good outcome for policy, by removing individuals' idiosyncrasies from the provision of health care.

Whether low correlations across payers for some measures is expected from MC and other system effects or is due to differences in care, access, or patient factors requires further research at the local and regional levels. The BQI analysis demonstrates, however, that combining datasets to achieve larger numbers of patients is only appropriate when the results are highly correlated and at similar levels across the populations in the datasets.

Other methodological and logistical issues can also erode confidence interpreting low correlations in the present context. For example, minimum patient counts make a moderately large difference in the correlations found across payers. When looking at correlations of actual performance rates across payers for providers in Minnesota, correlations increased about one quarter of the potential scale when the minimum patient count was increased from one to ten eligible patients. Also these types of changes in the magnitude of a correlation from a change in eligibility requirement make judgments about lower correlations perilous.

Further complicating the interpretation of correlations of provider rates across payers is the use of provider groups. The imprecision ascribable to assigning providers to provider groups, all within different payer structures, is further complicated by the attribution of patients to providers. Except in the case of high correlations, all of these factors erode whatever confidence there may be in an interpretation of a correlation of provider performance across payers to be a measure of provider consistency in the delivery of quality care.

Summary

In general, the correlations of performance measure rates across providers were neither stable, nor robust—two important characteristics that are generally required for national reporting. Yet, nothing in these findings on the lack of interpretability for low correlations across payers suggests that there is anything unreliable or invalid about these types of data for use within payers. It only suggests that rates with low correlations across payers should probably not have their datasets combined. The potential to combine datasets for larger patient counts per providers is still tenable when the correlations are sufficiently high and the actual performance rates are similar.

5

Reporting of Results to Physicians and Consumers

A primary goal of the BQI Project was to study how best to report quality measures—especially to Medicare beneficiaries and providers—and to assess the usefulness of the reported measures. To that end, the project offered a unique opportunity to gain insights into issues surrounding the reporting and use of quality measure data.

Two factors in particular contributed significantly to these insights:

1. The six pilot sites were at different stages of maturity in their reporting capabilities.
2. The sites made different decisions regarding reporting and data use in their various regional markets.

In this section we review the potential audiences or customers for quality measures, and the varying uses for the data each customer may have. For each use we specifically address several factors: the type of measures preferred; the level of detail needed; the frequency and currency of the necessary reporting; and whether the measures were reported at the level of the individual physician or at a more aggregated level (e.g., a medical group).

Differing Stakeholders Have Differing Needs and Preferences for Reported Quality Measures

- For most purposes, annual data reports are sufficient.
- Providers prefer the most detailed and current data on measures, and, for some measures, at a monthly or quarterly frequency.
- Patients and payers wish for individual provider data, but only if the results are valid and reliable. Group level is OK.
- Patients and payers prefer broader, disease-related composite measures that are easier to understand.

Customers for Reporting, Uses for Information

The audiences or "customers" for information about medical care quality are broadly divided into medical providers and health care consumers. Medical providers include physicians and other direct caregivers, medical managers of hospitals and offices, and the senior leadership and governing boards of medical organizations. Health care consumers include not only Medicare beneficiaries and other patients but also employers, health plans, policy-makers, and the media.

In the rest of this section we review each audience, describing its possible uses of data, the level of reporting desired (e.g., individual physician, physician group), the best frequency of reporting, and the preferred types of measures (e.g., process/outcome, individual measure/composite measure). We report, compare, and contrast the experience of the six BQI pilot sites in reporting measures to differing customers.

When comparing the experience of the BQI pilot sites with reporting, it is important to remember that they have made different decisions in their respective regions regarding each of these reporting issues. In addition, the sites are at different development stages in the lifecycle of their regional collaboratives. Some have been reporting to physicians and publicly for many

When comparing the experience of the pilot sites with reporting, it is important to remember that they have made different decisions in their respective regions regarding each of these reporting issues

years. Others only recently began reporting quality measures back to physicians, gaining their validation, and assessing their utility. Table 5-1 summarizes the past and current development of reporting practices among the six sites, and is referred to as their experience with each customer is discussed.

Table 5-1. Pilot Sites' Experience with Reporting Performance Measures, Past and Present

	Regional Collaborative Start Year	Year of First Feedback Reporting to Physicians and/or Groups	Year of First Public Reporting	Year of First Pay-for-Performance
CCHRI	1993	2001	2001	None
CHIR	1990	2008	None	None
IHIE	2004	2007	2008	2009
MHQP	1995	2002	2005	2006
MNCM	2001	2003	2004	2006
WCHQ	2002	2002	2003	None

CHIR and IHIE are involved in quality measurement efforts that build on extensive community databases. The other four pilots began specifically to measure medical quality in their communities, and have many years of experience

CHIR and IHIE are involved in recent quality measurement efforts that build on extensive, longstanding community databases. The other four pilots are regional collaboratives that began specifically to measure medical quality in their communities, and have many years of experience.

There are a number of other unique characteristics for each BQI pilot site as well. For example, MHQP has created an interactive web-based tool for providers so that physicians' offices can actively update their medical group memberships at will. MHQP uses this consistent provider registry to calculate group-level measure results in an accurate manner, and the physicians gain a sense of engagement and ownership of the process.

A distinct feature of IHIE was that it was specifically created to measure and report quality measures in 2004 using data from the Indiana Network for Patient Care (INPC). This regional database of clinical information was created in 1969 at the Regenstrief Research Institute. IHIE, through its Quality Health First RM program, began supplementing the clinical data with commercial and Medicaid claim data in 2007.

CHIR, a university-based research center with community partners, began calculating quality measures with the BQI project. CHIR grew out of a public health research program that began in 1990 at Arizona State University. The site uses a regional database, created in 1999, called Arizona HealthQuery (AZHQ). AZHQ contains patient-centric data of medical episodes regardless of payer across Maricopa County and much of Arizona, and allows for a longitudinal picture of health across provider and payer for much of Arizona.

WCHQ is the only pilot that collects all data from the provider groups that are members of their organization. This data collection methodology has also been recognized in MNCM, which has started collecting some measures in this manner and will be collecting more. The data collected from the physician groups allowed WCHQ and MNCM to provide outcome measures for the BQI Project. Using this methodology allowed WCHQ access to adequate look-back periods for

the Cancer Screening measure exclusionary criteria providing a more accurate patient list. CCHRI is attempting to publicly report individual physician-level performance data measures to complement medical group-level measures that are already publicly reported for managed care patients by the state of California.

Lastly, in some pilots (MNCM, WCHQ, IHIE, MHQP) medical groups receive reports with individual physician data. They may use these data for internal quality improvement and/or incentive purposes, though public reporting is only at the group level.

Experience with Use of Quality Data Reports by Providers

Individual physicians, other direct patient caregivers, and medical managers use quality measure results primarily for performance improvement purposes. They may also wish to compare themselves to other providers. In addition to professional motivation, such comparison helps providers identify currently achievable care benchmarks and offers insights into how to obtain them.

To improve quality, direct care providers and managers prefer data as current as possible. For measures that reflect rapidly-changing outcomes or processes, reporting quarterly or even monthly is useful to allow rapid management of improvement. However, many of the chronic care and preventive quality measures involve data look-back of 1-5 years. Reporting more frequently than annually on these measures is of little benefit for quality improvement purposes. For performance measures that change slowly over time, annual reporting is adequate and avoids a burdensome data workload. When very current and short-cycle measurements are needed for process improvement, managers obtained the measures from internal sources. These internal sources included their own administrative data, enhanced administrative data such as lab and pharmacy information, and data from clinical chart extraction.

To improve quality, direct care providers and managers prefer data as current as possible. For measures that reflect rapidly-changing outcomes or processes, reporting quarterly or even monthly is useful to allow rapid management of improvement

Managers may also use quality data to create incentives for their staff. Some medical groups in the Wisconsin, Minnesota, California, and Massachusetts markets used individual physician-level information internally both for comparison among employees as well as for incentive pay or bonuses.

The Boards of Directors (BODs) and senior management of health care institutions also use quality measure data for comparison and accountability purposes. Hospital BODs are required by the Joint Commission to receive regular reports about the quality of care provided, and in some cases, increases in the salaries of senior managers are linked to specific quality measures. The pilot sites in Minnesota and Wisconsin confirmed that medical group BODs also used the reported measures in this way.

Governing boards emphasize quality and safety in their organizations, and provide the necessary resources for improvement. Direct care providers in Minnesota and Wisconsin felt that reporting quality measures—particularly when combined with public reporting, pay-for-performance, or both—focused the attention of their governing bodies on issues of quality and safety and made obtaining necessary resources easier.

Medical groups in Wisconsin and Minnesota have reported improvement in both processes and quality measure scores using rapid-cycle data feedback. Additionally, IHIE is using their large electronic clinical database to calculate some measures as currently as two to three months. The Indiana pilot site is now in the process of validating these results with their providers.

Medical providers also use quality reports to negotiate improved payment rates with insurance plans. MNCM reports several cases in their market of medical groups using their quality measures in advertisements for patients and for prospective new doctors.

An important use of administrative data is the augmentation of clinical records to identify opportunities to improve patients' care and prevent harm. Clinical records have broader information...yet the two data sources are often complementary

An increasingly important use of administrative data by providers is the augmentation of the clinical records to identify opportunities to improve patients' care and prevent harm. Clinical records have broader information than data for administrative and billing purposes, yet the two data sources are often complementary. For example, a doctor may believe that a patient is taking his/her medication regularly, while the pharmacy benefit plan is aware that no medication has been charged for many months.

Health plans are increasingly aware of their active role in improving patient care, and are taking steps to make all possible information available to the responsible medical caregiver. IHIE has best demonstrated the power of a patient-centric database that combines clinical and administrative data from multiple sources for the benefit of patient care.

Experience with Use of Quality Data Reports by Consumers

Accurately assessing how individual patients use publicly reported quality measures is difficult. Although the pilot sites concur that public use of their sites is lower than they wish, they indicate that public interest in their reports is steady. Public traffic on websites that report quality data increases for new results, during annual selection of panels of physicians, or in response to media attention.

Beginning in 2001, WCHQ's public website has experienced traffic in the range of 1,400-2,300 visits/month. At MNCM, over four consecutive years (2005-2008), monthly visits range from lows of 3,000 to a spike of 30,000 after the Governor of Minnesota praised the website in a public address. MNCM has seen gradually increasing website traffic with each subsequent year.

MNCM used focus groups to evaluate the relevance of their reports to individual citizens. The participants found that measures reported at the level of large, corporate medical groups were not relevant to their experience of care. They preferred measure results at the individual physician level, but also found useful results at the level of the clinic site or location where they received care.

Consumers may have difficulty synthesizing multiple, technically detailed indicators regarding a single disease process. Disease-specific composite measures are perceived by patients as more relevant to their needs. Pilot sites also felt that further study into the context and presentation preferred by patients would improve the relevance of quality data to patients. Much of their current reporting and websites are targeted at medical professionals, and give specific measure results with little context.

In addition to broader disease-specific measures, other aspects of care such as access and patient experience rankings are interesting and relevant to patients. Presentation of a range of quality measures to patients should also include information regarding the context and relevancy of the results. For all these reasons, reporting of quality measures to patients and other consumers should be separate from reporting to providers, and target consumer interests and needs.

Reporting of quality measures to patients and other consumers should be separate from reporting to providers, and target consumer interests and needs

All sites felt that annual reporting current to within the most recent 12 to 18 months was acceptable for consumer use, and found no advantage to more frequent reporting. In several markets large employers have sent the publicly-available quality data to their employees at the time of annual plan enlistment. Doing so encourages patients to use the information when selecting a provider.

While health plans generally have their own pay-for-performance initiatives, some are beginning to align across plans using regional collaboratives. In Minnesota, BlueCross/BlueShield uses two of MNCM's measures for P4P, and recently state legislation has been passed to further align P4P programs using MNCM measure results.

Summary of the Value of Public Reporting

All the pilot sites hope to make their data and reporting style more relevant to individual beneficiaries. Despite low patient use, however, the sites have found many advantages to public reporting, such as maintaining trust, stimulating improvement, and empowering beneficiaries to make informed choices. Good medical care requires a trusting relationship between providers and patients, and lack of transparency invariably reinforces suspicion and doubt.

The six pilot sites also found that public reporting increased provider improvement efforts out of proportion to the actual number of individuals demonstrating interest in the data. Public reporting motivates performance improvement among individual providers; it also prompts governing bodies to increase improvement resources to their providers.

None of the pilot sites felt that their providers increased significant market share, or lost business, due to individual patients changing providers based on the publicly reported quality measure results

As a result of the BQI Project, four of the six pilot sites (IHIE, MHQP, MNCM, and WCHQ) successfully reported group-level quality measures publicly. These results, calculated from datasets that combined Medicare FFS data with their regional data sources, were posted on www.cms.hhs.gov/bqi. The four pilot sites also provided links to this data on their websites. The remaining two sites, CHIR and CCHRI, calculated quality measure at the level of the individual physician. They deferred public reporting while engaging the providers in their markets with reviewing and validating results.

None of the BQI pilot sites felt that their providers increased significant market share, or lost significant business, due to individual patients changing providers based on the publicly reported quality measure results.

Unintended Adverse Consequences of Public Reporting

While there is concern that public reporting could cause unintended adverse consequences, such as limiting access if physicians avoided treating high-risk patients, no pilot sites reported experiencing such behavior. The public reporting of all the BQI pilot sites has been done only for groups voluntarily participating with the sites. It also included the involvement and awareness of the physician groups prior to public reporting.

Summary of Reporting Preferences

The BQI Project provided many lessons learned regarding reporting quality measures to every type of data customer and for every potential use. Issues studied for each customer and use included: the preferred frequency of reporting; the preferred level of detail of the reported measures; and whether reporting at the individual physician level or group level was preferred.

The pilot sites preferred annual reporting of measures, which was sufficient for almost all customers and uses. For Cycle 3, four of the six pilots performed a semi-annual calculation of measures and found that the work burden was severe, while little additional information was derived. The only time frequent data reporting was found useful was when measures that can be measured quickly and that have the potential to make rapid improvement were used.

Such measures may be calculated quarterly or even monthly for use in rapid-cycle improvement. However, in order to be useful for process improvement, they must also be very current information. Because of the processing delay inherent in creating, aggregating, and validating administrative data, rapid-cycle reporting is best done from data already available internally to and collected by the providers.

In order to prevent caregiver resentment of overly simplified composites, the process for developing composite measures must be transparent and involve the providers.

As illustrated in Table 5-2, physicians, other direct care providers, and managers all preferred measure data at the most detailed and specific level. Although this level of detail is best for analyzing issues and designing corrective actions, most higher-level audiences find extremely specific measures overly technical, limited in scope, and difficult to understand. Measures which purport to represent an assessment of performance on some diagnosis or disease are more easily accepted by everyone above the direct caregiver level.

In order to prevent caregiver resentment of overly simplified composites, the process for developing composite measures must be transparent and involve the providers. Also, care must be taken to ensure that aggregating measures into a composite is statistically valid, accurate, and reliable.

Table 5-2. Preferred Level and Frequency of Reporting by Customer

Customer	Use		
	Preferred Frequency of Reporting	Preferred Level of Detail/ Specificity of Measure	Preferred Level of Reporting (Individual vs Group Level)
PROVIDERS			
Individual physicians and other direct caregivers	Annual for most uses, more frequent for some	Most granular detail	Individual for self-use, Group for public reporting
Managers of doctors and direct caregivers	Annual for most uses, more frequent for some	Most granular detail	Individual for self-use, group for public reporting
Health institutional governance, resource support	Annual	Composite, relevant to a recognized diagnosis	Individual for self-use, group for public reporting
CONSUMERS			
Individual patients	Annual	Composite, relevant to a recognized diagnosis	Individual if valid and relevant, clinic site is accepted
Organizations, purchasers, public media	Annual	Composite, relevant to a recognized diagnosis	Individual if valid and relevant, clinic site is better than none
Payers	Annual	Composite, relevant to a recognized diagnosis	Individual if valid and relevant, clinic site is acceptable

Data Feedback to Providers—Data Sharing Issues

An integral part of the BQI Project has been sharing data aggregation and measure calculation results with providers to validate attribution methodologies and measure results. Although time and resource-intensive, validation exercises are critical to ensuring credible results.

In order to share these results, however, Delmarva, CMS, and the pilot sites worked through laws and regulations governing how the pilots handled individually identifiable patient and practitioner information. This included both Medicare and non-Medicare data and any other data or information collected, acquired, or generated for the purposes of the project.

One set of laws and regulations in particular applies specifically to Delmarva as a QIO. For the BQI Project, since it was a QIO special study, Delmarva and our subcontractors (the pilots) were subject to the QIO confidentiality rules.[1] As business associates of HIPAA-Covered Entities or subcontractors of HIPAA business associates, Delmarva and the pilots were contractually required to comply with the HIPAA Privacy Rule.[2]

As described below, QIO and HIPAA laws and regulations generally operate simultaneously and independently of each other. In any instance in which the two laws impose different requirements, the QIO must adhere to the law that provides greater privacy protections (i.e., the one that is "more restrictive"). In general, QIO confidentiality provisions are more restrictive than the HIPAA provisions.

1 QIOs must collect, protect and disclose or release data only as authorized or required under the QIO statute in the Social Security Act (Sections 1154 and 1160), the QIO regulations in the Code of Federal Regulations (42 CFR Parts 476 and 480), and under the specific terms of their contracts.

2 45 CFR Parts 160 and 164 (Subparts A and E).

QIO Law and Regulations. Under their contracts with CMS, QIOs and their subcontractors obtain and analyze data for CMS on a routine basis. Any information that explicitly or implicitly identifies an individual patient or practitioner is considered "confidential" under the QIO regulations. This information is subject to strict confidentiality and disclosure requirements under the QIO law and regulations.

HIPAA Privacy Rule. Among other things, the HIPAA Privacy Rule provides certain patient privacy protections by controlling the use and disclosure of protected health information (PHI) by "covered entities" (and their business associates). PHI is a term of art under HIPAA, but is not a term that is used by the QIO statute and regulations[3].

Delmarva, CMS, and the pilots worked through several data sharing scenarios to ensure that all QIO and HIPAA regulations would be met throughout this project. For aggregate, non-identified data, the QIO regulations did not pose any issues. However, when considering the distribution of physician, physician group, and patient-identifiable data, the regulations affected our ability to thoroughly validate the measure results.

As already mentioned, any information that explicitly or implicitly identifies an individual patient or practitioner is considered "confidential" under the QIO regulations. In the case of individual patient-level information, this regulation prohibits Delmarva or the BQI pilot sites from disclosing patient information to anyone *other* than the practitioner who provided (i.e., billed for) the service. This restriction meant that we could not provide comprehensive patient-level measure listings.

For disease-specific measures, such as diabetes measures, we could not provide physicians with a list of his/her attributed diabetic patients if that physician had not diagnosed the patients. For all measures, we could not provide measure listings that included information about whether the patient received the service being measured (numerator events), or if the numerator event had been performed by another practitioner (for example, a screening mammography). While physicians understood this regulation, they expressed concerns about their ability to fully validate the measures. More importantly, this inability to share patient-specific information across all providers limited the ability to get complete information about the patient into the hands of the primary caregiver for the benefit of the patient's care.

In a similar fashion, this disclosure regulation affected our ability to provide individual physician-level reports to participating medical groups. We were not permitted to disclose individual-physician level reports to a group's representative, unless the physician had agreed to the disclosure, or the physician had designated the group's representative as an agent. Therefore, pilot sites had to go through a time-intensive process to obtain individual physician consent for disclosure, or they had to work through the legal department of groups to determine whether an agency had been established.

3 When acting as a business associate of the health care components of CMS, Delmarva is contractually required to comply with the provisions of the HIPAA Privacy Rule. These contractual obligations are carried through in Delmarva's subcontracts.

Despite these regulations, the BQI pilot sites were able to provide limited feedback reports. All conducted some level of validation with providers regarding the aggregated measure results. Details on the feedback to providers by the pilot sites can be found in Appendix 11.

6

Individual Physician-Level Measurement and Reporting

Experiences with Producing Individual Physician-Level Results

Initially, the BQI Project focused on reporting performance measures at the level of the individual physician or care provider. Its purpose was to evaluate the results of individual decisions of care—independent of medical team competency—and system support structures. Over the course of the project, however, interest in physician group-level public reporting increased. Currently, most of the pilot sites have expressed a strong preference for this level of reporting.

All the BQI pilot sites preferred to publicly report measure data at the level of physician groups

- Individual provider data is desired by patients and payers for accountability, but only if results are understandable, valid, and accurate
- As medical care is increasingly team-based, group-level reporting is more appropriate.
- Other reasons for group-level reporting include denominator size, attribution issues, data submission issues, and comfort and trust of providers.
- Barriers may be overcome in the future with identified strategies

There are several reasons for this preference. First, physician office care is often done in a group environment with individual physicians intentionally handing off specific types of care to other members of the group (care team). This indicates that the group level is a more natural and appropriate level to report on than the individual physician level. Additionally, there are concerns that small denominator sizes at the individual physician level produce performance measures that are not reliable. From a practical point of view, two of the pilot sites (Wisconsin and Minnesota) collect measures at the group level or clinic site level and cannot, therefore, drill down the measures to the individual physician level.

Most of the pilot sites have expressed a strong preference for physician group-level public reporting

Yet another concern that arises from individual physician-level reporting is requiring attribution of each patient's care to one physician. Many nurse practitioners, physician assistants, and other health care professionals make medical decisions; it is often not clear to which physician team member their decisions should be attributed for purposes of quality measurement.

Finally, physicians who are beginning to engage with quality measurement reporting are also more comfortable with the physician group level. For the regional medical collaboratives to be successful, physicians and providers must be actively and voluntarily involved. The desire here is not only to have the nominal approval of the regional providers; it is also to gain their discretionary voluntary effort, along with their active commitment to data-driven quality improvement. Therefore, collaboratives value and respect the concerns of the physicians and wish to develop individual-level public reporting when all stakeholders feel comfortable that the results reported are valid and add benefit to both the physicians and the consumers.

Figure 6-1. Pros and Cons of Individual Physician and Physician Group Reporting of Results

	PROs	CONs
Individual-Physician-Level Reporting of Quality Measure Results	• Patients and Payers who identify with a single physician find individual results most relevant for accountability and performance	• Small numbers may result in statistically unreliable data or "n/a" results • Questions of attributing patients to each provider, and allied providers to one doctor complicate validity of results • Individual doctors may avoid complex cases to improve scores, resulting in decreased access and unintended adverse consequences.
Physician-Group-Level Reporting of Results	• Medical care is provided in teams; therefore, measures should reflect team operation as a whole. • Larger numbers are more reliable. • Clinical data may be burdensome to collect at the individual physician level. • Physicians are more emotionally accepting of rankings as a team.	• Group-level data is not as emotionally satisfying to patients, media and payers as scores for an identifiable, individual person.

In response to these concerns, each pilot site at the outset of the project committed to three levels of reporting: 1) the level at which their community could calculate measures; 2) the level at which it could provide feedback about the measures to individual providers; and 3) the level at which the measures could be publicly reported (specifically, posted on the CMS website).

Two of the pilot sites, MHQP and MNCM, felt from the start of the BQI Project that, based on their current methods of collecting data (i.e., information from the health plans), that they could only commit to calculating and reporting at the physician group level. IHIE and WCHQ committed to work towards calculating measures at the individual physician level while maintaining reporting at the physician group level. The final two pilot sites (CCHRI and CHIR) made a commitment to calculate and report at the individual physician level.

At the beginning of the project, all of the pilot sites advocated adding language to their contracts stating any site committing to calculate and/or report at the individual physician level would work towards the stated goal. However, if for some reason calculation and reporting at the individual physician level was deemed not scientifically valid or reliable, or there was not buy-in from the physicians in the respective community, they would develop a plan to continue to work towards calculation and reporting at the individual level.

Table 6-1. Level of Measure Calculation and Feedback by Pilot Sites—Planned vs. Achieved

Pilot	Measure Calculation Level (Planned)	Achieved Y/N	Feedback to Providers Level (Planned)	Achieved Y/N	Public Reporting Level (Planned)	Achieved Y/N
CHIR	Individual Physician	Y	Individual Physician	Y	Individual Physician	N
CCHRI	Individual Physician	Y	Individual Physician	Y	Individual Physician	N
IHIE	Individual Physician	Y	Individual Physician	Y	Physician Group	Y
MHQP	Physician Group	Y	Physician Group	Y	Physician Group	Y
MNCM	Physician Group	Y	Physician Group	Y	Physician Group	Y
WCHQ	Individual Physician	N	Physician Group	Y	Physician Group	Y

All of the pilot sites that received data at the individual physician level were able to calculate measures for individual doctors. Their uniform experience was that the medical groups themselves found the individual-level data useful for internal self-improvement and in some cases for P4P at the group level. The medical groups dealt internally with issues of trust in the validity of data and attribution of patients to providers.

Their uniform experience was that the medical groups themselves found the individual-level data useful for internal self-improvement

Four of the six pilot sites (IHIE, MHQP, MNCM, and WCHQ) reported their results publicly on the CMS website (www.cms.hhs.gov/bqi). No pilot site wished to publicly report individual-level data for the reasons already stated. CHIR and CCHRI created files of physician-level reports without patient-identifiable data that would be suitable for public release. However, they are still in the process of reporting the data to their physicians and working with providers to understand, validate, and accept the results. Consequently, CHIR and CCHRI did not post publicly post results on the CMS website.

All pilot sites agreed there was a strong strategic risk to publicly releasing data that were still being validated and/or challenged by the providers. Disruption of necessary working relationships within the collaborative could be significant and long-lasting. Public perception that the data is inaccurate and controversial could damage local trust within the community for future reports.

Barriers to Individual-Level Reporting and Mitigation Strategies

Participants in the BQI Project identified mitigation strategies in response to four major barriers to reporting quality measures at the individual-physician level:

1. Medical care is provided in teams, therefore reporting individual-level results may be inappropriate or undermine team function.

   Mitigation Strategy: More analysis is needed for measures that may reflect team operations and systems support. Other measures that may more appropriately reflect individual decision-making and effectiveness also require further analysis. Different measures may have differing levels of most appropriate reporting.

2. Small numbers of cases may make individual physician-level results invalid or untrustworthy.

   Mitigation Strategy: Individual physicians may not have small numbers for certain measures that are broadly applicable. As electronic databases (such as labs, pharmacy and x-ray) become increasingly accessible, broader measures requiring clinical data may also become more available for calculation. The same applies as more clinical data are obtainable through increased use of electronic health records. Another strategy for stable measures is to extend the time across multiple years in order to increase numbers. Lastly, selecting measures with a 40 to 60 percent success rate—neither very high nor very low—necessitates a smaller population size for significant results.

3. Identifying responsibility and "ownership" of a particular patient's outcome to a single primary care provider, specialist, or both is problematic in real life. Further uncertainty is added by any computerized algorithms used to attribute patients to individual doctors.

   Mitigation Strategy: Advancing the concept of the "Medical Home," in which patients and providers share a mutual understanding of identified partnership, effectively resolves many issues related to patient-to-physician attribution. Another strategy involves notifying physicians with online rosters listing patients for whose care they are responsible. This technique, which permits physicians to correct, add, delete, and reconcile an agreed-upon roster, was used successfully in one pilot site to reconcile physician-to-physician group issues.

4. Voluntary, active physician participation is necessary to successful regional collaborative efforts, but physicians are uncomfortable with individually-reported measures.

   Mitigation Strategy: Experience suggests that providers may accept and even value reporting certain appropriate measures at an individual level if doctors are involved in a transparent process and can validate results.

An Example of Successful Individual Physician-Level Reporting Through Patient Experience of Care Measures

Through interviewing regional collaborative stakeholders, we identified a successful example of individual-level reporting derived from patients' experience of care (EOC), rather than from quality measures. In the following comparative analysis, we present how EOC measures overcame or addressed the barriers discussed above.

Team versus Individual Measures: The measurement tool for EOC measures makes a clear distinction between experience with a team or clinic site versus experience with an individual provider.

Small Numbers May Make Results Invalid: Extensive research has documented the validity and reliability of the EOC technique. There are accepted techniques that define the number and variation of responses needed for reliable and valid results. Only results that meet this threshold are reported.

Attribution of the Result to an Identified Provider: The EOC technique clearly and unambiguously identifies providers evaluated for the measure.

Provider Trust and Comfort with Reporting: Transparency and engaging providers in the process are important aspects of developing and reporting EOC measures. Consequently, these measures have been successfully deployed in several markets without adverse relationship consequences. Providers report that the successful experience has opened opportunities for more collaboration.

7

Project Evaluation and Conclusions

Pilot Sites' Assessment of the Effect of the BQI Project on Their Communities

At the conclusion of the BQI Project the pilot sites were asked to assess the impact of the experience upon their communities. Their perspectives are reported here in the context of each site's unique development.

California Cooperative Health Care Reporting Initiative (CCHRI)

CCHRI was established in 1993, the earliest of the six BQI pilot sites. The initial core group of champions included the Pacific Business Group on Health (PBGH), most of the largest health plans in the state, and representatives of the California Medical Society and the California Group Practice Association. Their goal was to collect and report standardized, reliable health plan and provider performance data, and to create efficiencies in data collection to reduce workload and cost. In 1997, CCHRI assumed responsibility to report publicly the results of the patients' experience of care survey. For ten years, it has continued to report these medical group-level survey results.

It is significant that in California there is a separate initiative called the Integrated Healthcare Association (IHA). Coordinated by the state of California, the IHA has collected quality performance measures submitted by MC plans and reported them publicly at the medical group level for eight years. To "fill the gap" not being met by group-level reports during the BQI Project, CCHRI calculated quality measures at the individual physician level from administrative data.

CCHRI receives its data from several commercial MC plans and, for the first time through the BQI Project, for Medicare FFS beneficiaries. **The BQI Project has enhanced relationships in California for data pooling and integrated performance measurement. CCHRI now recognizes that Medicare data is essential for reliably assessing physician performance. This is because Medicare beneficiaries are a large part of the patient population and use a large per-capita share of medical resources.**

California's providers, who have expressed concern with the validity of any new source of quality measures, want to validate both patient attribution and measure results. There is also interest in the CCHRI market to broaden quality measures and to include efficiency and patient experience of care measures. CCHRI feels that immediate challenges ahead for them include merging non-uniform data sets, dealing with issues of incompleteness in data sets, and mapping physicians to practice sites.

Massachusetts Health Quality Partners (MHQP)

MHQP, one of the earliest of the six BQI pilot sites, was established in 1995-96. The initial core group of champions included leaders of several large hospital/health systems, including physician executives, the Massachusetts Hospital Association, the Massachusetts Medical Society, and several leading health plans. Their goal was to provide public reporting of accurate and meaningful quality data. They initially focused on hospital-care performance, and for the first several years public reporting of quality measures was for hospitals only. As they gained experience and grew their network of stakeholders, MHQP's interest in reporting ambulatory care measures increased. By building trust across provider groups and working collaboratively with them, they successfully began reporting quality measures back to physicians for quality improvement in 2002. They began to publicly report physician group-level quality measures in 2005.

MHQP receives its regional information largely from health plans' administrative claims data. Prior to the BQI Project, MHQP received numerators and denominators for measures at the physician group level from various plans. **Beginning with the BQI Project MHQP used its established reputation for trust and confidentiality within their market to change the data use agreements with commercial plans. Doing so allowed MHQP to receive data at the individual physician level, though not patient-identifiable data.**

Patients are identified only by plan-specific ID numbers. If providers wish to get details regarding which specific patients have not received an indicated treatment, for example, the provider must query the responsible health plan directly as MHQP cannot provide that information. However, receiving data at the level of the physician allows MHQP to roll-up data into medical groups accurately using a single, standard directory of medical group membership. MHQP has carefully developed all its methodologies in a transparent manner that involves all stakeholders (including physicians) at every stage.

MHQP has aggregated physician-level data for primary care physicians across regional commercial health-plan data since 2003. Because MHQP reports at the physician group level, MHQP has deep experience with assigning physicians to groups and validating the information with providers. After finding significant disparities among the group assignments created from administrative data by the various health plans, MHQP created a web-based Master Physician Directory that providers can update, correct, and validate directly. MHQP feels that this method has generated accurate and valid physician group data, and has built relationships and trust directly with provider groups.

The BQI Project has expanded MHQP's expertise in aggregating claims data and increased their awareness and understanding of attribution issues and methodologies. They have found that sharing the quality and patient experience information back to providers is effectively improving ambulatory care in their community. The measures calculated from Medicare FFS data for the BQI Project, however, had limited impact in the Massachusetts medical community. This was due, in part, to numerous quality initiatives competing for attention within the region. More importantly, BQI data use restrictions limited the ability of MHQP to give feedback to physicians about care at the individual patient level. Lack of patient-level information made it difficult for providers to validate the results and limited their ability to determine which patients needed further care and services.

Minnesota Community Measurement (MNCM)

MNCM began in 2002 as a collaboration of all seven of Minnesota's non-profit health insurance plans. The regional collaborative rapidly partnered with medical groups and other stakeholders, and in 2005 the Minnesota Medical Association joined the effort, forming a new non-profit community-based organization. With the aggregated MC dataset, MNCM reports performance results on 90 primary care groups, 20 obstetrics/gynecology groups, 13 urgent care/convenience care clinics, three cardiology and two endocrinology groups. A total of 128 groups, therefore, were reported in the 2007 Health Care Quality Report. These medical groups represent over 700 clinics, providing primary care to over 90 percent of Minnesota's citizens. MNCM's 16-member Board of Directors is comprised of medical directors from the health plans, hospital representatives, physicians, employers, business groups, and consumer organizations.

MNCM's MC data is patient-level data submitted by the health plans. The measures use administrative data and also in some cases information extracted from medical records. Additionally, the MNCM dataset includes commercial fully- and self-insured MC products; Medicaid; General Assistance Medical Care; MinnesotaCare and Minnesota Senior Health Options; and Medicare MC. The BQI Project is the first time MNCM has had access to Medicare FFS data.

MNCM has demonstrated that reporting quality measures to providers and the public improves provider performance within their community. A composite measure for diabetes care, for example, has improved every year since 2004. **The BQI Project has allowed MNCM for the first time to compare care for Medicare beneficiaries to the commercially-insured population. The BQI Project has also allowed MNCM to begin a new, parallel method of quality measure collection with results directly submitted by the providers. This method allows richer, outcome-related measures to be captured, but has also introduced other requirements for data auditing and validation.**

Wisconsin Collaborative for Healthcare Quality (WCHQ)

WCHQ was started in 2002 by several health delivery systems, each with a large multi-specialty group and a tertiary hospital. Encompassing five geographically distinct markets, WCHQ now has more than 40 physician groups, hospitals, and health plans. It also includes several activated employers, the Wisconsin QIO MetaStar, the Wisconsin Hospital Association, and Wisconsin Medicaid.

Since its inception, WCHQ has focused on measuring and reporting defined quality metrics to the public as well as to providers. WCHQ is unique among the pilot sites in that it uses no administrative billing data, but rather has measure results submitted at the medical group level by the providers themselves. The reported measures are across not only "all payers" but "all patients" as well, including self-pay and charity patients. WCHQ does not receive any patient-specific or individual-physician level data, but only group-level numerators and denominators for each measure. The collaborative has put in place auditing and training functions to assure uniformity and accuracy of the provider-submitted measures.

Receiving Medicare FFS data through the BQI Project has allowed WCHQ and its participants to compare their all-patient measure results with the results for Medicare beneficiaries as a sub-population. Recently, WCHQ requested their provider-submitted data be stratified by payer category, including commercial and Medicare FFS, to gain insights into measure variation among payer categories. This has allowed providers to identify disparities in their quality measure results, and to begin studies and corrective action to improve them.

Indiana Health Information Exchange (IHIE)

IHIE is a community-based collaboration that builds on its long-established local experience in healthcare information technology (HIT). In the 1960s, local industrial philanthropist Sam Regenstrief established an institute for researching and developing improved HIT at the Indiana University's Medical Center. Voluntary clinical data-sharing and authorized research has been performed on this enormous base of clinical data for over 40 years. The Indiana Network for Patient Care (INPC) was created to improve sharing clinical information among hospitals, emergency rooms (ERs), and ambulatory sites. IHIE builds upon this base. When a patient now arrives in an ER in Indianapolis or its surrounding counties, for example, records including health history, meds, allergies, electrocardiograms (EKGs), and X-Rays are available to any ER doctor from any regional hospital.

Indiana Health Information Exchange

Building on this deep experience, it was natural for the Indiana community to establish the Quality Health First program in 2005 to begin outpatient quality measure calculation and a pay-for-performance program. Consequently, IHIE has only begun calculating and providing feedback for quality measures with the BQI Project. **The project created an opportunity for IHIE to develop a new, more complete source of administrative data for Medicare beneficiaries. The collaborative found that data should be as current as possible to be useful for process improvement by providers. They also found that data should be shared back to providers at the level of identifiable patients. Doing so validates the correct attribution of patients to providers. More importantly, it identifies patients who need interventions in care. All stakeholders in IHIE felt that the inclusion of Medicare FFS data gave a more complete and useful picture of care in their community.**

Center for Health Information and Research (CHIR)

CHIR has its origins in the Arizona HealthQuery (AZHQ) database. AZHQ began in 1999 through a grant from the charitable Flinn Foundation, targeting the number of uninsured children in the community of Yuma, Arizona near the Mexican border. Medical providers in the Yuma area were persuaded to voluntarily share their billing and coding data, with patient-specific information, and the AZHQ was created. This database is patient-centric, meaning that any patient, regardless of insurance status, could be tracked for any encounter with any medical provider or physician, even if they never paid for services. AZHQ has now grown to include the 60 percent of Arizona's population in Maricopa County surrounding Phoenix, as well as Medicaid and some state-wide datasets.

CHIR

Until the BQI Project, CHIR's primary focus was not on calculating ambulatory quality measures, providing feedback to physicians, or public reporting. Prior to 2005 their main focus

was public health research. CHIR has begun quality measure calculation and is progressing with validation of their results by physicians using the BQI Project. The region's medical market has only a few large medical groups, so most of the quality measure data calculated by CHIR related to solo- or small-group providers.

The BQI Project has created a dialogue among regional providers regarding the benefit of using quality measure data. CHIR found that physicians have a genuine interest and demand for information evaluating performance if they can be convinced that the results are valid and from a trusted source. CHIR feels that loss of trust or even the perception of invalid data would not only result in physicians discounting measure results, but would make obtaining data from providers impossible and undermine all community collaboration. CHIR feels the BQI Project provides the foundation for the developing reporting systems that meet their community's needs.

Conclusions from the BQI Project

Effective Methods of Aggregating Medicare Data and Other Regional Data

The BQI project's main objective was to provide recommendations for the most effective methods for aggregating Medicare claims data with other payer data. Such methods would produce more comprehensive measures to evaluate the quality of services Medicare beneficiaries received.

The BQI Project has demonstrated that, in all of six regional pilot sites, the inclusion of Medicare FFS data gave a more complete picture of care quality. More strongly, the pilot sites felt that having Medicare FFS data was essential to a complete picture of community care. Medicare beneficiaries represent a major population segment in all of the pilot markets, and senior citizens are frequent users of the healthcare system. Consequently, any picture of community care that does not include Medicare beneficiary information is significantly deficient.

The BQI Project has demonstrated that the inclusion of Medicare FFS data gave a more complete picture of care quality. More strongly, the pilot sites felt that having Medicare FFS data was essential to a complete picture of community care

The BQI Project has established that the technical issues surrounding aggregation of multiple data sets can be overcome. However, the lack of uniform data standards means doing so requires constant attention to detail and significant expertise. Uniform identifiers for medical providers and organizations (such as the National Provider Identifier) may make data aggregation less problematic in the future.

Datasets are best aggregated at least at the level of the individual physician or other provider, so that issues with medical group definition are minimized. If measure reporting is needed at the medical group level, consistent group membership rosters may be used to combine individual doctor measures.

Obtaining and combining data at the level of individual patients is best for purposes of data validation and for provider process improvement. Most importantly, data identifiable to the individual patient is necessary to recognize patients who need medical interventions or improved care. The BQI Project has demonstrated that it is possible to aggregate data across multiple datasets at the individual patient level while including appropriate protections of confidentiality for Protected Health Information as required by HIPAA. A patient-centric record could be more easily achieved if a uniform patient identifier was standardized across all medical datasets. In

addition, for other QIO projects, a patient-centric record could be more easily achieved if QIO regulations regarding QIO use of Medicare data were harmonized with HIPAA regulations.

Providing Information to Medicare Providers for Quality Improvement

The second major goal of the BQI project was to provide information to physicians and medical groups who treat Medicare beneficiaries for purposes of quality improvement.

The BQI Project demonstrated that it is possible to aggregate data across multiple datasets at the individual patient level while including appropriate protections of confidentiality for Protected Health Information

The regional medical collaboratives participating in the BQI Project have demonstrated that engaged providers do improve their performance by responding to feedback reporting of measurement data. Regions that have been reporting for several years have shown steady incremental increase in performance measure results. Once the providers are activated for process improvement and have validated the measurement process, their improvements appear to be accelerated by both public reporting and pay-for-performance incentives. These strategies also stimulate growth in the number of participating and engaged medical providers.

The BQI Project showed that performance measure results for Medicare FFS beneficiaries are in many cases not the same as results for patient populations with other insurers. Giving providers complete information about all their patient groups allows disparities to be addressed. Providers find that comprehensive data across all payer and patient groups is most reflective of their clinical practice, and most relevant to their perception of their own quality of care.

The BQI Project has shown that providers want detailed information about their performance measurement results as well as the most current and frequent reporting of measurement results. Current and detailed information is most useful for analysis of issues and for process improvement. Providers want individual patient-level data to validate their responsibility for the patient's care, and most importantly, to identify patients who may benefit from medical intervention.

An important lesson captured is that administrative data from payers improves providers' processes of care

An important lesson captured from the experience of the BQI pilot sites is that administrative data from payers improves providers' processes of care. It also can provide actionable information regarding deficiencies in the care of individual patients. If information "silos" can be broken down and comprehensive patient data put into the hands of providers, interventions can be taken to correct deficiencies, improve patients' care, and prevent harm.

Provide Medicare Beneficiaries with Physician Performance Information

Most BQI pilot sites agreed that patients find performance measurement results for large corporate medical groups of limited relevance and interest. Patients want to know how their own individual physician compares to others. If that level of detail is unavailable, patients want comparisons at the level of the clinic or practice site where they receive care.

The BQI Project participants concluded that patients generally prefer higher-level, understandable, and relevant information on medical performance measures. These measures should represent care for a condition or group, such as disease-specific measures for overall diabetes care or heart failure care. Patients are also interested in measures beyond clinical quality including patient experience

of care and access to care. Including quality measures as part of a balanced report with other patient-centric measures may make quality measures more useful to patients.

Lastly, several pilot sites suggested further research into the most effective techniques for presenting quality measures to patients in an appropriate context. The pilots' experiences with their own reporting, in addition to reporting of quality measures for the hospital industry and for health plans, suggest that reports to consumers should include explanatory notes and a contextual framework to allow consumers to decide how best to use the information for their own care.

Pilot sites that publicly report results agreed that public reporting was effective in catalyzing or further motivating providers in their improvement efforts. It also demonstrated transparency and partnership within their communities, maintaining the essential trusting relationship between medical caregivers and patients.

Conclusions Regarding Performance Measure Selection and Calculation

The BQI Project showed that administrative claims data, other electronic datasets such as laboratory results and pharmacy data, and information from clinical records are all useful and important in calculating performance measurement results

The BQI Project showed that administrative claims data, other electronic datasets such as laboratory results and pharmacy data, and information from clinical records are all useful and important in calculating performance measurement results. Some performance information is most reliably found in payment claims data. Whether a patient has had a mammogram performed within the last year, for example, may or may not be captured in any of several providers' records. However, it will be reliably included in her insurer's database if the test was charged and paid for. Other important quality information is only recorded in the clinical record (for example, a blood pressure measurement or a patient's improvement in pain).

Obtaining data from administrative sources is reliable and resistant to biases and human error. As more and larger administrative datasets become available for aggregation, the power of mining administrative data will only increase. However, the number and types of measures that can be calculated from claims data alone are very limited.

Submission of performance data by providers, based on their clinical records as well as their own administrative data, allows more measures with more outcome relevance to be calculated. Provider-submission of data also avoids many of the problems with attribution of patients to doctors and of doctors to medical groups. Assuring the accuracy of provider-submitted data requires standardized training of staff and an on-going auditing process. Electronic medical record systems may make clinical records more easily accessible for performance measurement in the future.

Implications of the BQI Project

In a short period of time the BQI Project has distilled many valuable insights into measure selection, combining large datasets, and performance measure reporting.

Of particular interest to CVEs in the Value-Driven Health Care Initiative are BQI insights into developing methods for aggregating data from multiple sources. Newly forming CVEs may also find the information in this report enhanced by the "Regional Coalition Collaboration Guide"

available at the AHRQ website (http://www.ahrq.gov/qual/value/collabguide.htm). This guide, authored by Delmarva in cooperation with the BQI sites, offers key lessons derived from the sites' success in building and managing regional coalitions.

On a larger level, the BQI Project has been the only initiative to combine multiple differing "raw" datasets to derive performance measures for ambulatory care. The project created an opportunity for three of the regional collaboratives (MHQP, MNCM, and WCHQ) to study Medicare FFS data at the individual patient level. They were able to combine it with data from other payers to make more complete the spectrum of care quality measurement in their communities. For the other three pilot sites (CHIR, CCHRI, and IHIE) the project has served as an innovative catalyst for new and larger data aggregation and measurement initiatives in their communities.

The BQI Project has been the only initiative to combine multiple differing "raw" datasets to derive performance measures for ambulatory care

Another important implication of the BQI Project is that the sites demonstrated the efficacy of using public-private regional coalitions to improve ambulatory medical care. The lessons learned in this regard will be applicable to other regional medical collaboratives committed to community improvement, including CVEs and the Robert Wood Johnson Foundation's Aligning Forces for Quality.

Finally, the pilot sites' experience has illuminated many of the questions and issues that the AQA identified in 2005. The responses to these questions from the BQI experience are reported in Appendix 1. As is always the case, each answer and every insight raises other issues that need further study. Opportunities for further research have also been included in Appendix 1.

The BQI Project has served as a platform for informing the creation of a sustainable and transparent system of physician performance reporting. The BQI Project has emphasized public-private collaboration at the local level. Beginning initially with a limited number of measures vetted through the NQF and AQA, the BQI pilot sites will continue to expand the scope and relevancy of their performance measures reporting .The BQI Project has demonstrated that physician performance measurement and reporting can be enhanced through the involvement of local collaboratives.

Acknowledgements

The Better Quality Information to Improve Care for Medicare Beneficiaries (BQI) Final Report is the product of collaborative work from Delmarva Foundation for Medical Care, the six BQI Pilot Sites, and Technical Subcontractors.

While this report presents the analysis, findings, and conclusions that emerged from the deliberations of the project participants and staff research conducted for this project, it is not a consensus document. However, we have done our best to present an accurate and faithful reflection of the thinking of the group as a whole. This report would not have been possible without the benefit of their insights and expertise.

Delmarva would like to specifically thank the following:

Delmarva Foundation for Medical Care

Dejene Ayele, MPH
Yan Bai
Don Broach, MS
Robert Brooks, MD, PhD
Tennille Daniels, MPH
Mariana Lesher, MS
Aziza Hull
Janel Handy
Dwayne Eutsey, MA
David Morrell, MS
Craig Weller

California Cooperative Healthcare Reporting Initiative

Cheryl Damberg, PhD
Julie France, BSN
Ann Marie Guisto, BSN
David Hopkins, PhD
Cathie Markow, BSN, MBA

Center for Health Information and Research: Arizona State University

Wade Bannister, PhD
William Johnson, PhD
Kathleen Russell, BS

Indiana Health Information Exchange

Jill LeMasters, RN
Chris Schultz, MBA
Jan Van Vlack, RN

Massachusetts Health Quality Partners

Polly Marvin, MPH
Barbra Rabson, MPH
Janice Singer, MA, MPH

Minnesota Community Measurement

James Chase
Diane Mayberry, MHA, RN, CPHQ
Anne Snowden, MPH, CPHQ

Wisconsin Collaborative for Healthcare Quality

Robert Stone Newsom, PhD
Chris Queram
Amy Topel
Kathy Werner, MS, RN

Technical Subcontractors

Center for Health Systems Research & Analysis
Halcyon Research
National Committee for Quality Assurance
Thomson Reuters
ViPs

Appendix 1

AQA Questions and Responses

AQA Questions: BQI Project Responses and Recommendations		
	BQI Responses	**BQI Recommendations**
Measurement		
1. Are the measures used in the pilots valid (i.e., precise and reliable) ?	The BQI pilot sites made decisions to allow precise and reliable measure calculation. They found that some of the nationally-recognized measure specifications were clearly applicable to some data sources (for example, commercial managed-care administrative data sets) but more ambiguous when applied to other data sources such as Medicare FFS data or clinical records. Therefore decisions regarding specifications for these sources had to be made locally, resulting in differing specifications which made comparison of results across regions difficult.	For each performance measure, specifications must be clearly defined and applicable across data sources.
2. Does the collection, aggregation, and reporting of the particular measures used in the pilots lead to specified goals and desired outcomes?	Regional collaboratives in the BQI project that have been reporting measures to providers as well as to the public for several years have seen steady improvement in the reported measures.	Measures that are more relevant to providers and patients, and measures that reflect more aspects of medicine than preventative and chronic care, must now be considered.
3. Which measures apply to primary versus specialty care? How do methodological rules differ based on this applicability?	For the BQI project measures were selected from national sources such as the NQF and the AMA/PCPI. Only primary care and those relevant specialties identified by the measure developers were included in measure calculation. **The analysis plan of the BQI final report describes the impact on measure results of decisions to attribute measure outcomes to one or multiple providers, and describes the options for doing so, including the possibility of assigning partial or weighted responsibility**	Measure developers should clearly specify the types of medical providers involved in the patient's care to whom partial or full responsibility for the outcome should be assigned.
Data Collection, Aggregation and Validation		
1. What is the most effective method for collecting and linking data from different sources (e.g., administrative data, including claims, lab, and pharmacy data; medical records data; survey data)?	The BQI project showed that the best way to combine and link data from different sources is at the level of identified individual patients. This method is most accurate and also allows information to be used by providers for improving the care of patients and preventing harm. If patient identifiers are not available to allow merging of patient-level data across data sets, then it is best to assign patients to identified individual physicians and merge the sets at this level. Merging data sets at the higher level of medical groups introduces complex errors due to non-standard assignments of physicians to groups.	**Consistent use of the National Provider Identifier must be mandated. Providers must be educated on the importance of verifying identifiers on a regular basis. A single uniform patient identifier should be developed to facilitate combining of data sets in a patient-centric manner. By moving toward a unified EHR for patients, this would speed quality improvement and improve medical care quality and safety.**
2. How should electronic and non-electronic sources be linked and how should public and private sector data be linked?	The BQI Project included pilot sites that primarily used electronic administrative data, as well as, pilot sites that collected performance measure directly from providers, including paper medical records. **Non-electronic sources must be converted into an electronic data set usually by manual data entry. Combining of differing sources may be accomplished regardless of the source as described in the previous response.**	Support for electronic health records implementation and continued development must be provided.
3. What were the potential methods you considered for collecting your measures? What were the barriers and solutions? What was the rationale for selecting the methods you chose? What was your experience using these methods?	The six BQI pilot sites used a variety of data sources available within their communities, and also included Medicare FFS data. Most of the sites included administrative data from commercial payers, and many included clinical data reported from physicians, hospitals and other providers. Most sources were electronic but some were from paper records. Some sites had experience with laboratory results and pharmacy/Medicare Part D data. Generally their choices were guided by what was available and practical. These issues are discussed in detail in the BQI Final Report.	A national consensus on measure specifications that are specific to various data sources is needed to prevent regional variation that impeded comparisons. A standard provider identifier and standard data element field definitions should be developed. **A unified patient identifier should be developed as recommended above.**

AQA Questions: BQI Project Responses and Recommendations		
	BQI Responses	**BQI Recommendations**
4. How did the pilot assess the "cost-benefit" trade-offs when exploring and linking multiple data sources and types? 5. How did the pilot address design issues to ensure that data collection was minimally burdensome and more cost effective (include consideration of whether/how to conduct administrative data collection, chart reviews, medical record abstraction, and patient survey efforts). How did the pilot balance burden of data collection with the need to assure valid, reliable information?	Some of the BQI sites (WCHQ, CHIR, and MNCM) used provider submitted data during the course of the project while others primarily used available electronic administrative data primarily. The BQI project did not directly address the cost-benefit tradeoffs, though the pilot sites do discuss the advantages and barriers of their methods of data collection.	Further research is needed into the cost-effectiveness of administrative, augmented-administrative, and clinical sources for measure result calculation.
6. What is the relationship between the pilots and existing clinical data exchange efforts?	Two of the pilot sites, IHIE and CHIR, are based on existing regional integrated clinical data repositories as described in the BQI final report. **The BQI project has shown that clinical data increases the number and relevancy of performance measures that can be calculated and reported, while administrative claims data may be a valuable augmentation of clinical data that identifies patients requiring medical intervention for care improvement or protection from harm**	Wherever possible, regional collaboratives for performance and quality measurement must build data sharing relationships with any regional clinical data exchange efforts.
7. Are the methods and infrastructure used by the pilot replicable and scalable to regional or national levels?	Yes. The varying data aggregation methods are described in detail. Specific methods with particular advantages, such as MHQP's interactive method of defining physician groups and IHIE's comprehensive patient-centric data repository are given particular emphasis.	Further research is needed about national standards for varying data sources and the infrastructure support needed to accommodate the increasing number and size of available data sets.
8 a. **Provider IDs:** What is the most effective method for establishing a common provider ID, determining the relationship of one physician to another, and the relationship of a physician to a particular organization or site? What is the most effective method for assigning a physician to a particular specialty?	For the BQI project, Medicare and local data sets were studied for the calendar years 2005, 2006, and 2007. The NPI has not yet been generally implemented. Extensive study of the use of UPINs, medical group UPINs, and Tax ID (TINs) was performed. Assigning physicians to groups using TINs is easiest because the TIN is available in all encounter forms while group UPINS are not. All methods of assigning physicians into groups by administrative data were found to have inaccuracies of 10% or greater. Methods to allow corrections of medical group memberships by the providers themselves were found to be most effective by the BQI pilot sites.	Further research into methods to allow active interaction with medical groups will improve the accuracy and validation of results.
8 b. **Sample size:** What is an adequate sample size for validating physician performance? i. On what basis did you select the sample size? Did it vary by measure and measure type? What was the rationale? What was your experience? ii. For a given measure or measurement type, how does sample size vary according to the degree of statistical confidence required by the pilot? iii. Does the sample size vary if reporting is based on single performance measures versus a "roll-up" of a procedure- or condition-specific cluster of measures versus an overall composite?	The BQI project resulted in extensive collaborative experience on the validation of measure results, as well as further statistical research into reliability issues. The validation by providers of results may be "good enough" for internal use for quality improvement even when based on very small numbers. However, when results are used for P4P or public reporting the perceived stakes are raised, with an increased need for statistical validation. The reliability of a result increases with larger sample population size, with measure result (low-average-high) and with the variation of measures around the result. It is different for different measures. The example of patient-experience-of-care surveys is discussed to suggest that a sufficient denominator size is best calculated on a case-by-case basis.	A great deal more research needs to be done to untangle the issues surrounding reliability of measures. Reliability of results can be affected by many different factors, and a simple rule-of-thumb is not sufficient. Also, the question of whether a single provider's performance result reliably predicts their future behavior is a distinct question from whether two providers' rankings can be reliably used to distinguish their performance.

AQA Questions: BQI Project Responses and Recommendations		
	BQI Responses	**BQI Recommendations**
8 c. **Attribution:** What is the most effective and efficient model for attributing performance to specific physicians when multiple physicians care for a patient? For example, should care be attributed to the physician who most frequently sees the patient, the physician with the highest costs, or the physician with the highest utilization?	Which method of attribution is most appropriate, most relevant to providers or to patients, or most reflective of true responsibility of care is a matter of judgment that may be reasonably made differently for different measures. The implications of the choice of attribution method on the calculated measure results are investigated at length in the BQI final report.	Further research is needed on these issues.
8 d. **Composite measures:** What is the appropriate method for developing a composite measurement set for a specific condition or across multiple conditions? How should physicians be scored based on the composite measurement set (e.g., point system, highest overall percentage, other)?	The BQI Project reported the experience of some of the pilot sites with composite measures, particularly all-or-nothing disease-specific composite measures. The preference of medical consumers for such broad measures was distinguished from the desire of providers for specific, detailed information.	Further research on composite measures especially for use by consumers, including patients and payers, is needed. Context, relevant to the target audience, validity and reliability for purposes of medical decision-making should be considered.
*8 e. **Risk-adjustment:** How did you address risk adjustment? Did it vary by measure and measure type? What was the rationale? What was your experience?*	Risk adjustment of patients and measures was not addressed in the BQI Project. Several of the pilot sites' physician stakeholders expressed their perception of the importance of risk-adjustment, especially for purposes of comparison and accountability.	Further research on risk-adjustment in ambulatory care outcomes and performance measurement and reporting is needed.
8 f. **Data analysis:** What are the common standards for benchmarking, performing correlation analysis and otherwise analyzing data?	The BQI Project included an analysis of the role strong correlation of rankings across providers can play in the decision to merge datasets.	Further research regarding the utility of benchmarking and correlation analysis in data aggregation is needed.
8 g. **HIPAA and data access:** Who has access to the data and at what level (aggregate versus raw data)? How will privacy and security issues be addressed? How has the project addressed ongoing control or ownership of aggregated data? Who will "own" the aggregated data?	All six pilot sites combined data from differing sources using Data Use Agreements and Business Associate Agreements that were compliant with all HIPPA and also QIO regulations. The pilot sites were able to maintain audits of every data element with respect to source, so that data access could be monitored and data can be returned if necessary.	Administrative data should be combined at the patient or at least provider level to maximize the opportunity for engagement of providers, action for process improvement, and utility for improving patient outcomes. Relevant regulations should be harmonized with HIPAA, and appropriate MOAs and BAAs executed to provide security, access control and patient ownership of information.
9. What mechanism was used to ensure data was validated (e.g., external review, physician attestation, other)?	The two BQI pilot sites (MNCM and WCHQ) that receive provider-calculated quality measures have defined training and auditing processes to ensure accuracy. Final measure results are given as feedback to providers who may challenge apparent errors or inconsistencies. All the BQI pilot sites emphasized the importance of trust in successfully accomplishing improvement.	Regional collaboratives should engage their providers transparently in the development and calculation of measures. Mechanisms for validation of measure results should be available when needed. An auditing process is required for provider-submitted measures.
Reporting		
1. What type of information should be reported back to physicians to facilitate quality improvement? What are the best ways to report this information back to physicians to promote change in clinical practice?	The BQI project showed that caregivers prefer the most specific and detailed level of measures in order to improve their processes. For some measures annual reporting is sufficient, for rapid-cycle improvement very current data at quarterly or even monthly intervals may be needed.	Regional collaboratives should constantly solicit feedback from their providers regarding the utility of feed back reports. Close interaction stimulates process improvement and creates opportunities for further measures.
2. How should information be portrayed so that it is most useful to consumers, employers and other stakeholders?	Medical care consumers prefer broader measures of care quality, presented in context that describes the relevance of the result to their needs. Consumers are also interested in other aspects of quality including access and service. Some pilot sites found focus groups useful for insight into consumers' desires and uses for information.	Reports to consumers must be designed for consumers' needs and understanding, and should be separate from provider reports. Consumer feedback should be obtained.

AQA Questions: BQI Project Responses and Recommendations		
	BQI Responses	**BQI Recommendations**
3. What role does time play? How much time is needed to collect a valid number of cases at the individual and group levels? How long a delay will exist between collecting data and finally reporting out?	One pilot site (IHIE) that receives administrative and clinical data locally has been able to calculate measures within 30-90 days of delivery. For receiving commercial and Medicare FFS data, annual calculation of measures was found to be sufficient. Results for most measures reported 9-18 months aged are satisfactory. Results over 18 months old become increasingly irrelevant to providers and patients.	Close interaction with care providers to identify opportunities for rapid-cycle improvement that may benefit from very current and frequent data reporting is needed. We recommend further research into methods to identify such opportunities.
Evaluation		
1. What are the most effective methods to measure the correlation between performance assessment based on measures of clinical quality, cost of care, structural capacity, and patient experience? What correlations were found? (Note: understanding the impact of structural/IT capacity in clinical quality and cost of care is a critical potential learning of these pilots.)	The final scope of the BQI project did not include cost of care or patient experience. The experience of several of the pilot site collaboratives confirms the interest and value of measuring and reporting not only quality of care but also efficiency, structural capacity and access, and patient experience of care.	Further research into the important relationships between measuring and reporting not only performance on clinical quality measures, but also cost of care, structural capacity, and patient experience, is needed.
2. What are the advantages and disadvantages of relating measures of quality, cost of care, and patient experience? Do composite and discrete measures of quality, cost of care, and patient experience result in useful and meaningful data representation?		
3. What are the most effective methods to develop composite scores within and across domains and at what level (e.g., individual physician or practice site)?	Composite scores were used by some of the BQI pilots for public reporting. Composites across domains to include cost and service were not studied. The BQI final report includes a detailed analysis of the issues surrounding group vs. individual level reporting, including an illustrative example comparing care quality reporting with patient experience of care.	Further research is needed into composite measures and their use to provide meaningful and relevant information that is at the same time accurate enough to inform medical decisions.
4. What correlations were found between physician performance measured for the Medicare population and the performance for the same physician based on commercial or Medicaid populations?	In some cases a provider's performance was well-correlated across payers, but in many cases the correlation was low. Correlations tended to be higher for very high and very low performers, lower for providers of average measure result. Factors found to possibly affect correlation include provider behavior, patient demographics and behavior, and system effects such as managed care and coverage options.	Further research is needed.
5. Based on the results of the pilots, what lessons/conclusions can be applied to future policy?	The BQI Final report makes several recommendations that may be applied to future policy.	The BQI Final report makes several recommendations that may be applied to future policy.

Appendix 2

Pilot Site Descriptions

CCHRI
CHIR
IHIE
MHQP
MNCM
WCHQ

Appendix 2
Pilot Site Descriptions

California Cooperative Healthcare Reporting Initiative
San Francisco, California

Mission

The California Cooperative Healthcare Reporting Initiative was convened to help consumers and employers make informed health care purchasing decisions through its mission to collect and report comparable, reliable performance data.

Overview

Convened in 1993 by the Pacific Business Group on Health, the California Cooperative Healthcare Reporting Initiative is governed by an executive committee with equal representation from purchasers, plans, and providers that oversees all projects and determines overall policy and strategy. A reporting committee advises the executive committee on all matters of internal and public reporting, whereas various project committees ensure that overall requirements and objectives of the projects are achieved. California Cooperative Healthcare Reporting Initiative stakeholders include Pacific Business Group on Health participating employers, representing nearly 3 million California employees, retirees, and their families; the major California health plans; and provider organizations.
The unique characteristics of CCHRI as a BQI site include:

- CCHRI is a statewide multi-stakeholder collaborative consisting of purchasers, health plans, and medical group/physician providers, with each of the three stakeholder groups having equal say in decision-making. All stakeholders agree to a set of rules regarding the use of data that are reported by CCHRI. These rules are set by the CCHRI Reporting Committee.
- CCHRI has a long history of collaboration on performance measurement and public reporting extending to its founding in 1994. For more than a decade, CCHRI has been working to produce performance measures at all levels of the health system and to ensure that those results are made transparent.
- CCHRI's agenda is responsive to the needs of purchasers and consumers in the State for better quality information. Its reporting is both internal – i.e., to the stakeholders themselves – and external – i.e., to the general public. Reporting to the general public is done through the State Office of the Patient Advocate, which obtains its data from CCHRI.
- CCHRI reports clinical performance measures at the health plan and individual physician levels and patient experience data at the health plan and physician-group levels. For the past three years, CCHRI has facilitated voluntary patient experience surveying at the individual physician level for physician groups. Separate collaboratives exist in the State for hospital measurement and reporting and for rewarding physician group performance.
- Public reporting of clinical performance at the physician-group level based on HMO and POS data has been occurring since 2004 through the California Integrated Healthcare Association (IHA) Pay-for-Performance program. As of 2008, HMO enrollment represented about one-half of the commercially insured market in California. The other half is largely represented by PPO health plans. The BQI Pilot enabled CCHRI to fill some major gaps in its performance measurement and reporting by addressing physician-level performance and assessing care delivered to commercial PPO and Medicare FFS beneficiaries.

Data Experience

The California Cooperative Healthcare Reporting Initiative has more than 10 years of experience collecting and pooling performance data at the health plan and medical group levels and, more recently, at the physician level. Eight health plans representing more than 85 percent of the commercial health maintenance organization population in California participate in a variety of the cooperative's data collection projects, and many plans participate in several different projects. In 2003, the Pacific Business Group on Health started collecting individual physician-level patient experience data and reporting performance feedback results to providers from 12 groups. In 2006, the California cooperative reported results from more than 3,000 individual physicians from 27 groups.

Performance Measurement

Since 1997 the California Cooperative Healthcare Reporting Initiative has advanced physician-level performance measurement in California and is currently using a survey tool based on the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Survey tool in physician-level measurement. The cooperative's work in California has informed the CAHPS survey development process. In particular, the cooperative and Pacific Business Group on Health staff have actively worked to develop the CAHPS as part of an effort to create a nationally standardized tool for measuring patient experience at the group, physician, and practice-site levels.

Reporting and Performance Improvement

The California Cooperative Healthcare Reporting Initiative has issued annual performance feedback reports since 1994. These reports, conducted through multiple reporting vehicles, compare the performance of the participating health plans on specific measures. Physicians receive feedback reports by individual health plan members of the cooperative and their respective medical groups. The cooperative's *2005 Report on Quality* included results for each participating plan as compared to the national mean, national 75th percentile, and national 90th percentile, for the following clinical topics:

Chronic Care

Diabetes care
Asthma care
Antidepressant medication
Mental illness
High blood pressure treatment
Beta blocker treatment
Cholesterol management
Appropriate treatment for children with upper respiratory infection
Appropriate testing for children with pharyngitis
Osteoporosis

Preventive

Immunizations for children and teens
Prenatal care
Postpartum care
Breast and cervical cancer screening
Chlamydia
Colorectal cancer screening
Smoking cessation
Influenza immunizations

Impact and Accomplishments

- The California Cooperative Healthcare Reporting Initiative is the main source of data for California's Office of the Public Advocate consumer report card, the official health plan and medical group report card for the State.
- Employers, such as CalPERS and Wells Fargo, use the cooperative's data in their plan chooser tools for employees and their dependents.

- The California Cooperative Healthcare Reporting Initiative produces physician group performance data that are used by Integrated Healthcare Association pay for performance.
- The California Cooperative Healthcare Reporting Initiative spawned a multistakeholder quality improvement collaborative, the California Quality Collaborative, whose mission is to advance the quality and efficiency of patient care in California through collaboration.
- The Pacific Business Group on Health has been actively involved in the Agency for Healthcare Research and Quality-funded CAHPS development work to create a nationally standardized tool for measuring patient experience with care at the group, physician, and practice-site levels.

Center for Health Information and Research-Arizona State University Phoenix, Arizona

Mission

The Center for Health Information and Research (CHIR) provides all health care stakeholders with information that enables quality improvement activities at the policy and the practice level.

Overview

CHIR is an academic based research center that began around 1990 and focused its efforts on health policy research. One of its ongoing projects is Arizona HealthQuery (AZHQ), a massive relational database that contains health information for millions of individuals who have received health care in Arizona. AZHQ began in 1999 through a grant from the charitable Flinn Foundation, targeting the number of uninsured children in the community of Yuma, Arizona near the Mexican border. Dr. William Johnson and his staff convinced essentially all the medical providers in the Yuma area to voluntarily share their billing and coding data, with patient-specific information, and the AZHQ was created. This database was patient-centric, meaning that any patient, regardless of insurance status, could be tracked for any encounter with any medical provider or physician, even if they never paid for services.

The worthy public health goals of this database and the trusted reputation and relationship of the CHIR group at Arizona State University (ASU) gained the voluntary participation of most major hospital systems as well as other providers and payers in Maricopa County, Arizona, surrounding the Phoenix metropolitan area and containing approximately 60 percent of Arizona's population. In the past year, AZHQ has also expanded significantly in Pima County area (includes Tucson) constituting significant coverage in the most populated areas of the state. Health plans including Arizona's Medicaid program now submit statewide data to AZHQ. CHIR's staff have been able to accept encounter code data in almost any format from all Arizona providers and interface the data into one consistent, searchable data repository. Remaining patient-centric, AZHQ is a unique medical database that allows tracking of cohorts of similar patients regardless of insurance across changing providers and across time in Maricopa County and largely, Arizona. CHIR currently pursues a variety of research initiatives, now including quality measurement, and uses AZHQ to inform much of its research.

The primary care delivery model of Maricopa County is over 90 percent solo practioners, unlike the large corporate medical groups prominent in the other BQI pilot regions. These issues have to be addressed by CHIR as they progress down the development lifecycle demonstrated by all the other regional medical collaboratives of the BQI pilot sites.

Until the BQI project, CHIR's primary focus was not on calculating ambulatory quality measures, providing feedback to physicians, or public reporting. Prior to 2005 their primary focus was on public health research. Among the other BQI pilot sites the only similar facility might be the Regenstrief Institute in Indiana before the creation of the IHIE.

CHIR's respected reputation within the Arizona provider community and the on-going AZHQ database create a unique opportunity for calculating quality measures across patients regardless of insurance, and also for studying the effects and issues raised by combining the provider-encounter data of AZHQ with the Medicare Fee-for-Service (FFS) data of the BQI Project.

The leaders of CHIR are now engaging physician champions, thought-leaders within their professional community, with the goal of realizing the value of accurate comparative data for improving the care doctors provide in their offices. Once a core group of physicians have validated the calculated quality measures, and go on to champion their use within the regional medical community, CHIR hopes to broaden the involvement of other stakeholders. Consideration of issues of pay-for-performance and public reporting are intended to be considered at a more mature stage. By this progressive plan CHIR hopes to build a trusted collaborative community of stakeholders similar to those already in existence at the other five BQI pilot sites.

Data Experience

The aim of the AZHQ project is to develop and maintain a community health data system that houses essential and comprehensive health information for each resident of Arizona. The system is unique for its ability to provide continuously updated health care information and to link patients across systems and over time. The AZHQ database has been in production in its current form since 2003, and currently integrates data from a large number of data sources.

Reporting and Performance Improvement

Many studies have been conducted using the AZHQ data warehouse. Most recently, *The Effect of AHCCCS Disenrollment on Health Care Utilization in Maricopa County* found that enrollment in the Arizona Health Care Cost Containment System, or AHCCCS, was associated with decreased emergency department use and decreased hospitalization as enrollees received more care through routine visits. The report also found that the decision to change eligibility and to disenroll members produces the opposite effect: more emergency department use, more hospitalizations, and less routine care.

Impact and Accomplishments

The Center for Health Information & Research and Arizona HealthQuery have an established track record in data aggregation and reporting at the community level, offering a strong platform on which to build a broader data aggregation and quality improvement agenda. The existing database already includes administrative and encounter data for more than 9 million individuals who have received health care in Arizona, including statewide Medicaid administrative data.

The Center for Health Information & Research has a long history of robust stakeholder participation by providers, health plans, and employers (including some actively involved in national quality measurement efforts such as Health Net of Arizona) and are working towards engagement of other major entities such as CIGNA and Humana, enabling natural synergies with quality measurement activities.

Indiana Health Information Exchange
Indianapolis, Indiana

Mission

The Indiana Health Information Exchange (IHIE) is committed to using information technology and shared clinical information to improve health care in Indiana, enhance health research, and be a national model of health information exchange.

Overview

IHIE was founded by a unique collaboration of 13 institutions representing hospitals, providers, researchers, public health organizations, and economic development groups. Its vision is to use information technology and shared clinical information to:

- Improve the quality, safety, and efficiency of health care in the State of Indiana.
- Create unparalleled research capabilities for health researchers.
- Exhibit a successful model of health information exchange for the rest of the country.

The Regenstrief Institute, BioCrossroads, and the five charter hospital systems are key stakeholders. Other stakeholders include local and state health departments, the state medical society, community health networks, the local quality improvement organization, and the Employers' Forum of Indiana.

The Quality Health First ® (QHF) program was developed to reward improvement in performance, support quality improvement efforts by providers (e.g., the adoption of Electronic Medical Records (EMR)), and, once a reasonable threshold has been reached, concentrate financial incentives from multiple Health Plans and base these rewards on one set of physician-endorsed, community-wide reports.

Data Experience

IHIE covers a nine-county Indianapolis metropolitan statistical area in central Indiana. The patient population consists of approximately 12 percent Medicare Beneficiaries, 12 percent Medicaid Recipients, 16 percent uninsured, with the remainder being commercially insured. Approximately 60 percent of the population is covered by the QHF program. IHIE's data aggregation efforts are built upon those of the Indiana Network for Patient Care (INPC), the oldest, largest, and most successful health information exchange.

Through the initiative with INPC, the exchange aggregates clinical data from several different sources, including hospitals, laboratories, and public health agencies. These data are then complemented with claims data from payers and point of care data from physician offices for the exchange's pay-for-performance project.

Performance Measurement

A consensus of medical directors from primary care practice groups and health plans determined the measures used to start the pay-for-performance program. IHIE drew from established nationally accepted measures, such as those developed by the AQA, the Centers for Medicare & Medicaid Service's Doctor's Office Quality Information Technology program, and the Healthcare Effectiveness.

Data and Information Set.

The criteria for choosing measures include national acceptability, clinical validity, relevance to payers and consumers, likelihood for improvement, and ability to measure outcomes. As the measures are finalized, the AQA's "Parameters for

Selecting Ambulatory Care Performance Measures" are used as a guideline.

Reporting and Performance Improvement

IHIE, through its Docs4Docs Program, delivers a community-wide clinical messaging service providing physicians with a single source for clinical results for laboratory and pathology, radiology, electrocardiogram reports, transcriptions, and emergency department and hospital encounter information from all participating central Indiana hospitals. Through its QHF Program, IHIE reports to providers at the physician and group levels, measure result summaries of provider performance on the included measures as well as individual patient-level alerts and reminders to be used in patient care. Consumer reports will include physician group and community-level data.

Impact and Accomplishments

- IHIE's stakeholder and partner, the Regenstrief Institute, is an internationally recognized informatics and health care research organization. Regenstrief's research scientists have developed the Regenstrief Medical Records System, one of the nation's first electronic medical record systems. Bridges to Excellence citations from investigators at the Regenstrief Institute account for approximately one third of Bridges to Excellence evidence.
- The exchange collaborates with the Indianapolis Patient Safety Coalition to address several important patient safety issues in the inpatient setting.
- IHIE participates in nationwide knowledge-sharing efforts, such as Connecting Communities for Better Health.
- IHIE has been designated a Chartered Value Exchange and is an approved PQRI registry.

Massachusetts Health Quality Partners
Boston, Massachusetts

Mission

Massachusetts Health Quality Partners (MHQP) improves the quality of health care services to the residents of Massachusetts through broad-based collaboration among health care stakeholders.

Overview

MHQP was established in 1995 by Massachusetts health care leaders who recognized the importance of valid, comparable measures to drive improvement. As a coalition of physicians, hospitals, health plans, consumers, purchasers, and government agencies working together to promote improvement in the quality of health care services, the coalition provides physicians and consumers with comparative performance information on physician groups and practices. The coalition brings together a large number of Massachusetts health care organizations, including the State's Executive Office of Health and Human Services, its medical society, hospital association, physician leaders, and several major health plans, all of which collaborate to endorse and disseminate a variety of evidence-based practice guidelines and quality improvement tools.

Data Experience

MHQP has been aggregating physician-level data for primary care physicians across health plans since 2003. The coalition has reported on the comparative performance of primary care physicians on both Healthcare Effectiveness Data and Information Set (HEDIS®) and patient experience measures.

The Massachusetts coalition has developed a unique algorithm to group each individual physician into the appropriate practice site, medical group, and physician network. This allows Massachusetts Health Quality Partners to aggregate and report data at various levels of care from individual physician practices to physicians' offices, medical groups, and networks.

The coalition's data reporting covers roughly 5,000 adult and pediatric primary care physicians in five health plans serving commercially insured enrollees in health maintenance organizations and point-of-service products. More than 50 percent of commercially insured residents were enrolled in these plans during the period covered by the most recent report.

Performance Measurement

MHQP's online report, *Quality Insights: Health Care Performance in Massachusetts*, presents both clinical performance measures and patient experience measures. The clinical measures are drawn from the HEDIS® Measure Set developed by the National Committee for Quality Assurance.

Patient experience measures are fielded from a survey instrument comprised of the best performing items from two validated surveys, Ambulatory Care Experiences Survey and the Consumer Assessment of Healthcare Providers and Systems Clinician and Group Survey. The instrument covers domains characterizing patients' experiences with their primary care physicians, including quality of physician-patient interactions and organizational features of care.

To further analyze performance measures and quality metrics, MHQP has forged partnerships with Tufts New England Medical Center, Massachusetts eHealth Collaborative, Harvard Medical School and Harvard School of Public Health, and the RAND Corporation.

Reporting and Performance Improvement

MHQP completed four cycles (2003–2006) of comparative HEDIS® clinical performance reports to physician groups in Massachusetts, encompassing 2001 to 2006. In February 2005, the Massachusetts coalition issued its first public comparative performance report of 9 physician networks for 16 measures, including preventive care and management of certain chronic diseases. In February 2006, MHQP publicly released HEDIS® clinical results for 150 medical groups. Performance stars are assigned to each group based on the group's performance against three benchmarks: the national 50th percentile, the national 90th percentile, and the Massachusetts statewide rate.

The current report includes over 26 HEDIS® measures, including both chronic care and prevention measures such as Asthma medication for children and adults; Cholesterol screening after a heart attack; Depression in adults; Diabetes care for adults; Well visits for infants, children, and adolescents; Breast cancer screening; Cervical cancer screening; Chlamydia screening; and Colorectal cancer screening.

In March 2006, Massachusetts Health Quality Partners publicly launched results from its first statewide survey of patient experience. The reports present results for more than 400 practice sites. Performance stars for patient experience reporting tells how a physician's office compares to all the other physicians' offices in the State that were part of the coalition's survey.

The current report includes the following patient experience measures:

Quality of Physician-Patient Interactions	**Organizational Features of Care**
Communication	Organizational access
Integration of care	Visit-based continuity
Knowledge of the patient	Clinical team
Health promotion	Office staff

Impact and Accomplishments

- For more than 10 years, MHQP has brought together multiple stakeholders, often with disparate agendas, who have effectively worked together to produce trusted, comparable performance measures that help drive health care quality improvement in Massachusetts.
- MHQP has successfully implemented five public releases of performance information: four with physician performance information and one with hospital performance information. The coalition has designed a user-friendly Web site vetted by a health literacy specialist. The site incorporates findings from consumer focus groups the Massachusetts coalition has conducted to make information accessible and useful. Massachusetts Health Quality Partners has also developed a process to vet the public report and the press release with the coalition's multistakeholder members.
- MHQP has developed a Web-based reporting process to provide physician organizations with performance reports. The coalition reports annually to primary care physicians about performance on clinical HEDIS® measures at the physician network, medical group, practice site, and, if requested by the medical group, at the individual physician level. The Massachusetts coalition reports to primary care physicians and, in 2008, also reported to cardiologists, orthopedists and obstetricians/gynecologists about performance on the patient experience survey.
- MHQP has successfully aggregated health plan claims data and attributed commercial health maintenance organization, point of service, and preferred provider organization patients to primary care and specialist physicians using a visit-based methodology to assign patients. It is the first organization in the country to implement this methodology, which enables it to broaden the patient experience survey to include preferred provider organization members and patients seeing specialists. The physician support and buy-in Massachusetts Health Quality Partners has garnered through its collaborative process means that physicians are more likely to embrace the measures resulting from this process.
- The Massachusetts eHealth Collaborative selected MHQP, in partnership with Computer Sciences Corporation, to pioneer efforts to capture data from electronic health records and translate them into clinical performance measures for use in comparative performance reporting for physicians. The collaborative brings together the State's major health care stakeholders to establish an electronic health record system that enhances quality, efficiency, and safety for health care in Massachusetts.
- MHQP is a founding member of the Network for Regional Healthcare Improvement, an association of regional health improvement collaboratives from around the country.
- Supported by Agency for Healthcare Research and Quality, MHQP is providing data to Harvard researchers to analyze the impact of electronic health records on clinical quality.

Minnesota Community Measurement
St. Paul, Minnesota

Mission

Minnesota Community Measurement (MNCM) accelerates the improvement of health by publicly reporting health care information.

Overview

Since 2002, MNCM's collaborative, community approach has encouraged medical groups to improve health care quality by publicly reporting on several measures. All seven of Minnesota's nonprofit health insurance plans participated in developing the coalition's initial reports, with the Minnesota Medical Association joining the effort in 2005. Since that initial report publication, a health plan in South Dakota and two county-based purchasing organizations have provided data to Minnesota Community Measurement.

The resulting nonprofit community-based organization has a 17-member board of directors, with representation from health plans, hospitals, physicians and medical groups, employers, local quality organizations and consumers. The Reporting Advisory Committee (consisting of physicians and health care quality improvement experts) advises the MNCM President on the scope of data and measures. A separate workgroup of data experts from health plans looks at technical issues around data and reporting.

Data Experience

MNCM data reporting covers more than 100 primary care provider groups representing 700 clinic sites in Minnesota and bordering counties. These groups cover roughly 90 percent of primary care delivered in the State. MNCM has aggregated data across eight health plans and two county-based purchasing organizations, including commercial health maintenance organizations, point of service organizations, preferred provider organizations, Medicaid and State Children's Health Insurance Program managed care, Medicare Advantage, Medicare Cost, and Medicare dual eligibles. The Minnesota coalition reports all measures at the community and medical group levels and is committed to adopting and incorporating nationally accepted standards. Presently, the coalition is working with eight health plans as well as with several of the Bridges to Excellence program's health information networks to obtain additional, more granular level data. These data are submitted directly to MNCM from the medical groups and posted on the coalition's Web site at the clinic-site level.

Performance Measurement

The MNCM 2007 report assessed 13 clinical topics and included more than 40 individual measures. Measures include two composite measures: one for optimal diabetes care and one for optimal cardiovascular care. The diabetes composite measures patients who have met all five treatment targets to decrease their risk of developing cardiovascular disease and other complications of diabetes and the cardiovascular composite measures patients who have met four treatment goals. The coalition uses physician-defined standards of care endorsed by the Institute for Clinical Systems Improvement as well as matching national standards for the measurement of care whenever possible (e.g., NCQA's HEDIS® and NQF endorsed measures).

Reporting and Performance Improvement

Since 2003, MNCM has been reporting on medical groups, with its first public report released in 2004. The coalition's 2007 Health Care Quality Report, based on calendar year 2006 data, compared each medical group against a State benchmark as well as against all other medical groups. The report included the following measures:

Chronic Care Measures

Optimal diabetes care composite (overall diabetes care)
Optimal Cardiovascular care composite
Use of effective medications for asthma
High blood pressure treatment

Preventive Measures

Immunizations for children and teens
Well-baby visits
Breast and cervical cancer screening
Chlamydia screening
Colorectal cancer screening
Cancer screening composite

Acute Care Measures

Appropriate treatment for children with upper respiratory infection
Appropriate testing for children with pharyngitis

Impact and Accomplishments

- MNCM is uniquely situated to rapidly implement data aggregation and reporting on new measures and to demonstrate the impact these efforts can have on improving the health of the community. Key accomplishments and impact on Minnesota's health care system include:
- Providing performance information that encompasses roughly 90 percent of primary care delivered in the State.
- Being one of the first organizations in the country to publicly report community-wide health care measure results by medical group.
- Having strong physician support as a cornerstone of success, as evidenced by physician board representation and leadership roles in advisory groups.
- Having as founding members all licensed Minnesota health plans that continue to provide direct financial support as well as a significant level of in-kind support through data collection and reporting.
- Participating with the Minnesota Business Partnership on developing cost-of-care measures.
- Participating with the Minnesota employer coalition (provide clinic- level performance rates on two quality programs) on implementing a state-wide payment for performance program
- Collaborating with the State's Department of Human Services to develop a useful data collection and reporting process and tool for the State's Medicaid program.
- Operating a Web site for 2 years that provides consumer information on medical groups across the State, with the site receiving an estimated 30,000 visits in November 2005.
- Compiling evidence that overall community rates, including all children's health measures, are improving in Minnesota.

Wisconsin Collaborative for Healthcare Quality
Madison, Wisconsin

Mission

The Wisconsin Collaborative for Healthcare Quality (WCHQ) is a voluntary consortium of organizations learning and working together to improve the quality and cost-effectiveness of health care for the people of Wisconsin.

Overview

The nonprofit, statewide collaborative was founded in 2002 by several health delivery systems, each with a large multispecialty group clinic and tertiary hospital. Encompassing five geographically distinct markets, the collaborative now includes more than 41 physician groups, hospitals, and health plans, including two of the state's largest integrated delivery systems.

WCHQ is governed by a board of directors and funded by member dues and grants. It has the active support and participation of the clinical and administrative leadership of most of Wisconsin's large, multispecialty groups, representing approximately 42 percent of the licensed physicians in the state. With more than 40 reporting entities from virtually every region in the state, the Wisconsin collaborative actively solicits the participation of public and private sector purchasers in its work, ensuring consumer perspective is considered in selecting measures and in preparing the public report.

Data Experience

Although membership includes several health plans and hospitals, the Wisconsin collaborative's primary focus is measuring and reporting on physician groups. Members have tested and verified reliable methods of data collection and aggregation within a broad range of physician group practices. The measurement methodology emphasizes the specification of a denominator that is population based, representing "all patients, all payers" for a given condition. This method of reporting generates highly accurate and actionable information, which in turn has generated a high degree of support for the collaborative within the physician community in Wisconsin.

Performance Measurement

WCHQ has extensive experience in performance measurement at the physician-group level. Over the past 2 years, quality specialists from the collaborative's organizations have developed ambulatory care specifications that join administrative data with more robust clinical results, enabling health systems to collect and report quality of care results on all patients under their care. The collaborative's system and method measure the quality of care administered by health care providers on a given patient population. This approach offers several valuable outcomes, including the following:

- provides a system view of performance with the ability to drill down to provider level.
- includes all patients within a system in the population.
- represents all payers.
- delivers a ready-made patient registry.
- delivers a roadmap for improvement.
- provides the foundation for physician pay for performance.

To date, WCHQ has developed and reported quality on a number of conditions, including diabetes, coronary artery disease, uncomplicated hypertension, postpartum care, and preventive services (colorectal, mammography, pneumococcal vaccination and cervical cancer screening). Its Web-based Performance & Progress Report (www.wchq.org/reporting/) consists of a broad and growing collection of performance measures that compare more than 40 reporting provider

organizations. Each measure represents a specific aspect of care for a defined period that provides a "snapshot" of a given health care organization's performance in relation to an evidence-based standard as well as in relation to one another.

Reporting and Performance Improvement

The Wisconsin Collaborative for Healthcare Quality 2007 *Performance & Progress Report* organized the measures into the following categories:

- Access
- Patient satisfaction
- Critical care
- Pneumonia
- Diabetes
- Surgery
- Health information technology
- Women's health
- Heart care

Reports are available at the physician group, health plan, and hospital levels. The reporting process, which enables physician groups to submit results using a secure Web-based data submission tool, includes two innovative components: a "preview report mechanism" for all reporting entities to use before data are published and a scalable infrastructure that supports significant expansion in measures without changing the reporting platform itself.

The collaborative continues to add participating entities and expand its measures with its unique measures structured so that participants can collect data on all patients within a health system regardless of payer sources, electronic medical record platform, or electronic medical record level of implementation. In 2007, the Wisconsin Collaborative released results on preventive care services, such as breast, cervical, and colorectal cancer screening.

Chronic Care Measures	**Preventive Measures**
Blood sugar control and screening for diabetics	Breast cancer screening
Controlling hypertension	Cervical cancer screening
Kidney function monitoring for diabetics	Colorectal cancer screening
Low-density lipoprotein c holesterol testing and monitoring for diabetics	Pneumococcal vaccination
Low-density lipoprotein cholesterol testing and monitoring for patients with cardiovascular conditions	

Impact and Accomplishments

- WCHQ is a founding member of the Wisconsin Health Information Organization and is responsible for catalyzing its inception. The Wisconsin Health Information Organization is building a data repository to support an expansion in reporting on ambulatory performance. The data will allow providers, employers and consumers to use measures of resource use and cost of care. When these data are combined with the clinical

quality measures generated by the Wisconsin collaborative, stakeholders will be able to assess the value of care by looking at cost and quality over an entire episode of care.

- WCHQ is one of 14 grant recipients of the Robert Wood Johnson Foundation's Aligning Forces for Quality initiative and has been recently recognized as a recipient for the extension of this project.
- WCHQ, in partnership with the Wisconsin Medical Society, will begin the CMS-Electronic Health Record demonstration project in 2009
- WCHQ has been selected as an Improving Performance In Practice site. IPIP is a state-based, nationally led quality improvement program that provides support for physicians to improve care in the practice setting. The IPIP project is funded in part by a grant from the Robert Wood Johnson Foundation with oversight and leadership based within the American Board of Medical Specialties. The goal of IPIP is to equip physicians in smaller practice environments with the systems and support to close the quality gap in asthma and diabetes care in primary care settings through measurement and analysis of care practices followed by improvement activities.
- The Wisconsin Hospital Association's Check Point and Price Point initiatives represent a progressive association-based response to the market's demand for information on the quality and cost of hospital services.
- The National Committee for Quality Assurance's 2005 State of Health Care Quality Report ranked six Wisconsin health maintenance organizations among the top 50 health plans. In addition, a 2006 report by the Agency for Healthcare Research and Quality listed Wisconsin as number one among the 50 States for overall quality of health care services.
- WCHQ has established relationships with key strategic partners, including the Wisconsin Health Information Organization; the Wisconsin Hospital Association; the Wisconsin Medical Society; the University of Wisconsin; the Medical College of Wisconsin; and MetaStar, the Wisconsin Quality Improvement Organization.

Appendix 3

All Measures Considered

Appendix 3
Measures Considered for BQI

					Feasibility		
Used in BQI	Measure Name	Developer	NQF Endorsement Status	AQA Starter Set	Applicable to Medicare population	Data Sources Required[1]	Notes/Considerations
	Childhood Immunization Status	NCQA (HEDIS 2007)	Y	N	N	N/A	
	Adolescent Immunization Status	NCQA (HEDIS 2007)	N	N	N	N/A	
	Appropriate Treatment for Children w/Upper Respiratory Infection	NCQA (HEDIS 2007)	Y	Y	N	N/A	
	Appropriate Treatment for Children w/Pharyngitis	NCQA (HEDIS 2007)	Y	Y	N	N/A	
	Inappropriate Treatment for Adults w/Acute Bronchitis	NCQA (HEDIS 2007)	Y	N	N (age 18-64)	Claims, including pharmacy	Measure could apply to Medicare population that is under 65
√	Colorectal Cancer Screening	NCQA (HEDIS 2007)	Y	Y	Y	Claims	10-year lookback as per clinical guidelines, variation in administrative-only vs. hybrid methods of data collection
√	Breast Cancer Screening	NCQA (HEDIS 2007)	Y	Y	Y	Claims	
	Cervical Cancer Screening	NCQA (HEDIS 2007)	Y	Y	N (age 24-64)	Claims	Measure could apply to Medicare population that is under 65
	Chlamydia Screening in Women	NCQA (HEDIS 2007)	Y	N	N (age 16-25)	N/A	
	Osteoporosis Management in Women Who Had a Fracture	NCQA (HEDIS 2007)	Y	N	Y (aged 67 or older)	Claims, including pharmacy	Sufficient sample size?
√	Controlling High Blood Pressure	NCQA (HEDIS 2007) CMS	Y	N	Y (18-85)	Hybrid only	Chart review for confirmation of hypertension diagnosis and representative blood pressure
√	Beta-Blocker Treatment after Heart Attack	NCQA (HEDIS 2007)	Y	Y	Y (aged 35 and older)	Claims, including pharmacy	Measure retired with release of HEDIS 2008
√	Persistence of Beta-Blocker Treatment after Heart Attack	NCQA (HEDIS 2007)	Y	Y	Y (aged 35 and older)	Claims, including pharmacy	
√	Cholesterol Management for Patients with Cardiovascular Conditions	NCQA (HEDIS 2007)	Y	N	Y (18-75)	Claims, laboratory data (test results)	Laboratory data (test results) required
√	Diabetes – HbA1c testing	NCQA (HEDIS 2007)	Y	Y	Y (18-75)	Claims	
	Diabetes – HbA1c poor control	NCQA (HEDIS 2007)	Y	Y	Y (18-75)	Claims, laboratory data (test results)	Laboratory data (test results) required
√	Diabetes – HbA1c good control	NCQA (HEDIS 2007)	N	N	Y (18-75)	Claims, laboratory data (test results)	Laboratory data (test results) required
√	Diabetes – Retinal eye exam	NCQA (HEDIS 2007)	Y	Y	Y (18-75)	Claims	
√	Diabetes – LDL-C screening	NCQA (HEDIS 2007)	Y	Y	Y (18-75)	Claims	
√	Diabetes – LDL-C control	NCQA (HEDIS 2007)	Y	Y	Y (18-75)	Claims, laboratory data (test results), hybrid	Laboratory data (test results) required
√	Diabetes – Medical attention for nephropathy	NCQA (HEDIS 2007)	Y	N	Y (18-75)	Claims	

					Feasibility		
Used in BQI	Measure Name	Developer	NQF Endorsement Status	AQA Starter Set	Applicable to Medicare population	Data Sources Required[1]	Notes/Considerations
√	Diabetes – Blood pressure control (<140/90)	NCQA (HEDIS 2007)	Y	Y	Y (18-75)	Claims, laboratory data (test results), hybrid	Laboratory data (test results) required
√	Diabetes – Blood pressure control (<130/80)	NCQA (HEDIS 2007)	N	N	Y (18-75)	Claims, laboratory data (test results), hybrid	Laboratory data (test results) required
	Use of Appropriate Medications for Patients with Asthma	NCQA (HEDIS 2007)	Y	Y	N (5-56)	N/A	Measure could apply to Medicare population that is under 65
√	Use of Spirometry Testing in the Assessment and Diagnosis of COPD	NCQA (HEDIS 2007)	N[2]	N	Y (aged 42 and older)	Claims	
	Follow-up After Hospitalization for Mental Illness	NCQA (HEDIS 2007)	N	N	Y (aged 6 and older)	Claims	Obtaining sufficient sample size may be challenging
√	Antidepression Medication Management – Acute Phase	NCQA (HEDIS 2007)	Y	Y	Y (aged 18 and older)	Claims, including pharmacy	
	Antidepression Medication Management – Continuation phase	NCQA (HEDIS 2007)	Y	Y	Y (aged 18 and older)	Claims, including pharmacy	
	Follow-up Care for Children Prescribed ADHD Medication	NCQA (HEDIS 2007)	Y	N	N (aged 6-12)	N/A	
	Glaucoma Screening in Older Adults	NCQA (HEDIS 2007)	N	N	Y (67 and older)	Claims	Feasibility for and applicability to non-Medicare population uncertain
	Use of Imaging Studies for Low Back Pain	NCQA (HEDIS 2007)	Y	N	N (18-50)	Claims	Measure could apply to Medicare population that is under 65
	Disease-Modifying Anti-Rheumatic Drug Therapy in Rheumatoid Arthritis	NCQA (HEDIS 2007)	Y	N	Y (aged 18 and older)	Claims, including pharmacy	
√	Annual Monitoring for Patients on Persistent Medications	NCQA (HEDIS 2007)	Y	N	Y (18 and older)	Claims, including pharmacy	
	Drugs to be Avoided in the Elderly	NCQA (HEDIS 2007)	Y	N	Y (65 and older)	Claims	Feasibility for and applicability to non-Medicare population uncertain
	Potential Harmful Drug-Disease Interactions in the Elderly	NCQA (HEDIS 2007)	N	N	Y (65 and older)	Claims	Feasibility for and applicability to non-Medicare population uncertain
	Medical Assistance with Smoking Cessation (Advising Smokers to Quit)	NCQA (HEDIS 2007)	Y	Y	Y (18 and older)	Survey	
	Flu Shots for Adults Aged 50-64 (Influenza Vaccination)	NCQA (HEDIS 2007)	Y	Y	N (50-64)	Survey	Measure could apply to Medicare population that is under 65
	Flu Shots for Older Adults	NCQA (HEDIS 2007) CMS	Y	N	Y (65 and older)	Survey Claims (CMS)	Feasibility for and applicability to non-Medicare population uncertain
	Pneumonia Vaccination Status for Older Adults	NCQA (HEDIS 2007)	Y	N	Y (65 and older)	Survey	Feasibility for and a pplicability to non-Medicare population uncertain
√	Cardiovascular – LDL testing	NCQA (HEDIS 2007)	Y	N	Y	Claims	
√	Cardiovascular – LDL control	NCQA (HEDIS 2007)	Y	N	Y	Claims, Laboratory results, EMR	Laboratory test results not available from Medicare FFS data

					Feasibility		
Used in BQI	Measure Name	Developer	NQF Endorsement Status	AQA Starter Set	Applicable to Medicare population	Data Sources Required[1]	Notes/Considerations
√	Pneumonia Vaccination	CMS/NCQA	Y	Y	Y	Survey Claims (CMS)	Feasibility for and applicability to non-Medicare population uncertain
	Tobacco Use	AMA / PCPI	Y	Y	Y (18 and older)	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
√	CAD - Drug Therapy for Lowering LDL Cholesterol	AMA/PCPI/ ACC/AHA	Y	Y	Y (18 and older)	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
√	HF - ACE Inhibitor /ARB Therapy	AMA/PCPI	Y	Y	Y (18 and older)	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
√	HF - LVF Assessment	AMA/PCPI	Y	Y	Y (18 and older)	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
	Asthma – Pharmacologic Therapy	AMA/PCPI	Y	Y	N (5-40)	N/A	
	Screening for Human Immuno-deficiency Virus	AMA/PCPI	Y	Y	N (prenatal care)	N/A	
	Anti-D Immune Globulin	AMA/PCPI	Y	Y	N (prenatal care)	N/A	
	Asthma Assessment	AMA/PCPI	Y	N	N (Aged 5-40)	N/A	
	Management plan for people with asthma[3]	IPRO	Y	N	Not specified	Medical records	
	COPD – assessment of oxygen saturation	AMA/PCPI	Y	N	Y (aged 18 and older)	Claims, Medical records	Medical record required for numerator determination
	COPD – inhaled broncho-dilator therapy	AMA/PCPI	Y	N	Y (aged 18 and older)	Claims, Medical records	Medical record required for numerator determination
	Osteroarthritis: Functional and Pain Assessment	AMA/PCPI	Y	N	Y (aged 18 and over)	Claims, Medical records	Medical record required for numerator determination
	Osteroarthritis: Assessment for Use of Anti-inflammatory or Analgesic OTC Medications	AMA/PCPI	Y	N	Y (aged 21 and over)	Claims, Medical records	Medical record required for numerator determination, or consistent use of CPTII code
	Low back pain: Use of Imaging Studies	NCQA	Y	N	N (aged 18-50)	Claims	Measure could apply to Medicare population that is under 65
	Osteoporosis Management in women who had a fracture	NCQA (HEDIS 2007)	Y	N	Y (aged 67 and older)	Claims	Feasibility for and applicability to non-Medicare population uncertain
	Diabetes: Foot exam	NCQA[4]	Y	N	Y (aged 18-75)	Claims, Medical records	Requires consistent use of CPTII code or medical record review
	Diabetes: HbA1c test for pediatric patients	NCQA	Y	N	N (aged 5-17)	N/A	
	CAD – Symptom and activity assessment	AMA/PCPI/ ACC/AHA	Y	N	Y (18 and older)	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
	CAD – ACEI/ARB therapy	AMA/PCPI/ ACC/AHA	Y	N	Y (18 and older)	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
	CAD – Antiplatelet therapy	AMA/PCPI/ ACC/AHA	Y	N	Y (18 and older)	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
	Ischemic Vascular Disease (ICD): Use of Aspirin or other Antithrombotic	AMA/PCPI/ ACC/AHA	Y	N	Y (18 and older)	Claims, EHR, medical record review	Requires chart review for numerator determination
	CAD – Beta Blocker therapy – Prior MMI	AMA/PCPI/ ACC/AHA	Y	N	Y (18 and older)	Claims, EHR, medical record review	Requires chart review and/or electronic health record data

					Feasibility		
Used in BQI	Measure Name	Developer	NQF Endorsement Status	AQA Starter Set	Applicable to Medicare population	Data Sources Required[1]	Notes/Considerations
	CAD: BP Management	NCQA	Y	N	Y	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
	IVD: Complete Lipid Profile and LDL control <100	NCQA	Y	N	Y	Claims, EHR, medical record review	Requires chart review and/or electronic health record data
	CAD: Percentage of members who have optimally managed modifiable risk	Health Partners	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	HF: Assessment of Activity Level	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	HF: Assessment of Clinical Symptoms of Volume Overload (Excess)	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	HF: Left Ventricular Function Assessment	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	HF: ACEI/ARB therapy	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	HF: Patient Education	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	HF: Beta-blocker therapy	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
√	HF: Warfarin therapy patients with atrial fibrillation	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	HF: Weight measurement	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	Hypertension: Blood pressure management	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	Hypertension: Plan of care	AMA / PCPI / ACC / AHA	Y	N	Y	Likely Claims, EHR, medical record review	Likely requires chart review and/or electronic health record data
	Medication management: documentation of medication list in the outpatient record	CMS / SCRIPT	Y	N	Y	Medical records	Requires chart review and/or electronic health record data
	Medication management: documentation of allergies and adverse reactions in the outpatient record	CMS / SCRIPT	Y	N	Y	Medical records	Requires chart review and/or electronic health record data
	Mental health: major depressive disorder, diagnostic evaluation	AMA / PCPI	Y	N	Y	Medical records	Requires chart review and/or electronic health record data
	Mental health: major depressive disorder, suicide risk assessment	AMA / PCPI	Y	N	Y	Medical records	Requires chart review and/or electronic health record data
	Diagnosis of ADHD in primary care for school age children and adolescents	ICSI	Y	N	N	N/A	
	Management of ADHD in primary care for school age children and adolescents	ICSI	Y	N	N	N/A	

					Feasibility		
Used in BQI	Measure Name	Developer	NQF Endorsement Status	AQA Starter Set	Applicable to Medicare population	Data Sources Required[1]	Notes/Considerations
	Bipolar disorder & major depression: Assessment for manic or hypomanic behaviors	STABLE	Y	N	Not specified	Medical records	Requires chart review and/or electronic health record data
	Bipolar disorder & major depression: Appraisal for alcohol or chemical substance abuse	STABLE	Y	N	Not specified	Medical records	Requires chart review and/or electronic health record data
	Bipolar disorder & major depression: Appraisal for risk of suicide	STABLE	Y	N	Not specified	Medical records	Requires chart review and/or electronic health record data
	Bipolar disorder & major depression: level-of-function evaluation	STABLE	Y	N	Not specified	Medical records	Requires chart review and/or electronic health record data
	Bipolar disorder & major depression: assessment for diabetes	STABLE	Y	N	Not specified	Medical records	Requires chart review and/or electronic health record data
	Initiation and engagement of alcohol and other drug dependence treat-ment	NCQA / WC	Y	N	Not specified	Medical records	Requires chart review and/or electronic health record data
	Obesity: BMI in adults >18 yrs of age	NYC – DHMH	Y	N	Y	Medical records	Requires chart review and/or electronic health record data
	Obesity: BMI 2 through 18 years of age	NICHQ	Y	N	N	N/A	
	Blood groups (ABO), D (Rh) type	AMA/PCPI	Y	N	N (prenatal care)	N/A	
	Blood group antibody testing	AMA/PCPI	Y	N	N (prenatal care)	N/A	
	Tobacco Use Prevention and Cessation for infants, children and adolescents	AMA / PCPI	Y	N	N (infants – adoles-cents)	N/A	
	Tobacco Use Assessment and Intervention	AMA / PCPI	Y	N	Y	Medical records	Requires chart review and/or electronic health record data
	Counseling on physical activity in older adults	NCQA	Y	N	Y (65 and older)	Medical records	Requires chart review and/or electronic health record data
	Urinary incontinence management in older adults	NCQA	Y	N	Y (65 and older)	Medical records	Requires chart review and/or electronic health record data
	Fall risk management in older adults	NCQA	Y	N	Y (75 and older)	Medical records	Requires chart review and/or electronic health record data
	Osteoporosis screening	NCQA	Y	N	Y (65 and older)	Medical records	Requires chart review and/or electronic health record data
	Influenza vaccination	AMA / PCPI	Y	N	N	Likely Claims, EHR, medical record review	Likely requires chart review and/ or electronic health record data
	Pneumococcal vaccine needed for all adults aged 65 years and older	RHI	Y	N	Y (65 and older)	Not specified	
√	Optimal Diabetes Composite (HbA1c control, LDL control, BP control, daily aspirin use, documented tobacco free)	MNCM	N	N	Y	Claims, medical record review, EMR	MNCM collects using hybrid methodology and provider direct data submission
√	Diabetes – smoking status	MNCM	N	N	Y	Claims, medical record review, EMR	MNCM collects using hybrid methodology and provider direct data submission
√	Diabetes – daily aspirin use	MNCM	N	N	Y	Claims, medical record review, EMR	MNCM collects using hybrid methodology and provider direct data submission

(Footnotes)
1 Unless otherwise specified, claims = physician, inpatient and outpatient claims
2 AMA / PCPI measure "COPD – Spirometry Evaluation" has been endorsed.
3 http://projects.ipro.org/index/asthma_indicators
4 Diabetes Physician Recognition Program

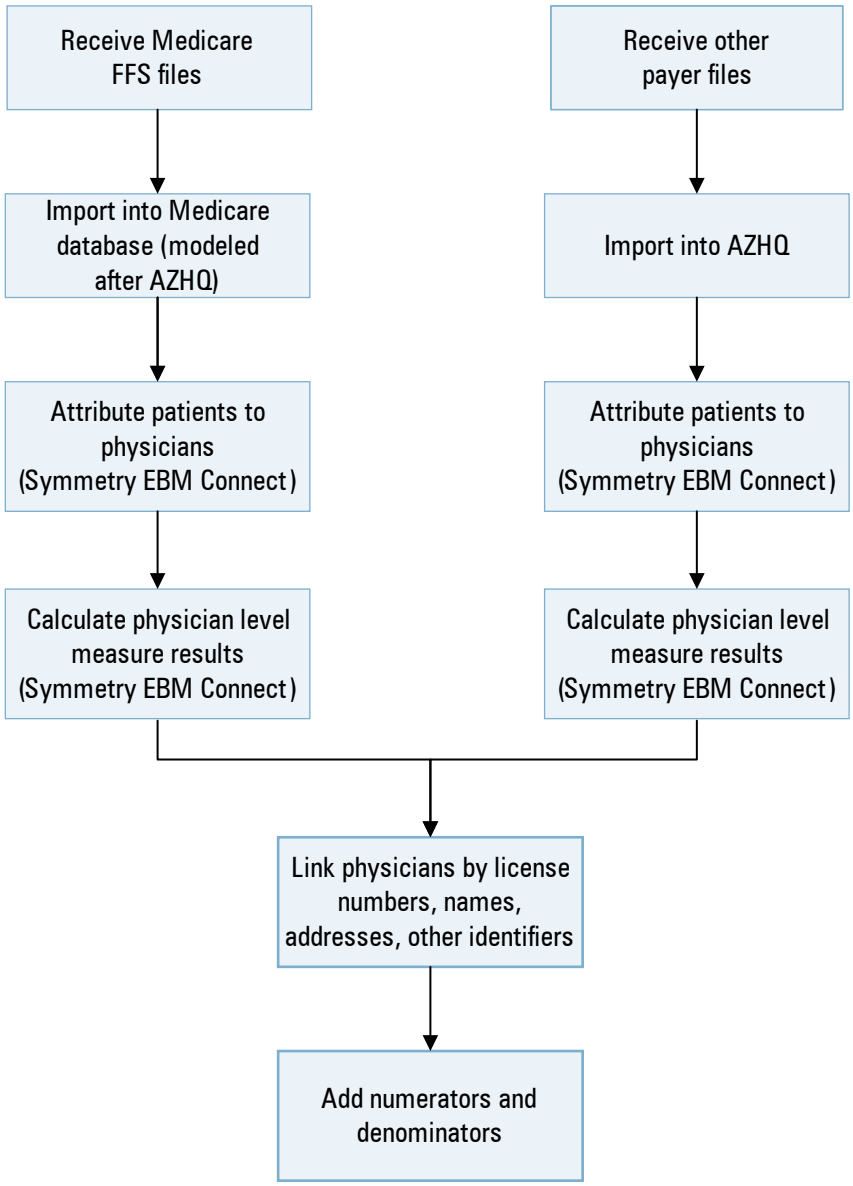
Appendix 4

Pilot Medicare Data Processes
Pilot Attribution Methods

Appendix 4
BQI Pilot Sites Medicare Data Processes

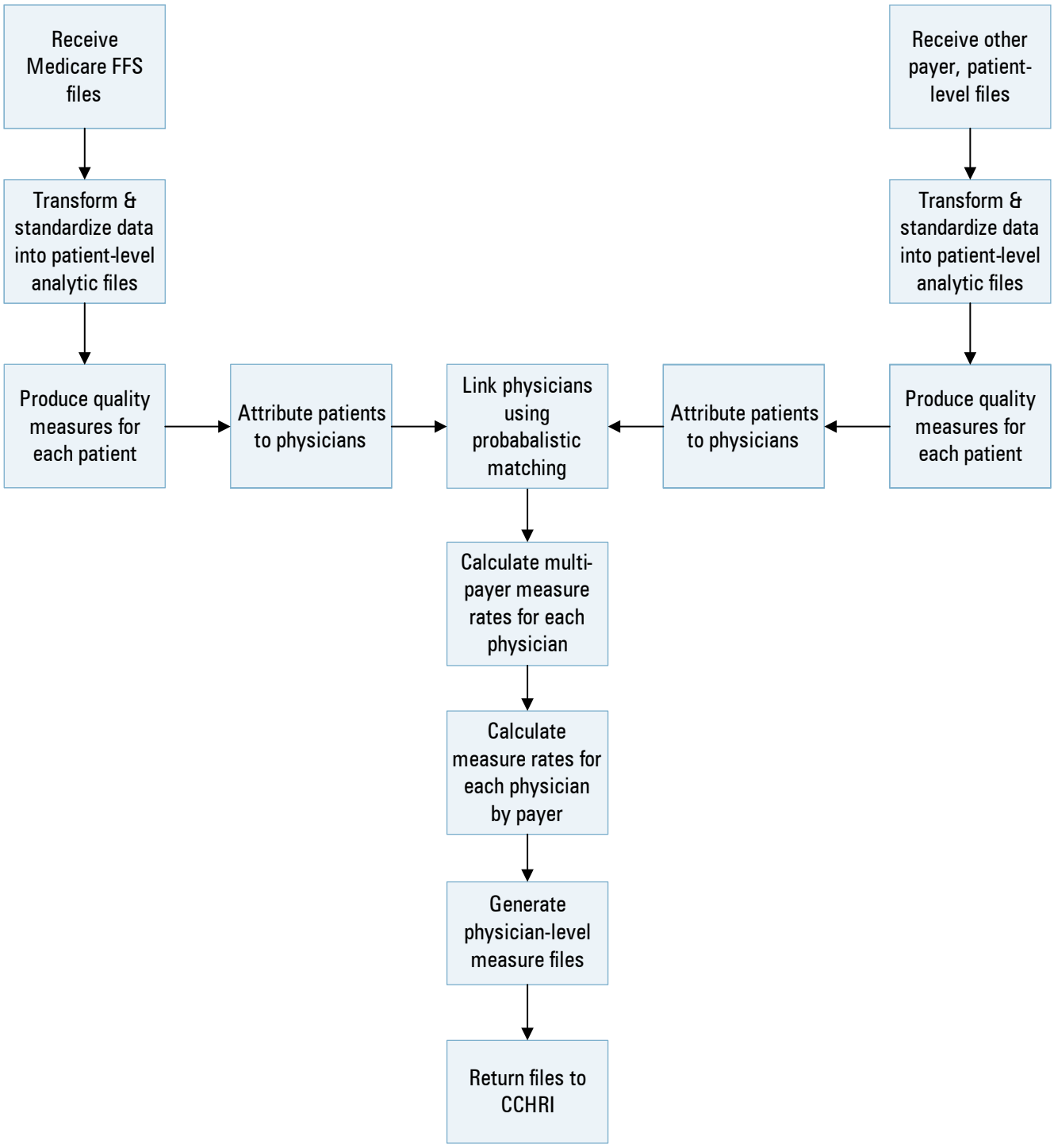
Center for Health Information & Research (CHIR)

CHIR received the raw Medicare FFS data and imported it into a Medicare-only database which was modeled on its existing multi-payer database, Arizona HealthQuery (AZHQ). CHIR then used Symmetry EBM Connect ® software on the patient-identified data in both databases to attribute patients to physicians and generate physician level results. CHIR linked physicians across both result sets by using identifiers such as license numbers, names, addresses, to create aggregated, multi-payer measure results.



California Cooperative Healthcare Reporting Initiative (CCHRI)

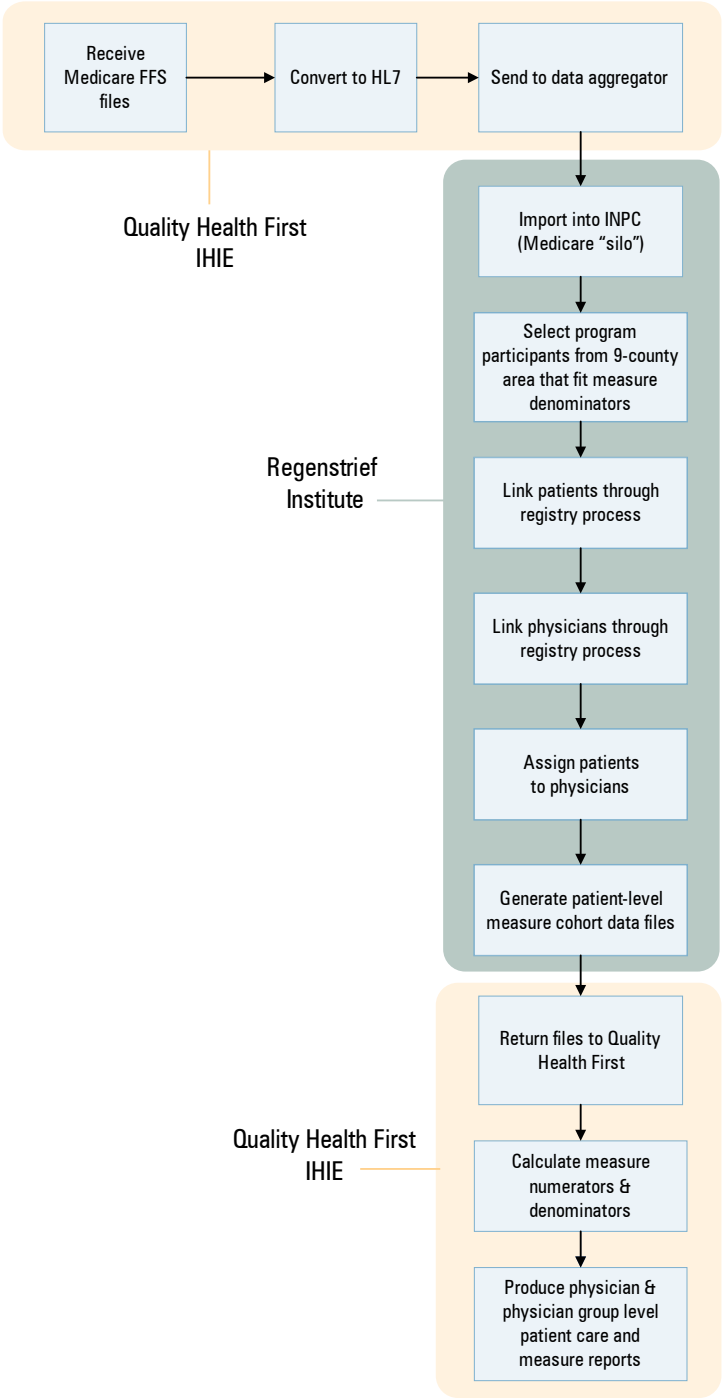
CCHRI contracted with Thomson Reuters to perform its data aggregation and analysis. Thomson Reuters received the raw Medicare FFS data and created standardized, patient-level analytic files. Thomson Reuters calculated measure results by patient, then attributed patients to physicians. Using probabilistic matching, Thomson linked physicians across payers and created multi-payer measurement results, as well as results by payer.



Indiana Health Information Exchange (IHIE)

For IHIE, Delmarva first pulled the Medicare FFS raw, EBCDIC files off the CMS-supplied tape cartridges, then converted the EBCDIC compressed, variable-length files to fixed-width ASCII files containing patient-level, identified data.

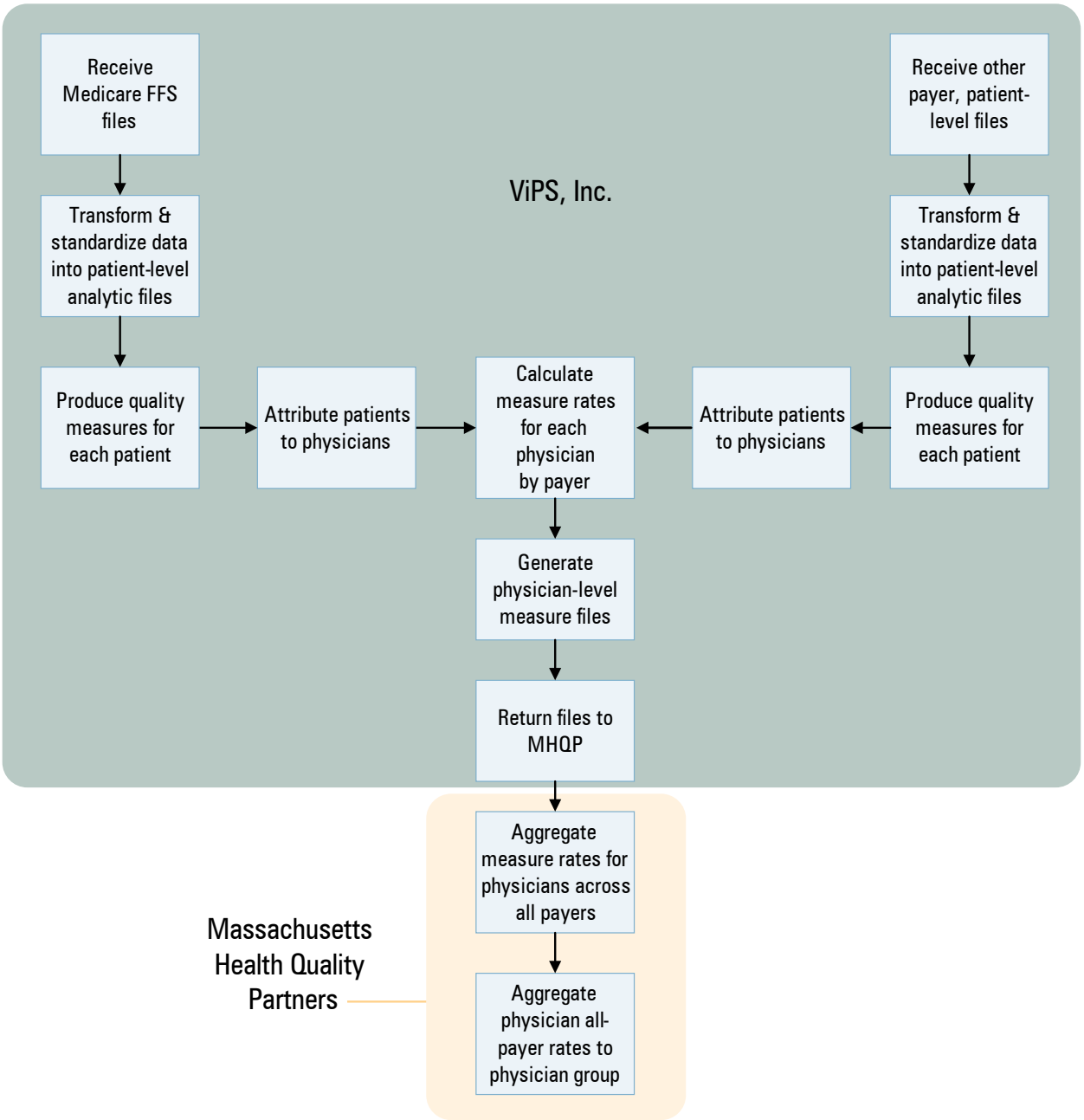
Delmarva sent these files to IHIE, which processed the data, mapped it to a standard, synchronized format internally, then converted the data to a standard, patient-level HL7 (Health Level 7) format that its aggregator, Regenstrief Institute (RI) requires. Regenstrief, in turn, imported the data into a specific area in the Indiana Network for Patient Care (INPC) database system. Then, patients and physicians were linked across payers prior to patient attribution and measure calculation. Physician group level results were generated from the individual physician level results.



Massachusetts Health Quality Partners (MHQP)

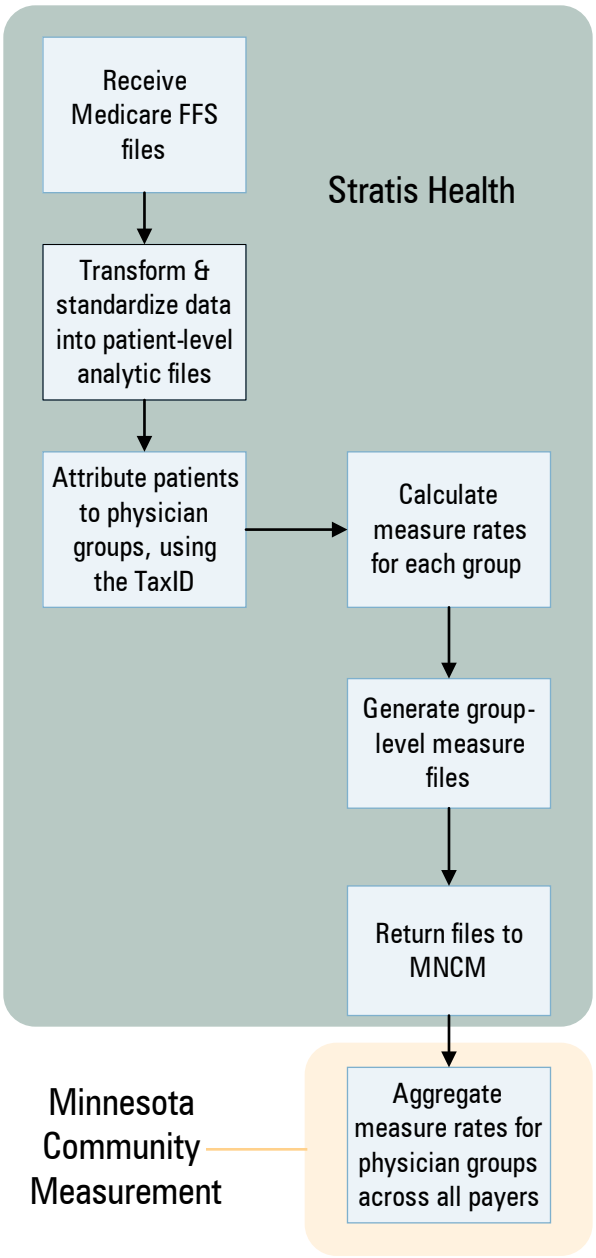
For MHQP, Delmarva first pulled the Medicare FFS raw, EBCDIC files off the CMS-supplied tape cartridges, then converted the EBCDIC compressed, variable-length files to fixed-width ASCII files containing patient-level, identified data.

Delmarva sent these files to ViPS, with which MHQP had contracted to perform data aggregation and calculate measure results. ViPS first transformed the patient-level Medicare FFS data into standardized patient-level analytic files. Then, ViPS calculated the quality measure results for each patient, attributed the patients to particular physicians, and returned a physician-level measurement file to MHQP. MHQP then aggregated the physician-level results to physician-group results, using its internally developed and maintained Master Provider Directory.



Minnesota Community Measurement (MNCM)

MNCM subcontracted with Stratis Health, the Minnesota Quality Improvement Organization (QIO), to process the Medicare FFS patient-level data. Using the Carrier claims, Stratis extracted the relevant patient-level variables into an analytic dataset. Stratis subsequently attributed patients to physician groups, using the TaxID number (TIN) as the means of identifying a group. Then, Stratis calculated measure rates at the group level and returned a physician-group level file to MNCM for aggregation with its other payer results.

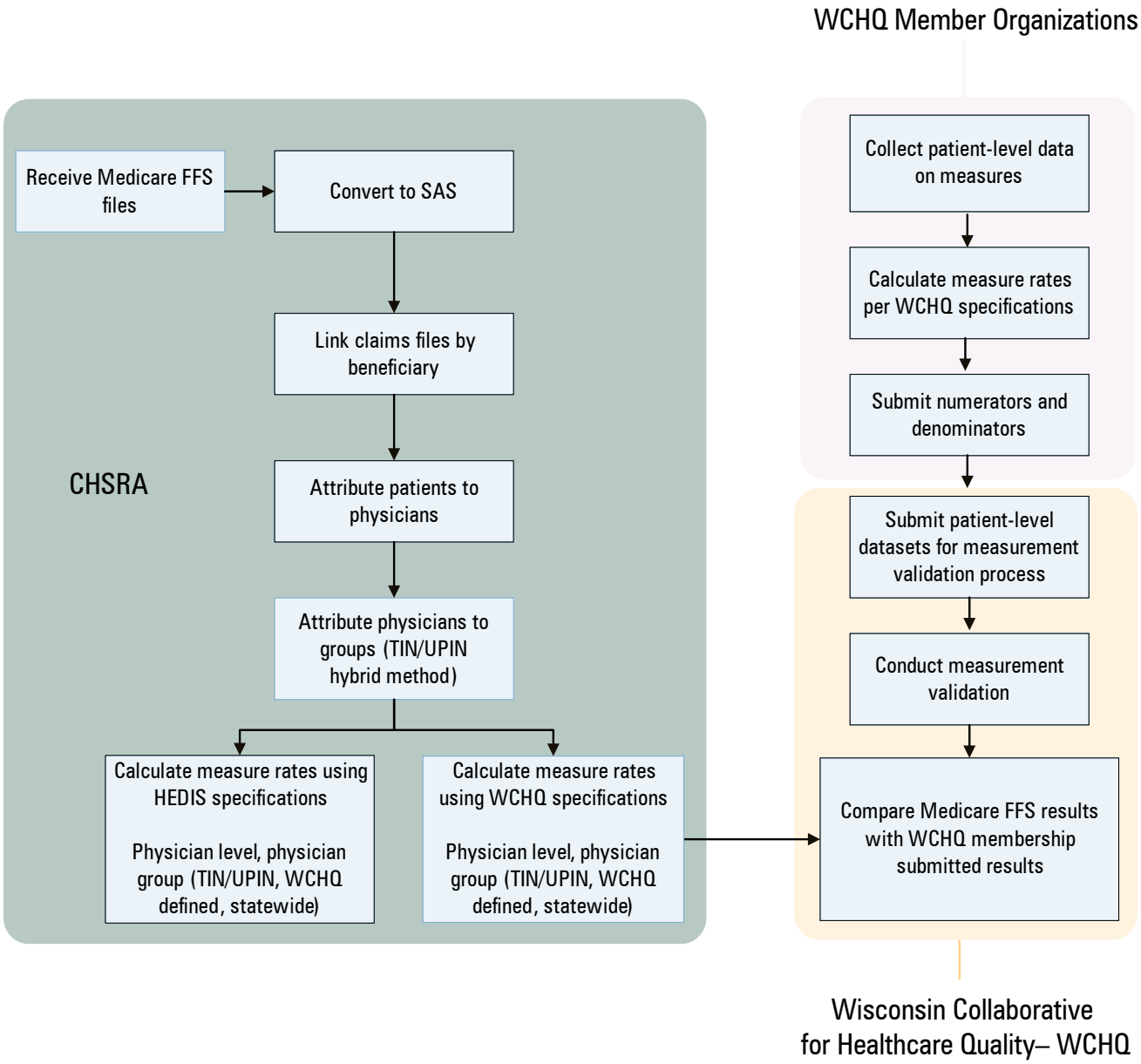


Wisconsin Collaborative for Healthcare Quality (WCHQ)

WCHQ contracted with the Center for Health Systems Research & Analysis (CHSRA) at the University of Wisconsin-Madison to process the Medicare FFS data and apply the WCHQ measurement methodology to the FFS data.

CHSRA extracted specific patient-level data elements from the Medicare FFS data to create patient-level analytic files. They linked the claims (Inpatient, Outpatient, Carrier, Skilled Nursing Facility) files by beneficiary, then attributed patients to physicians. Physicians were attributed to groups using a hybrid TIN and UPIN approach, after which CHSRA calculated measure rates at the physician group level.

CHSRA provided WCHQ with physician-group level measurement results, which WCHQ used to compare with the data directly submitted by its membership.



Attribution of Patients to Physicians

Pilot	Method
CHIR	EBM Connect – PCP Imputation Process Physical exam or assessment performed by physician w/ allowed specialty (limited to selected specialties) and who is the PCP assigned via enrollment process. Most recent physical exam or assessment performed by physician other than assigned PCP, limited to allowed specialities Physician who is allowed specialty who is other than the assigned PCP and who performed largest number of evaluation and management type visits (i.e., maximum frequency rule)
CCHRI	The CCHRI BQI project tested three different methods of assigning patient events to physicians as part of the Cycle 1 work, and one method was selected for use in scoring work for all Cycles of the BQI project. The selected method ("Method 2") assigns denominator eligible patients to every physician of a relevant specialty for that measure who had at least one E&M visit with the patient during the measurement period. (i.e., one-touch rule) By the end of the project, the decision was made to align the numerator and attribution periods so that a physician must have seen the member for an E&M visit during the time period in which they were to have received the numerator service. In Cycles 1 and 2, the attribution period was consistent with the denominator period, not the numerator period.
IHIE	In order to attribute the patient to a provider, IHIE created an algorithm that creates a rank ordered list of physician associations with the patient. IHIE then uses data about the providers including their specialty to identify the primary care providers (versus the oncologist who may be seeing the patient weekly to administer and monitor chemotherapy). The current version of the algorithm relies on actual encounters that occurred (not appointments), laboratory results and prescriptions. Patients fall into one of several categories: A. Patients who have not had interactions with any providers B. Patients who have had interactions with only one provider that meets criteria to be a PCP C. Patients who have had interactions with multiple providers that meet criteria to be PCPs.
MHQP	Attribute care for patients in managed care products to the PCP assigned to them by their health plan Attribute care for patients in PPOs and Medicare FFS to the PCP who had the highest volume of E&M office visits with that patient in the measurement period. (i.e., maximum frequency rule) For PPO and FF patients with no E&M office visits to a primary care physician in the measurement period, attribute diabetes care only to endocrinologists where the member had E&M office visits to one specialist only and that specialist was an endocrinologist and attribute breast cancer screenings to gynecologists and cardiac care measures to cardiologist where the member had E&M visits to one specialist only andthat specialist was a gynecologist in the first case or a cardiologist in the second. Note: The endocrinologist will not be held responsible for breast cancer and colorectal cancer screenings but will be held accountable for diabetes measures.
MNCM	Each member in the eligible population for each measure will be attributed to one Medical Group (based on TIN) for the measurement year based on claims/encounter data for selected services (E&M codes and Preventive codes) received in that measurement year. Non-diabetes measures: Attribute the member to the TIN with the highest number of E&M claims/encounters (i.e., maximum frequency rule) only if those claims/encounters are associated with the following specialties: general practice, family practice, internal medicine, pediatrics, geriatric medicine, obstetrics and gynecology, cardiology, physician assistant, nurse practitioner. If there is a tie between a primary care and specialist provider, pick the TIN with the primary care provider visit. If there is a tie between 2 primary care providers, select the one with the most recent date of service. Primary Care defined as general practice, family practice, internal medicine, pediatrics, geriatric medicine, physician assistant, and nurse practitioner. For those members that have claims/encounters that are not associated with one of the specialties listed above, assign those members to Medical Group 0 (zero). Eligible members not attributed to a valid Medical Group ID in the previous steps will be attributed to Medical Group ID 999. Additional steps for diabetes measures: For those members that are attributed to Medical Group 999, repeat steps above and attribute members who cannot be attributed by the above specifications, to the TIN with the highest number of E&M or diabetes claims/encounters (i.e., maximum frequency rule) during the measurement year regardless of specialty. Map these newly assigned TINs to Medical Group IDs as described in the annual MNCM Medical Group Aggregation Matrix. Some members may still be attributed to Medical Groups 0 and 999.

Pilot	Method
WCHQ	WCHQ members determine responsibility for their patients based on three questions: For disease-specific measures: 1) Is this a patient with the disease or condition? – Patients require a defined number of office visits for their condition to qualify for the measure. 2) Is this patient whose care is managed within the physician group? Patients are required to be managed by the physician group in order to be eligible for the measure. 3) Is this a patient currently managed in our system – Patients must be currently managed by the physician group in order to be included in the measure. For other measures: 1) Is this a patient we manage? - Patients are required to be managed by the physician group in order to be eligible for the measure. 2) Is this a patient that is current in our system? Patients must be currently managed by the physician group in order to be included in the measure. 3) Is this a patient that is eligible for the measure? – Patients who meet the defined measure eligibility criteria.

Attribution of Physicians to Physician Groups

Pilot	Method	Comments / Example
CHIR	Using available physician licensing data, in conjunction with other available public information, to supplement the provider information obtained via the CMS data. We then arranged the physicians into groups by practice location, using a variety of text matching methodologies. The results were then validated both internally and externally.	
CCHRI	Not applicable	CCHRI is calculating measures at the individual physician level.
IHIE	IHIE requires Physician Groups (can be 1 to many providers in a group), to attach a list of physicians in the Group when they enroll in the Quality Health First (QHF) program. QHF collects the physician name, Tax ID, NPI, Specialty and maintains that information in the QHF system. Physicians submit updates as needed – in order to receive QHF reports and incentives.	
MHQP	MHQP's method for assigning physicians to Groups has evolved over the four years MHQP has been doing this. MHQP's initial physician grouping was based on information from the health plans. In 2002 and 2003, MHQP allowed physician groups to submit rosters to clarify discrepancies and correct errors. In 2004, MHQP built a simple web-based facility to allow physician groups to directly edit their rosters, which MHQP has enhanced over the years. • Groupings for existing physicians (PCPs) are carried forward. • Newly added physicians are assigned to practices (where possible) and medical groups based on plan information. • Medical groups manage their organization's information, physician rosters and physician details.	
MNCM	Not applicable.	MNCM relies on the TaxID to identify its reporting groups.
WCHQ	Not applicable.	WCHQ membership determines which physicians are in a group, and submits group-level results.

Appendix 5

The Effects of Differing Attribution Methods—Beneficiaries to Physicians

Appendix 5
The Effects of Differing Attribution Methods—Beneficiaries to Physicians

One of the most challenging features of measuring physician performance when using administrative claims data from Fee-for-Service (FFS) environments is that no single method for assigning the responsibility for a beneficiary to a physician is clearly superior to other possible methods. When using Medicare FFS data, each method functions to optimize one aspect or another of the methodology through quantitative differences in the approaches to the various options. The goal of the analyses in this report, therefore, is to clarify the quantitative differences between a selection of attribution methods for their effects on physician rates for five HEDIS® process measures.

Methods

The analyses were conducted using Medicare FFS administrative data for six states (i.e., Arizona, California, Indiana, Massachusetts, Minnesota, and Wisconsin) for the years 2003 through 2006. The five quality measures were calculated using Medicare FFS data for the six states from the 2006 data. The prior years' data were used as look-back periods for some of the 2006 rates. The 2007 HEDIS® specifications were used to calculate the rates for five process measures: breast cancer screening, colorectal cancer screening, HbA1c testing, LDL testing, and nephropathy testing. These five measures were selected because they are consistently defined across all six states.

To test for differences in results that could arise from the methodologies associated with both perspectives, we examined five differing attribution rules (i.e., algorithms) that are in common use. Two rules attribute care to all physicians who have a minimum number of evaluation and management (E&M) visits with the eligible patients (one-touch, two-touch). The other three rules attribute care to physicians based on whether a certain proportion of the patient's E&M visits were with the physician(s) (30 percent, 50 percent) or whether the physician had the most number of E&M visits with the patient (maximum frequency). Table 1 lists these rules, with a brief description of each.

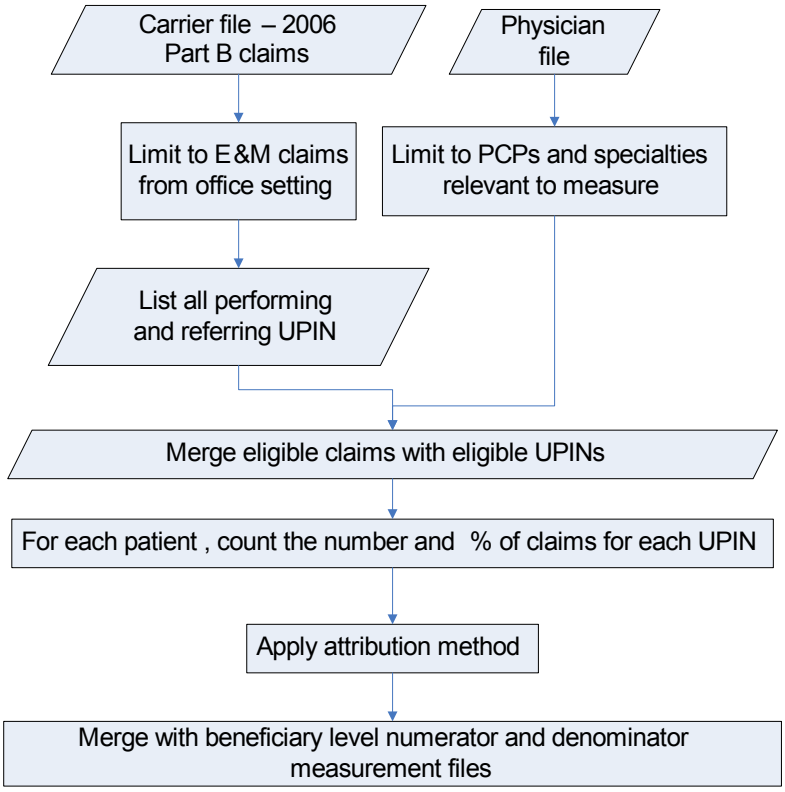
Table 1. Attribution Rules Tested

Attributing Beneficiaries to Physicians with Administrative Data	
Rule	**Description**
Potentially Attributes One Patient to Multiple Physicians	
One-touch rule	Beneficiary is attributed to every physician touched at least once (based on the number of E&M visits) by the beneficiary.
Two-touch rule	Beneficiary is attributed to every physician touched at least twice (based on the number of E&M visits) by the beneficiary.
30% rule	Beneficiary is attributed to each physician who submitted at least 30% of the total office visits (based on the number of E&M visits) for that beneficiary – to a maximum of three physicians.
Attributes One Patient to Only One Physician	
50% rule	Beneficiary is attributed to the physician who submitted at least 50% of the total office visits (based on the number of E&M visits). If tied, 1 physician was randomly selected.
Maximum frequency	The physician with the highest number of claims based on the number of E&M visits) for a beneficiary is attributed to them. If tied, the most recently performing physician is selected.

The data flow for calculating the attribution methods is described in Figure 1. A base file was created by starting with the 2006 carrier file for each state and then limiting the data to only E&M visits in an office setting (see Appendix E of the Cycle 2 Report for codes used to identify E&M visits). The list of eligible Unique Physician Identification Numbers (UPINs) was limited to physicians who were either performing or referring. UPINs on an eligible claim and who had a specialty related to the measure (see Appendix F of the Cycle 2 Report for list of specialties per measure).

This step of limiting UPINs to those with relevant specialties required three runs of attribution algorithms – once each for the breast cancer screening, colorectal screening, and the diabetes measures. The list of eligible UPINs is then merged to the list of eligible claims and each of the five attribution algorithms is in turn applied. At that point, the resulting beneficiary – physician pairing list is merged with a beneficiary level file that contains the numerator and denominator information for each measure.

Figure 1. Data flow for attribution methods



To compare the five attribution methods, the means of the physician rates, the percentages of eligible physicians with a rate, percentage of eligible beneficiaries attributed to a physician with a rate, and the beta-binomial reliability are calculated for each attribution rule for each state and process measure. To evaluate the impact of a minimum denominator on physician process measure rates, the attribution rules are applied to each measure after including only physicians with a minimum denominator of 10, 20, 30 and 50 eligible beneficiaries (i.e., minimum patient count).Complete analysis results can be found at the end of this discussion, beginning on page 19.

The analytic methodology for this section needs to clarify the independent effects from:

- the minimum acceptable eligible physician denominator,
- attribution rule applied,
- the specific process measure, and
- combined state-site effects on:
  - physician process measure rates,
  - the percentage of eligible physicians with a rate,
  - the percentage of eligible beneficiaries attributed to a physician with a rate, and
  - the average beta-binomial reliability of physicians' rates.

As such, multiple regression analysis was selected and dummy variables were created for each attribution rule, process measure, and state. The coding for the minimum acceptable physician denominator is intrinsically in a ratio scale and, therefore, did not require recoding. For the multiple regression, the one-touch rule (i.e., an attribution rule), HbA1c exams for identified beneficiaries with diabetes (i.e., a process measure), and Minnesota (i.e., a state location) were chosen as the relative frames of reference. The one-touch rule was chosen as the most inclusive attribution method. HbA1c exams and Minnesota were chosen due to their unadjusted average physician rates exceeding the average physician rates of the other process measures and states, respectively. These choices of frames of reference thereby provide an anchor for one end of each characteristics' distribution along the primary dependent variable of interest (i.e., physician process measure rates).[1]

Four non-weighted multiple regression analyses were conducted: one analysis for each of the four identified dependent variables of interest in the project (i.e., physician process measure rates, the percentage of physicians with a rate, the percentage of eligible beneficiaries attributed to a physician with a rate, and the average beta-binomial reliability of physicians' rates). The decision was made not to weight the analyses for several reasons. Perhaps the most important reason was that the choice of weighting variables depends more upon the nuances of the policy questions being addressed by the analysis than it does on the structure of the data. The average number of attributed beneficiaries per physician, the total numbers of physicians, and the total numbers of beneficiaries within each category were available weighting choices. Each weighting choice skews the results somewhat towards a specific policy perspective. For example, weighting by either the total number of physicians or the total number of beneficiaries shifts the calculated values towards California's results, due to the state's size. Furthermore, none of the available weighting variables is essentially independent of the minimum denominator constraints. Lastly, few substantively important changes were seen in either the unstandardized or in the standardized regression coefficients from any of the types of weighting discussed. The issue of appropriate weighting for these types of aggregates is, therefore, left for future research on the topic.

Results and Discussion

The following sections of the report separately examine the independent effects of (1) the minimum acceptable eligible physician denominator, (2) the attribution rule applied, (3) the specific process measure, and (4) combined state-site effects on each of the four dependent variables of interest. This Appendix compares the relative importance of each independent variable across the four analyses just described, discusses the limitations for the findings, and provides suggestions for future work on the topic.

Effects on Physician Process Measure Rates

Table 2 presents results for the regression analysis for the average physician rates. The table shows both the unstandardized and the standardized regression coefficients, along with their significance level in a population of 600 observations. The unstandardized regression coefficients represent the average percentage point change in physician process measure rates for a one unit movement in the independent variable, adjusting or accounting for the effects of every other variable in the model. The standardized regression coefficients also represent the expected change in physician process measure rates, but in a common metric, so the relative contributions of each variable to explaining average physician rates can be compared. The p-values show the estimated probability that the results are ascribable to random error within the measurement system.

1 As frames of reference, the anchoring characteristics (i.e., 1-touch, HbA1c, Minnesota) are not separately presented in the tabled results but, instead, are used as frames of reference for their substantively related entries.

Table 2. Effects on Physician Process Measure Rates

Explanatory variables	Unstandardized Coefficients	Standardized Coefficients	P-value	Relative Importance
Patient Count				
Minimum Patient Count	0.022%	0.024	0.002	12
Attribution Rule				
2-Touch Rule	0.76%	0.022	0.020	13
30% Rule	-2.1%	-0.061	< 0.001	10
50% Rule	-2.3%	-0.067	< 0.001	8
Maximum Frequency Rule	-2.0%	-0.057	< 0.001	11
Process Measure				
Breast Cancer Screening	-15.8%	-0.459	< 0.001	3
Colorectal Cancer Screening	-38.7%	-1.125	< 0.001	1
Diabetes - LDL Screening	-5.1%	-0.149	< 0.001	5
Diabetes - Nephropathy Screening	-18.0%	-0.522	< 0.001	2
State				
Arizona	-2.7%	-0.065	< 0.001	9
California	-4.2%	-0.114	< 0.001	6
Indiana	-7.6%	-0.206	< 0.001	4
Massachusetts	-3.1%	-0.084	< 0.001	7
Wisconsin	-0.05%	-0.001	0.885	14

Notes: **Frames of reference: 1-Touch Rule, Diabetes – HbA1c Testing, and Minnesota**
N = 600 aggregate physician process measure rates
$R^2 = 0.967$, $AdjR^2 = 0.966$
Listed percents are interpreted as percentage points
P-values less than 0.05 are statistically significant

The table shows statistically important contributions from each of the categories of measures to explaining physician rates. Starting with the minimum patient count, the variable is statistically significant but shows a relatively low standardized coefficient, lending to it being the 12th most important contributor to explaining physician rates of the 14 variables in the model. Only the relative contribution of the 2-touch rule and having data from Wisconsin (i.e., statistically no difference to being from Minnesota) were less important to explaining categorized physician process measure rates. Furthermore, the unstandardized coefficient suggests that every increase of a single person to the minimum required patient count (i.e., denominator size) results in an average increase of 0.022 percentage points in physician process measure rates. This difference in choices between minimum denominators that could range from 10 to 50 eligible patients results in only an expected change in average physician rates of 0.88 percentage points.

The choice of attribution rules results in a substantively larger impact on the resultant physician process measure rates than does the minimum patient count, as much as three times the impact in a z-score-type metric. Although a small gain in average physician rates is seen when switching to the 2-touch rule (i.e., 0.76%), substantively larger losses are seen when the 30%, 50%, and maximum frequency rules are used, with 2.1%, 2.3%, and 2.0% average reductions seen in physician process measure rates, respectively. The effect of each of the attribution rules on physician process measure rates is statistically different from the results of the 1-touch rule.

Overall, the choice of process measure has the greatest impact on physician process measure rates. Of the 14 variables in the model, the process measures are the first, second, third, and fifth most important contributors to explaining the variation seen in physician process measure rates. The relative impact of the choice of process measures lowers the average physician score from that of the Diabetes - HbA1c test scores by 5.1 (i.e., Diabetes – LDL screening) to 38.7 (i.e.,

Colorectal cancer screening) percentage points. The importance of the choice of process measures clearly dominates the explanation of the variability of average physician process measure rates, with the lone exception of the data coming from Indiana, discussed next.

Of the four categories of choices in this section of the report (i.e., minimum patient count, attribution rule, process measure, and state), the state generating the data was, overall, a less important contributor to physician process measure rates than the specific process measure examined but more important than the attribution rule followed or the minimum patient count required. On average, the choice of states can result in a change in average physician process measure rates of as much as 7.6 percentage points, adjusting or controlling for every other variable in the model. Yet, Wisconsin's average physician rates showed no statistical difference from Minnesota's average physician rates, controlling for the other variables in the model. Nonetheless, an examination of the relative importance of the state variables shows that state-by-state variation, overall, overwhelms considerations of attribution rule and minimum patient counts as important predictors or explanatory variables for differences in average physician process measure rates for four of the five states compared with Minnesota's average physician process measure rates.

One critically important feature of the analysis for interpreting its importance to policy considerations is its accuracy. The values for R^2 and $AdjR^2$ indicate the degree of association between variation in the dependent variable (i.e., physicians' process measure rates) and the combined effects of the other variables in the model. The R^2 for the model was 0.967, with an adjusted R^2 of 0.966. Those exceptionally high values mean that the multiple correlation between the explanatory variables and physician process measure rates was over 0.98, which is an extraordinarily high multiple correlation. These findings are especially important when the range in the average physician process measure rates is considered. These rates (for cells not found in Tables A9-1 through A9-30 listed at the end of this appendix) range from 38.9 percent for the average rate found under the 10-patient minimum count and 50% attribution rules for colorectal cancer screening in Indiana to 92.6 percent for the average rate for the 50 patient minimum count and 50% attribution rule for Diabetes – HbA1c testing in Minnesota. Furthermore, the trivially small reduction in the R^2 of 0.967 to 0.966 through adjusting for the likely non-randomness in the model due to the number of explanatory variables suggests that the accuracy of the results is not due to over-fitting the model. Although the model was not built to explain individual physician rates, it is likely to be remarkably accurate for explaining average category rates for physician process measures within the included states, and possibly for the nation.

With one of the most important goals of public health being to equalize the health care given to Medicare beneficiaries across the country by improving disparate quality care, a resolution to these state differences would seem to be a higher priority than future research on minimum patient counts and attribution rules at this stage of understanding. These state differences cannot be explained by differences in the measures (i.e., all are appropriately structured HEDIS® measures), attribution methods, or minimum denominator counts, as all of these factors were entered into the model. Furthermore, the measures used for this analysis were process measures. All eligible patients are supposed to receive the services associated with process measures. These results, therefore, are not influenced by patients who have not seen a physician; rather, the results are likely, but not exclusively, influenced by differences in the following: physician or office staff behaviors (including a differential ability to persuade patients), levels of beneficiary support, access to ancillary facilities (e.g., labs), data processing, and data aggregating techniques and methodologies. At this point in time and considering only the constructs tested by the model, increased efforts are needed for explaining and resolving state differences in average physician process measure rates more than those efforts are needed for gaining a better understanding of issues with a substantively smaller relative impact on those rates (i.e., minimum patient count and attribution method).

Effects on percentage of physicians with a rate

Table 3 presents results for the regression analysis for the percentage of physicians with a process measure rate due to having a sufficient number of attributed patients. This table shows both the unstandardized and the standardized regression coefficients, along with their significance level in a population of 600 observations. The interpretations of the unstandardized and the standardized regression coefficients along with their p-values remain unchanged from previously discussed, with the exception of the percentage of physicians with a process measure rate being the target or explained variable instead of physicians' process measure rates (i.e., the dependent variable in the multiple regression analysis).

Table 3. Effects on the Percentage of Physicians with a Process Measure Rate

Explanatory variables	Unstandardized Coefficients	Standardized Coefficients	P-value	Relative Importance
Patient Count				
Minimum Patient Count	-0.65%	-0.592	< 0.001	1
Attribution Rule				
2-Touch Rule	-8.8%	-0.216	< 0.001	7
30% Rule	-14.3%	-0.352	< 0.001	5
50% Rule	-18.4%	-0.451	< 0.001	3
Maximum Frequency Rule	-17.4%	-0.428	< 0.001	4
Process Measure				
Breast Cancer Screening	-1.0%	-0.025	0.069	12
Colorectal Cancer Screening	23.0%	0.563	< 0.001	2
Diabetes - LDL Screening	< -0.01%	< 0.001	0.996	13
Diabetes - Nephropathy Screening	< -0.01%	< 0.001	0.997	14
State				
Arizona	3.1%	0.064	< 0.001	10
California	3.3%	0.075	< 0.001	9
Indiana	14.8%	0.339	< 0.001	6
Massachusetts	1.4%	0.033	0.014	11
Wisconsin	8.4%	0.192	< 0.001	8

Notes: **Frames of reference: 1-Touch Rule, Diabetes – HbA1c Testing, and Minnesota**
N = 600 aggregate physician process measure rates
$R^2 = 0.931$, $AdjR^2 = 0.930$
Listed percents are interpreted as percentage points
P-values less than 0.05 are statistically significant

The table clearly shows that the minimum patient count required for generating a physician process measure rate is the most important contributory variable for explaining differences in the percentage of physicians with a process measure rate. Over the tested range of potential minimum patient counts (i.e., 50 - 10 = 40 patient difference in minimum requirements), the average effect of moving from one extreme to the other (i.e., from 10 to 50 patients required to generate a rate), the expected percentage of physicians with a calculable rate lowers by 26 percentage points (i.e., -0.65x40= -26.0).

The second most important contributor to explaining the variation in the percentage of physicians with a calculated process measure rate is the presence of colorectal cancer screening tests, which improved the percentage by 23.0 percentage points, controlling for the independent influence of all other variables in the model. Recall that the impact of colorectal cancer screening also serves to decrease physician process measure rates by 38.7 percentage points and

was the most influential explanatory variable in that previous model. The results from knowledge of colorectal cancer screening are assessed as substantively important but qualitatively in a different light, according to the policy question at hand.

The section of the table on the impact of the various attribution rules shows their relative importance as a category is second only to the minimum patient count requirement for explaining difference in the percentage of physicians with a calculated process measure rate. Three of the four attribution rules are in the top five for relative importance from the 14 variables in the model. The other variable in the attribution rule category (i.e., 2-touch rule) is the seventh most important, not eroding the overall importance of the category to explain variation in the percentage of physicians with a calculated process measure rate. According to the attribution rule used, the percentage of physicians with a calculated process measure rate can fall as much as an average of 18.4 percentage points when using the 50% rule instead of the one-touch rule (i.e., the frame of reference for the category). Even the least most impactful attribution rule (i.e., the 2-touch rule) decreases the percentage of physicians with a calculated process measure rate by an average of 8.8 percentage points, accounting for the independent impact of every other variable in the model. Clearly, the attribution rule that is followed has a substantively large impact on the percentage of physicians with a calculated process measure rate.

The choice of process measures only substantively matters for explaining variation in the percentage of physicians with a calculated process measure rate for colorectal cancer screening. Knowledge of colorectal cancer screening increases the percentage of physicians with a calculated process measure rate by 23.0 percent – again, the second most influential variable in the model. With that exception, the category is assessed as having weak explanatory ability due to it containing three of the four variables that have the lowest explanatory power of any of the 14 variables in the model. Knowledge of the state of origin is of varying policy relevance from the perspective of the range in effects on the percentage of physicians with a calculated process measure rate. From the perspective of explaining the percentage of physicians with a calculated process measure rate, four states had substantively close adjusted percentages of physicians with process measure rates (i.e., Arizona, California, Massachusetts, and, as the reference state, Minnesota). These states were all within 3.3 adjusted percentage points of each other for this measure. Yet, the percentage of physicians with process measure rates for Wisconsin and for Indiana were 8.4 and 14.8 percentage points higher, respectively, than for Minnesota (i.e., the reference state). All of the differences in state rates from the frame of reference (i.e., Minnesota) were statistically significant.

The R^2 for the model was 0.931, with an adjusted R^2 of 0.930. Those exceptionally high values mean that the multiple correlation between the explanatory variables and physician process measure rates was over 0.96, which is an extraordinarily high multiple correlation. These findings are especially important when the range in the percentage of physicians with a calculated process measure rate is considered. These rates range from 1.48 percent for the average rate found under the 50 patient minimum count and 50% attribution rules for breast cancer screening in Minnesota to 68.1 percent with the 10 patient minimum count and 1-touch attribution rule for colorectal cancer screening in Wisconsin. Furthermore, the trivially small reduction in the R^2 of 0.931 to 0.930 through adjusting for the likely non-randomness in the model due to the number of explanatory variables suggests that the accuracy of the results is not due to overfitting the model.

The third, fourth, and fifth most important variables for explaining differences in the percentage of physicians with a process measure rate are clustered within the section for the attribution rule that was followed. Clearly, understandings of the functioning of the attribution rule followed and of the minimum patient count required (i.e., the most influential explanatory variable) are needed to synthesize the policy implications of the present model and analysis.
The effect of an increasing minimum patient count requirement was to reduce the percentage of physicians with a

calculated process measure rate by as much as 26 percentage points over the range of the assessed data, as previously discussed. The manner by which this variable functions is by progressively removing physicians who have at least 10 eligible patients but fewer than 20, then fewer than 30, and then fewer than 50 eligible patients. At a 50 minimum patient count, the physicians with small patient populations are no longer represented.

The functioning of the attribution method to impact the percentage of physicians with a calculated process measure rate has even wider-spread policy ramifications. Only the 50% rule and the maximum frequency rule assign a patient to a single physician. The other three rules can, and most often do, assign patients to multiple physicians. These assignments to multiple physicians function to inflate the size of the physicians' apparent eligible populations. These apparently larger populations then make smaller-population physicians eligible for measures requiring larger minimum patients counts, which they might now have by sharing other physicians' patients.

The impact on degrading the meaning of an individual physician's process measure rates is hard to predict. For example, Physician 1 is Patient A's PCP; both parties would agree. Patient A goes for a timely mammogram but also sees Physician 2 for a wrenched back twice during Physician 1's two-week vacation, as the 'on call' physician for the one-physician practice. Attribution scenarios that attribute one patient to multiple physicians could credit both physicians with the mammogram, even though Patient A only saw Physician 2 twice for a single acute issue. Alternatively, Patient A might not have received a mammogram but would reduce the process measure result for Physician 2 – the 'on-call' physician. Yet, groups of physicians might share patients and utilize the physician whose specialty is closest to the required medical situation at the time. Sharing both the benefits and the disadvantages of that arrangement might be seen as fully appropriate. Neither the current data nor the current understanding of ramifications of both scenarios is sufficiently mature to reach a conclusion on an optimal resolution to this issue.

Although the outcomes of attribution rules that assign one patient to multiple physicians show more physicians with a calculated process measure rate, the effects of these types of rules could include a degrading of the specificity of some individual physician rates and a confounding for the total number of eligible beneficiaries. For example, if 10 patients each saw 3 eligible physicians once and no other physicians during the time period studied, all three physicians would get credit for the 10 patients under the 1-touch attribution rule for 10 eligible patients for the minimum patient count. Summing the number of patients gives us 30 patients, rather than the 10 that exist. Accounting for the complicating relationships to untangle the issue and arrive at the true number of patients through weighting or some other means is a challenging methodological topic. Yet, even if accomplished, the attribution method can still result in the apparent rates of all three physicians being primarily based on the actions or inactions of two other physicians who could easily be unknown to each other. Untangling the policy ramifications intrinsic to selecting an attribution of patient to physician rule, therefore, extends well beyond the bounds of the current project.

Effects on Percentage of Beneficiaries with a Physician Rate

Table 4 presents results for the regression analysis for the percentage of beneficiaries attributed to a physician with a process measure rate. As before, the table shows both the unstandardized and the standardized regression coefficients, along with their significance level in a population of 600 observations. The interpretations of the standardized and the unstandardized regression coefficients along with their p-values remain unchanged from previously discussed, with the exception of the percentage of beneficiaries attributed to a physician with a process measure rate being the target or explained variable (i.e., the dependent variable in the multiple regression analysis).

Table 4. Effects on the Percentage Of Beneficiaries Attributed to a Physician with a Process Measure Rate

Explanatory variables	Unstandardized Coefficients	Standardized Coefficients	P-value	Relative mportance
Patient Count				
Minimum Patient Count	-1.0%	-0.626	< 0.001	1
Attribution Rule				
2-Touch Rule	-13.6%	-0.231	< 0.001	7
30% Rule	-18.6%	-0.317	< 0.001	6
50% Rule	-29.3%	-0.498	< 0.001	2
Maximum Frequency Rule	-23.4%	-0.398	< 0.001	4
Process Measure				
Breast Cancer Screening	-5.3%	-0.090	< 0.001	11
Colorectal Cancer Screening	22.2%	0.377	< 0.001	5
Diabetes - LDL Screening	< -0.01%	< 0.001	0.998	13
Diabetes - Nephropathy Screening	< -0.01%	< 0.001	0.999	14
State				
Arizona	10.7%	0.151	< 0.001	10
California	11.6%	0.184	< 0.001	8
Indiana	25.7%	0.407	< 0.001	3
Massachusetts	5.3%	0.083	< 0.001	12
Wisconsin	10.1%	0.160	< 0.001	9

Notes: **Frames of reference: 1-Touch Rule, Diabetes – HbA1c Testing, and Minnesota**
N = 600 aggregate physician process measure rates
$R^2 = 0.831$, $AdjR^2 = 0.827$
Listed percents are interpreted as percentage points
P-values less than 0.05 are statistically significant

The table shows that, as in the previous model, the minimum patient counts chosen has more impact on the percentage of beneficiaries attributed to a physician with a calculated process measure rate than any of the other 13 variables in the model. Over the 10-patient minimum to the 50-patient minimum rules, the impact on the percentage of eligible beneficiaries attributed to a physician with a calculated process measure rate can vary by an adjusted average of approximately 40 percentage points.

Following the minimum patient count in relative importance, the attribution rule category shows consistently large impacts on the percentage of beneficiaries attributed to a physician with a calculated process measure rate. Compared with the 1-touch rule, the percentage of beneficiaries attributed to a physician with a calculated process measure rate drops by 29.3 percentage points when the 50% rule is applied. Yet, although the 50% rule would correctly sum the total number of beneficiaries accounted for in the study and would also correctly reproduce the process measure rates for single physicians, some of the individual physicians functioning within coordinated groups of physicians could see their process measures decline. Nonetheless, the attribution rules that assign one patient to multiple physicians would include more of the smaller physicians and their patients who would be missed through minimum patient count methodologies. Substantive conclusions on the more appropriate approach depend, therefore, on the policy perspective at hand.

The least influential category, overall, is the choice of process measure selected. Nonetheless and as discussed in the previous results as well, knowledge of colorectal cancer screening strongly influences the percentage of beneficiaries attributed to a physician with a calculated process measure rate, in this case increasing it by 22.2 percentage points. Yet, the diabetes LDL screening and nephropathy measures have practically no discernable effect at all compared with their

HbA1c testing frame of reference and knowledge of breast cancer screening drops the percentage of beneficiaries by only 5.3 percentage points.

The state effect, while not as large as either the minimum patient count or attribution rule effects, is stronger than the process category effect, overall. Although state effects can be as high as 25.7 percentage points (i.e., Indiana compared with Minnesota), the typical state effects were closer to 10 percentage points.

The R^2 for the model was 0.831, with an adjusted R^2 of 0.827. Those high values mean that the multiple correlation between the explanatory variables and physician process measure rates was over 0.90, which is a high multiple correlation. Furthermore, the relatively small reduction in the R^2 of 0.831 to 0.827 through adjusting for the likely non-randomness in the model due to the number of explanatory variables suggests that the accuracy of the results is most likely not due to over-fitting the model. Although these results show that the percentage of beneficiaries attributed to a physician with a process measure rate is less well explained than the previous two models, the explanatory power is still high for data of this nature.

The discussions in the previous section accounting for variability in the percentage of physicians with a calculated process measure rate through the functioning of operationalizing a minimum patient count and choosing an attribution rule are consistent with the effects seen and the relationships explained for the current model. The methodology for counting beneficiaries depends upon how they are attributed to physicians (i.e., one-to-one or one-to-many) in conjunction with the minimum patient count required to produce a physician process measure rate. For this reason, it is both reasonable and expected that the results from the present and the previous models should be somewhat similar. This similarity is discussed in the section that follows, which is a presentation of the results for the effects on a two parameter model for determining reliability – beta-binomial reliability.

Effects on Beta-Binomial Reliability

Table 5 presents results for the regression analysis to explain the variability in average beta-binomial reliability. As before, the table shows both the unstandardized and the standardized regression coefficients, along with their significance level in a population of 600 observations. The interpretations of the unstandardized and the standardized regression coefficients along with their p-values remain unchanged from previously discussed, with the exception of average beta-binomial reliability being the target or explained variable (i.e., the dependent variable in the multiple regression analysis).

Table 5. Effects on Beta-Binomial Reliability

Explanatory variables	Unstandardized Coefficients	Standardized Coefficients	P-value	Relative Importance
Patient Count				
Minimum Patient Count	0.0032	0.392	< 0.001	5
Attribution Rule				
2-Touch Rule	-0.0007	-0.002	0.922	14
30% Rule	0.016	0.055	0.028	13
50% Rule	0.024	0.080	0.001	11
Maximum Frequency Rule	0.018	0.059	0.019	12
Process Measure				
Breast Cancer Screening	0.054	0.182	< 0.001	9
Colorectal Cancer Screening	0.227	0.763	< 0.001	1

Explanatory variables	Unstandardized Coefficients	Standardized Coefficients	P-value	Relative Importance
Diabetes - LDL Screening	0.072	0.241	< 0.001	7
Diabetes - Nephropathy Screening	0.158	0.532	< 0.001	2
State				
Arizona	0.112	0.312	< 0.001	6
California	0.138	0.431	< 0.001	4
Indiana	0.139	0.436	< 0.001	3
Massachusetts	0.068	0.211	< 0.001	8
Wisconsin	0.029	0.090	< 0.001	10

Notes: **Frames of reference: 1-Touch Rule, Diabetes – HbA1c Testing, and Minnesota**
N = 600 aggregate physician process measure rates
R^2 = 0.773, AdjR^2 = 0.767
P-values less than 0.05 are statistically significant

The table shows the minimum acceptable patient count to be the fifth most influential variable of the 14 variables in the model. Over the range of the data tested (i.e., 10- to 50-minimum patient counts), as the minimum patient count increases, the average beta-binomial reliability increases by approximately 0.128 (i.e., 40-patient difference in minimum patient count x 0.0032 = 0.128), which is a substantively important amount. The minimum acceptable patient count, therefore, can have important policy ramifications for the reliability, and therefore acceptability, of the results. The effect of the attribution rule chosen is generally statistically significant and positive (i.e., 2-touch rule being the exception), but the size of these effects is substantively small. All four of the smallest effects for the 14 variables in the model were found for the four variables in this category. The largest effect within the category on the average beta-binomial reliability was seen for the 50% rule. In this case, average beta-binomial reliability increased by .024, which is still substantively small.

The relative importance of the process measures to explaining variation in beta-binomial reliability is seen through the range of the unstandardized coefficients, from 0.054 to 0.227. Even the smallest of the values (i.e., 0.054) is large enough to potentially be of substantive importance. The largest coefficient (i.e., 0.227) is of a sufficient magnitude to suggest potentially important differences in the beta-binomial reliability of the HbA1c tests (i.e., the frame of reference) compared with that of colorectal cancer screening, which is 0.227 higher than the average beta-binomial reliability of HbA1c tests.

With the exception of Wisconsin compared with Minnesota (i.e., the frame of reference), all other states showed substantively important differences from Minnesota's average beta-binomial reliability. The magnitudes of these differences ranged from 0.068 for Massachusetts to 0.139 for Indiana, again excepting Wisconsin. The causes of these differences could be extraordinarily broad. Yet, the differences found would suggest that national systems for physician reporting could be somewhat affected by differences in the beta-binomial reliability of process measures when the calculated reliability is over a given cut-point (e.g., 0.70) in one state but below it in another for a given combination of attribution and minimum patient count rules.

Overall, the chosen attribution method is trivially related to the average beta-binomial reliability of the tested physician process measures. Yet, both the specific process measure chosen and the state generating the data can have important effects on beta-binomial reliability. The reliability of the process measures are impacted by the actual average rate for each measure as part of the mathematics involved. Yet, the impact of differing process measure rates would be accommodated within the regression model through the inclusion of the process measures themselves. Additional research into these relationships, is needed to more fully explain state effects on beta-binomial reliability. Such research

would require data beyond those available in the BQI Project.

Overall Importance of Each Variable Across the Four Models

Table 6 presents the relative importance of each explanatory variable for each of the four models just discussed. The right-most column in the table presents the ranked average relative importance of each explanatory variable across these same four models. Critically important for understanding the results in the table, the relative importance of an explanatory variable across the four models is not necessarily in a consistently positive direction. Furthermore, the same variable can function to an important extent in a positive direction in one model and in a negative direction in other models. Relative importance, therefore, refers to a variable's explanatory ability alone and not to its contribution to a potentially better outcome.

Table 6. Relative Importance of the Explanatory Variables Across Models

Explanatory variables	(1) On Physician Process Measure Rates	(2) On Percentage of Physicians with a Rate	(3) On Percentage of Beneficiaries Assigned to a Physician with a Rate	(4) On Average Beta-Binomial Reliability	Ranked Average Relative Importance Across the Four Models
Patient Count					
Minimum Patient Count	+12	-1	-1	+5	3
Attribution Rule					
2-Touch Rule	+13	-7	-7	-14	13
30% Rule	-10	-5	-6	+13	8
50% Rule	-8	-3	-2	+11	4
Maximum Frequency Rule	-11	-4	-4	+12	6
Process Measure					
Breast Cancer Screening	-3	-12	-11	+9	9
Colorectal Cancer Screening	-1	+2	+5	+1	1
Diabetes – LDL Screening	-5	-13	-13	+7	11
Diabetes – Nephropathy Screening	-2	-14	-14	+2	7
State					
Arizona	-9	+10	+10	+6	10
California	-6	+9	+8	+4	5
Indiana	-4	+6	+3	+3	2
Massachusetts	-7	+11	+12	+8	12
Wisconsin	-14	+8	+9	+10	14

Notes:

Frames of reference: 1-Touch Rule, Diabetes – HbA1c Testing, and Minnesota

N = 600 aggregate physician process measure rates

'+' or '-' indicate the direction of the effect

The table shows that the minimum patient count variable is the most influential variable in two of the four models, the fifth most influential in a third model and not very influential in a fourth. Overall, this variable is the third most influential variable across all four models; yet, critically, the direction of its influence is different across the models. While having very little impact on physician process measure rates, increasing the required minimum patient count strongly and negatively influences the percentage of physicians with a rate and the percentage of beneficiaries attributed to physicians with a process measures rate – more so than any other variable in either model. Yet, this same variable influences beta-binomial reliability in the opposite way – and also relatively strongly being the fifth most influential variable in the model. These results strongly suggest a compensatory effect for the establishment of minimum patient

count rules. Larger required patient counts appear to decrease the percentage of physicians with a rate and the percentage of beneficiaries attributed to physicians with calculated process measure rates while simultaneously improving the beta-binomial reliability of the results.

A discussion of the effects of the attribution rules on beta-binomial reliability cannot escape the degree to which the two models (i.e., the percentage of physicians with a rate and the percentage of beneficiaries assigned to physicians with a rate) have essentially equivalent results. The ranked relative importance of each of the explanatory variables across the two models is identical for two of the variables and differs by one ranking out of 14 for the other two variables. Furthermore, across all 14 variables, the ranked order of relative importance never differs by more than 3 ranking positions and is identical for 6 variables and differs by only a single ranking position for 6 more variables. When accounting for sampling and specification errors inherent to both models, their functioning is practically identical. Only the total explanatory power of the two models substantively differs, with the model for the percentage of physicians with a calculated process measure rate more precisely explained. Yet, the other model explains more than an adequate proportion of the variation in the percentage of beneficiaries attributed to a physician with a calculated process measure rate as well. So, discussions on the functioning of the categories with one model adequately discussed the other model within the accuracy of the data used to generate the results.

On balance, the attribution rule is only modestly related to the variability in physician process measure rates but is much more strongly related to the percentage of physicians with a calculated process measure rate and the percentage of beneficiaries attributed to physicians with a calculated process measure rate and practically unrelated to average beta-binomial reliabilities. Compared with the 1-touch rule, the 50% and maximum frequency rules have the largest overall effect, but that result would be expected through the manifest differences between attributing a single beneficiary to a single physician versus attributing a single beneficiary to multiple physicians on the models of interest to the current project. Overall, the 50% and maximum frequency attribution rules most strongly function to decrease the percentage of physicians with a calculated process measure rate and the percentage of beneficiaries attributed to physicians with a calculated process measure rate, as would be predicted through their computational logic.

As a group, the impact of the selection of a specific process measure can vary but is generally quite strong across the models. First, this finding suggests that important differences in the probability that a given methodology will detect that an eligible patient received an appropriate service. Although almost a trivial finding due to its previous elucidation in a vast number of sources, the effect of the choice of these measures after controlling for the other types of effects in the models has been studied much less often. The outcome of the present work, therefore, is to provide more accurate estimates of the studied effects than have been previously available. This effect size can be as large as 38.7 percentage points in physician process measure rates.

Second, of all of the categories, the attribution rule followed most strongly influences the percentage of physicians with a rate and the percentage of beneficiaries attributed to physicians with a calculated process measure rate. The policy ramifications of the differences in the potential attribution rules are likely to cluster in two camps: those who favor the attribution of a single patient to a single physician versus those who favor the ability to assess coordinated efforts across multiple physicians. The stage of development of this aspect of attribution methodology is at a point where it seems that a middle ground has yet to be, but must ultimately be, found. Both camps present policy-relevant evidence to suggest important reasons for their position. Choosing between these positions is beyond the bounds of the present work. Nonetheless, the present explanation of the differences in the effects on important policy issues can inform both policy and future research in this area.

Third, the importance of the process measure category is almost opposite to that of the attribution category across the four models. Although the attribution category is relatively unimportant to explaining differences in physician rates and beta-binomial reliability, with exceptions previously noted, the opposite is true for the process measure category. The process measure category showed some of the strongest relationships in these two categories.
The impact of the data's state of origin is moderate, overall, across the entire spectrum of models. While rarely one of the most or least influential variables in any of the models, its impact was rarely small. As previously discussed, research to accurately model these adjusted state differences will form an important contribution to understanding the differences that research finds in the provision of health care services.

Limitations

One limitation of multiple regression models is their general lack of interaction terms. Although potentially a valid criticism of simple models, models with the complexity of those presented herein require more robust data (i.e., from larger and more representative populations) to support the fitting of the interaction terms than could be provided within the present framework. As such, the results should be interpreted more as guidance for future work than as data on which to base national policy.

Next Steps

The present work has delineated both the compensatory nature of the models and the specific importance of the explanatory variables with and across each model. In the process, several gaps in present knowledge became apparent. The compensatory nature of the models in combination with the explanatory variables functioning in opposite directions and with very different levels of importance (with the exception of the two models discussed earlier) suggests that one of the most important next steps will be to find attribution models to accommodate advocates of both the single patient to single physician and the single patient to multiple physicians camps. The functional differences highlighted herein suggest ongoing difficulties with national reporting of physician measures until a consensus can be reached on this issue.

Secondly, the state-level differences found in the four models cannot be ignored for a multitude of national policy and measurement theory reasons. The substantive differences that are attributable to the state generating the data after controlling for the other variables in the models are too large, to be ignored. The list of potential reasons for these differences is lengthy and will be difficult to untangle. Yet, by untangling the contributors to the state differences, far more meaningful models of physician rates, percentage of physicians with a rate, percentage of beneficiaries attributed to physicians with rates, and beta-binomial reliabilities could be constructed. Overall, state differences were always substantively important to the four models presented. Through an effective understanding and modeling of those differences, the state effects found would be re-distributed throughout the rest of the variables in the model. This redistribution of effects would yield more meaningful results for many types of stakeholders.

On balance, though, it is the substantive difference in the purpose of each model that determines which of the explanatory variables are most important and to which the most attention should be paid. This issue represents a qualitative discussion and series of decisions at a policy level and not at a level to be well informed by data. As the present analyses have shown, the data can inform aspects of the choices inherent to each of the models but it cannot rate the importance of the policy choices that would result in one or more of the models being more important than the others.

Table A9-1: Calendar Year 2006 Measure Rates and Reliability, By State, 1-Touch Rule

Measure	State	Rate	Denominator	Eligible physicians[1]	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[2]	Percent beneficiaries in denominator w/a physician[3]	Beta binomial reliability
Breast Cancer Screening	AZ	57.5%	52,659	5,848	65.2%	26.7%	71.9%	83.3%	0.63
	CA	54.0%	299,436	34,721	65.4%	25.4%	73.7%	81.6%	0.66
	IN	54.9%	99,112	7,630	61.0%	23.4%	79.4%	89.6%	0.63
	MA	64.5%	86,210	9,197	72.3%	23.4%	71.8%	79.7%	0.50
	MN	65.9%	52,686	7,425	71.5%	23.4%	79.2%	82.4%	0.46
	WI	64.3%	73,122	6,862	68.9%	22.3%	79.0%	83.9%	0.54
Colorectal Cancer Screening	AZ	38.6%	265,562	5,353	47.6%	25.0%	89.6%	82.5%	0.88
	CA	37.1%	1,324,729	31,495	47.2%	25.4%	89.8%	81.7%	0.88
	IN	38.3%	463,760	6,979	45.9%	23.2%	93.0%	88.3%	0.90
	MA	40.8%	381,786	8,450	49.1%	25.5%	88.3%	79.4%	0.80
	MN	41.6%	267,757	6,949	48.9%	21.9%	91.3%	81.3%	0.71
	WI	41.7%	364,606	6,360	48.4%	22.3%	91.7%	82.8%	0.84
HbA1c Testing	AZ	73.7%	48,155	5,202	79.1%	22.6%	74.0%	85.4%	0.49
	CA	77.0%	281,941	30,956	80.3%	21.2%	74.6%	89.1%	0.61
	IN	78.3%	107,790	6,880	79.3%	20.3%	81.0%	93.4%	0.66
	MA	78.7%	74,319	8,239	80.0%	21.7%	71.6%	83.2%	0.47
	MN	85.2%	44,742	6,836	87.0%	16.5%	78.6%	86.2%	0.21
	WI	86.9%	70,320	6,240	87.6%	15.6%	79.9%	89.0%	0.35
LDL Testing	AZ	69.7%	48,155	5,202	75.6%	24.1%	74.0%	85.4%	0.60
	CA	76.0%	281,941	30,956	80.1%	21.2%	74.6%	89.1%	0.59
	IN	72.2%	107,790	6,880	73.8%	21.7%	81.0%	93.4%	0.70
	MA	71.7%	74,319	8,239	73.4%	24.7%	71.6%	83.2%	0.57
	MN	77.5%	44,742	6,836	79.7%	20.4%	78.6%	86.2%	0.37
	WI	79.0%	70,320	6,240	79.4%	19.4%	79.9%	89.0%	0.48
Nephropathy Testing	AZ	62.2%	48,155	5,202	70.8%	24.5%	74.0%	85.4%	0.64
	CA	62.8%	281,941	30,956	69.6%	24.4%	74.6%	89.1%	0.69
	IN	55.7%	107,790	6,880	62.4%	24.4%	81.0%	93.4%	0.78
	MA	64.6%	74,319	8,239	69.4%	25.0%	71.6%	83.2%	0.61
	MN	66.2%	44,742	6,836	72.9%	22.0%	78.6%	86.2%	0.50
	WI	65.0%	70,320	6,240	69.3%	22.4%	79.9%	89.0%	0.67

[1] Eligible physicians is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

[2] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[3] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-2: Calendar Year 2006 Measure Rates and Reliability, By State, 2-Touch Rule

Measure	State	Rate	Denominator	Eligible physicians[1]	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[2]	Percent beneficiaries in denominator w/a physician[3]	Beta binomial reliability
Breast Cancer Screening	AZ	57.5%	52,659	5,848	66.9%	26.6%	63.5%	77.0%	0.59
	CA	54.0%	299,436	34,721	66.4%	26.0%	65.1%	74.5%	0.61
	IN	54.9%	99,112	7,630	62.4%	23.8%	70.9%	83.5%	0.59
	MA	64.5%	86,210	9,197	73.7%	23.8%	61.6%	70.1%	0.43
	MN	65.9%	52,686	7,425	72.5%	25.2%	68.1%	72.1%	0.41
	WI	64.3%	73,122	6,862	69.8%	23.3%	69.9%	74.4%	0.49
Colorectal Cancer Screening	AZ	38.6%	265,562	5,353	48.0%	25.5%	80.3%	77.0%	0.86
	CA	37.1%	1,324,729	31,495	47.7%	25.4%	79.3%	75.6%	0.86
	IN	38.3%	463,760	6,979	46.3%	23.5%	83.3%	82.9%	0.88
	MA	40.8%	381,786	8,450	49.3%	25.3%	74.7%	71.3%	0.76
	MN	41.6%	267,757	6,949	49.4%	23.4%	80.1%	71.7%	0.65
	WI	41.7%	364,606	6,360	49.0%	23.0%	81.3%	74.2%	0.80
HbA1c Testing	AZ	73.7%	48,155	5,202	80.5%	22.3%	66.6%	81.1%	0.46
	CA	77.0%	281,941	30,956	81.1%	21.2%	66.3%	84.1%	0.58
	IN	78.3%	107,790	6,880	80.3%	20.0%	72.4%	89.9%	0.66
	MA	78.7%	74,319	8,239	81.1%	21.5%	61.8%	76.2%	0.45
	MN	85.2%	44,742	6,836	87.7%	18.4%	68.0%	79.0%	0.20
	WI	86.9%	70,320	6,240	88.4%	15.7%	71.0%	82.9%	0.34
LDL Testing	AZ	69.7%	48,155	5,202	77.0%	24.1%	66.6%	81.1%	0.56
	CA	76.0%	281,941	30,956	81.0%	21.0%	66.3%	84.1%	0.56
	IN	72.2%	107,790	6,880	75.1%	21.4%	72.4%	89.9%	0.70
	MA	71.7%	74,319	8,239	74.3%	25.1%	61.8%	76.2%	0.55
	MN	77.5%	44,742	6,836	81.0%	21.8%	68.0%	79.0%	0.35
	WI	79.0%	70,320	6,240	80.5%	19.7%	71.0%	82.9%	0.48
Nephropathy Testing	AZ	62.2%	48,155	5,202	71.3%	25.0%	66.6%	81.1%	0.64
	CA	62.8%	281,941	30,956	69.6%	25.0%	66.3%	8.4%	0.67
	IN	55.7%	107,790	6,880	62.3%	24.9%	72.4%	89.9%	0.78
	MA	64.6%	74,319	8,239	70.0%	25.5%	61.8%	76.2%	0.60
	MN	66.2%	44,742	6,836	73.3%	24.4%	68.0%	79.0%	0.49
	WI	65.0%	70,320	6,240	69.4%	23.9%	71.0%	82.9%	0.66

[1] Eligible physicians is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

[2] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[3] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-3: Calendar Year 2006 Measure Rates and Reliability, By State, 30% Rule

Measure	State	Rate	Denominator	Eligible physicians[1]	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[2]	Percent beneficiaries in denominator w/a physician[3]	Beta binomial reliability
Breast Cancer Screening	AZ	57.5%	52,659	5,848	59.5%	29.3%	56.7%	82.2%	0.59
	CA	54.0%	299,436	34,721	59.4%	29.0%	59.3%	79.8%	0.56
	IN	54.9%	99,112	7,630	56.2%	26.5%	64.9%	88.6%	0.58
	MA	64.5%	86,210	9,197	68.8%	26.6%	60.6%	79.1%	0.44
	MN	65.9%	52,686	7,425	68.1%	28.1%	64.8%	80.4%	0.42
	WI	64.3%	73,122	6,862	66.6%	25.2%	68.6%	82.9%	0.49
Colorectal Cancer Screening	AZ	38.6%	265,562	5,353	38.9%	26.1%	73.7%	81.9%	0.84
	CA	37.1%	1,324,729	31,495	39.1%	25.9%	73.6%	80.5%	0.84
	IN	38.3%	463,760	6,979	38.3%	23.8%	77.8%	87.7%	0.86
	MA	40.8%	381,786	8,450	44.4%	25.9%	74.7%	79.1%	0.75
	MN	41.6%	267,757	6,949	43.5%	24.1%	78.7%	80.3%	0.67
	WI	41.7%	364,606	6,360	43.1%	23.7%	80.8%	82.2%	0.79
HbA1c Testing	AZ	73.7%	48,155	5,202	74.2%	27.8%	59.0%	84.6%	0.54
	CA	77.0%	281,941	30,956	76.8%	25.4%	59.2%	87.4%	0.59
	IN	78.3%	107,790	6,880	76.3%	23.8%	64.7%	92.4%	0.68
	MA	78.7%	74,319	8,239	79.0%	23.3%	59.8%	82.7%	0.47
	MN	85.2%	44,742	6,836	84.9%	22.1%	63.5%	84.8%	0.24
	WI	86.9%	70,320	6,240	85.6%	19.7%	67.6%	88.2%	0.35
LDL Testing	AZ	69.7%	48,155	5,202	69.6%	29.5%	59.0%	84.6%	0.62
	CA	76.0%	281,941	30,956	75.7%	25.8%	59.2%	87.4%	0.59
	IN	72.2%	107,790	6,880	70.7%	24.9%	64.7%	92.4%	0.70
	MA	71.7%	74,319	8,239	71.7%	26.8%	59.8%	82.7%	0.56
	MN	77.5%	44,742	6,836	77.7%	25.5%	63.5%	84.8%	0.37
	WI	79.0%	70,320	6,240	77.4%	23.0%	67.6%	88.2%	0.50
Nephropathy Testing	AZ	62.2%	48,155	5,202	65.2%	28.8%	59.0%	84.6%	0.64
	CA	62.8%	281,941	30,956	63.8%	28.5%	59.2%	87.4%	0.66
	IN	55.7%	107,790	6,880	56.4%	27.0%	64.7%	92.4%	0.76
	MA	64.6%	74,319	8,239	66.9%	27.3%	59.8%	82.7%	0.58
	MN	66.2%	44,742	6,836	68.7%	28.0%	63.5%	84.8%	0.48
	WI	65.0%	70,320	6,240	66.2%	26.1%	67.6%	88.2%	0.65

[1] Eligible physicians is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

[2] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[3] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-4: Calendar Year 2006 Measure Rates and Reliability, By State, 50% Rule

Measure	State	Rate	Denominator	Eligible physicians[1]	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[2]	Percent beneficiaries in denominator w/a physician[3]	Beta binomial reliability
Breast Cancer Screening	AZ	57.5%	52,659	5,848	57.5%	31.3%	51.0%	74.5%	0.58
	CA	54.0%	299,436	34,721	57.4%	30.9%	52.7%	69.8%	0.53
	IN	54.9%	99,112	7,630	54.3%	28.6%	58.9%	81.1%	0.57
	MA	64.5%	86,210	9,197	68.1%	28.1%	54.6%	73.9%	0.42
	MN	65.9%	52,686	7,425	66.9%	30.4%	57.4%	71.8%	0.41
	WI	64.3%	73,122	6,862	65.9%	27.3%	63.0%	76.3%	0.47
Colorectal Cancer Screening	AZ	38.6%	265,562	5,353	36.2%	26.5%	66.8%	75.9%	0.83
	CA	37.1%	1,324,729	31,495	36.5%	25.9%	66.5%	72.6%	0.82
	IN	38.3%	463,760	6,979	36.0%	23.9%	70.8%	81.9%	0.86
	MA	40.8%	381,786	8,450	43.0%	25.9%	68.6%	75.7%	0.73
	MN	41.6%	267,757	6,949	42.2%	25.2%	72.6%	74.2%	0.66
	WI	41.7%	364,606	6,360	41.3%	23.5%	74.9%	77.8%	0.77
HbA1c Testing	AZ	73.7%	48,155	5,202	72.9%	29.9%	52.9%	77.4%	0.55
	CA	77.0%	281,941	30,956	76.0%	27.1%	52.6%	77.4%	0.59
	IN	78.3%	107,790	6,880	75.5%	25.6%	58.6%	84.5%	0.69
	MA	78.7%	74,319	8,239	78.8%	24.4%	54.8%	78.4%	0.46
	MN	85.2%	44,742	6,836	85.0%	23.5%	56.5%	76.9%	0.24
	WI	86.9%	70,320	6,240	85.3%	21.2%	62.2%	82.3%	0.35
LDL Testing	AZ	69.7%	48,155	5,202	68.2%	31.6%	52.9%	77.4%	0.63
	CA	76.0%	281,941	30,956	74.8%	27.4%	52.7%	77.4%	0.57
	IN	72.2%	107,790	6,880	69.8%	26.7%	58.7%	84.5%	0.70
	MA	71.7%	74,319	8,239	71.9%	27.5%	54.7%	78.4%	0.55
	MN	77.5%	44,742	6,836	77.7%	27.0%	56.5%	76.9%	0.37
	WI	79.0%	70,320	6,240	77.0%	24.7%	62.2%	82.3%	0.50
Nephropathy Testing	AZ	62.2%	48,155	5,202	63.9%	30.6%	53.1%	77.4%	0.64
	CA	62.8%	281,941	30,956	61.9%	30.3%	52.6%	77.4%	0.65
	IN	55.7%	107,790	6,880	55.2%	28.3%	58.7%	84.5%	0.76
	MA	64.6%	74,319	8,239	66.2%	28.6%	54.8%	78.4%	0.57
	MN	66.2%	44,742	6,836	67.7%	30.0%	56.3%	76.9%	0.49
	WI	65.0%	70,320	6,240	65.1%	28.0%	62.3%	82.3%	0.65

[1] Eligible physicians is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

[2] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[3] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-5: Calendar Year 2006 Measure Rates and Reliability, By State, Maximum Frequency Rule

Measure	State	Rate	Denominator	Eligible physicians[1]	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[2]	Percent beneficiaries in denominator w/a physician[3]	Beta binomial reliability
Breast Cancer Screening	AZ	57.5%	52,659	5,848	59.0%	30.3%	53.4%	83.3%	0.58
	CA	54.0%	299,436	34,721	59.3%	29.9%	55.7%	81.6%	0.55
	IN	54.9%	99,112	7,630	55.4%	27.6%	61.1%	89.6%	0.57
	MA	64.5%	86,210	9,197	68.6%	27.9%	56.7%	79.7%	0.42
	MN	65.9%	52,686	7,425	68.0%	28.9%	61.2%	82.4%	0.41
	WI	64.3%	73,122	6,862	66.2%	26.6%	65.1%	83.9%	0.48
Colorectal Cancer Screening	AZ	38.6%	265,562	5,353	38.5%	26.3%	69.5%	82.5%	0.83
	CA	37.1%	1,324,729	31,495	39.0%	25.9%	69.0%	81.7%	0.82
	IN	38.3%	463,760	6,979	37.6%	23.7%	73.0%	88.3%	0.85
	MA	40.8%	381,786	8,450	44.1%	25.9%	70.2%	79.4%	0.73
	MN	41.6%	267,757	6,949	43.6%	24.5%	75.4%	81.3%	0.66
	WI	41.7%	364,606	6,360	42.3%	23.3%	76.5%	82.8%	0.77
HbA1c Testing	AZ	73.7%	48,155	5,202	73.7%	28.9%	55.5%	85.4%	0.55
	CA	77.0%	281,941	30,956	86.8%	25.9%	55.5%	89.1%	0.59
	IN	78.3%	107,790	6,880	76.0%	24.6%	60.8%	93.4%	0.69
	MA	78.7%	74,319	8,239	78.8%	24.2%	56.3%	83.2%	0.46
	MN	85.2%	44,742	6,836	85.2%	22.4%	60.1%	86.2%	0.24
	WI	86.9%	70,320	6,240	85.3%	20.9%	64.1%	89.0%	0.35
LDL Testing	AZ	69.7%	48,155	5,202	69.3%	30.3%	55.5%	85.4%	0.63
	CA	76.0%	281,941	30,956	75.9%	26.1%	55.5%	89.1%	0.57
	IN	72.2%	107,790	6,880	70.2%	25.6%	60.8%	93.4%	0.71
	MA	71.7%	74,319	8,239	71.5%	27.7%	56.3%	83.2%	0.56
	MN	77.5%	44,742	6,836	77.9%	26.2%	60.1%	86.2%	0.37
	WI	79.0%	70,320	6,240	77.0%	24.2%	64.1%	89.0%	0.50
Nephropathy Testing	AZ	62.2%	48,155	5,202	65.3%	29.4%	55.5%	85.4%	0.64
	CA	62.8%	281,941	30,956	64.0%	28.9%	55.5%	89.1%	0.65
	IN	55.7%	107,790	6,880	56.8%	27.4%	60.8%	93.4%	0.76
	MA	64.6%	74,319	8,239	66.9%	28.1%	56.3%	83.2%	0.57
	MN	66.2%	44,742	6,836	68.9%	28.8%	60.1%	86.2%	0.48
	WI	65.0%	70,320	6,240	65.8%	27.3%	64.1%	89.0%	0.65

[1] Eligible physicians is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

[2] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[3] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-6: Calendar Year 2006 Breast Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, 1-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Breast Cancer Screening	AZ	10	39.3	67.4%	15.4%	43.2%	82.1%	0.73	0.41
		20	48.3	68.6%	13.7%	31.6%	78.5%	0.76	0.56
		30	58.5	69.8%	13.0%	22.3%	72.1%	0.79	0.66
		50	78.4	70.9%	12.6%	11.2%	56.3%	0.84	0.77
	CA	10	39.2	65.5%	15.8%	46.5%	80.5%	0.74	0.42
		20	48.5	65.7%	15.0%	33.9%	77.2%	0.78	0.59
		30	58.7	65.8%	14.6%	23.9%	71.4%	0.81	0.69
		50	81.2	65.3%	14.1%	11.4%	56.3%	0.86	0.79
	IN	10	48.5	61.8%	13.7%	55.6%	89.2%	0.70	0.33
		20	54.2	61.9%	12.9%	47.7%	88.0%	0.72	0.49
		30	61.2	61.8%	12.3%	38.6%	84.7%	0.74	0.58
		50	78.3	61.6%	11.5%	21.7%	69.8%	0.77	0.68
	MA	10	35.5	72.4%	12.9%	42.9%	77.2%	0.60	0.30
		20	44.0	72.8%	11.8%	30.7%	72.2%	0.65	0.46
		30	53.0	73.0%	11.1%	21.0%	64.1%	0.68	0.55
		50	74.3	73.8%	10.0%	8.7%	42.2%	0.74	0.65
	MN	10	30.1	72.1%	13.2%	48.1%	80.0%	0.56	0.30
		20	37.8	71.8%	12.2%	32.4%	74.3%	0.62	0.46
		30	46.4	72.3%	11.2%	19.8%	63.6%	0.65	0.54
		50	65.0	72.4%	10.0%	6.2%	37.0%	0.69	0.63
	WI	10	35.9	70.1%	13.3%	55.7%	82.5%	0.62	0.31
		20	42.6	70.3%	12.1%	42.4%	78.5%	0.64	0.46
		30	50.4	70.5%	11.5%	29.6%	70.1%	0.68	0.55
		50	69.1	71.4%	10.0%	11.7%	46.0%	0.70	0.63

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-7: Calendar Year 2006 Breast Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, 2-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Breast Cancer Screening	AZ	10	31.8	68.6%	15.6%	35.8%	74.8%	0.69	0.42
		20	41.4	70.1%	14.1%	23.1%	67.5%	0.74	0.58
		30	51.7	71.0%	13.2%	14.4%	57.1%	0.77	0.66
		50	73.9	72.1%	12.7%	5.5%	34.8%	0.83	0.77
	CA	10	30.6	66.7%	16.5%	37.3%	72.1%	0.71	0.44
		20	40.7	66.8%	15.5%	23.1%	64.3%	0.77	0.62
		30	51.6	66.7%	14.9%	14.0%	53.8%	0.80	0.70
		50	74.7	65.6%	13.8%	5.2%	33.4%	0.84	0.78
	IN	10	37	62.7%	14.8%	48.7%	82.7%	0.67	0.36
		20	43.9	62.3%	13.8%	37.3%	79.3%	0.70	0.52
		30	51.9	62.3%	13.1%	26.4%	71.0%	0.73	0.61
		50	71.5	61.8%	12.3%	10.8%	45.1%	0.78	0.71
	MA	10	27.8	72.9%	13.6%	33.0%	66.2%	0.55	0.31
		20	37.2	72.9%	12.4%	19.5%	56.4%	0.62	0.47
		30	46.5	73.5%	11.3%	11.3%	43.4%	0.66	0.55
		50	67.7	73.5%	9.8%	3.3%	20.2%	0.70	0.63
	MN	10	22.1	73.0%	14.4%	32.1%	66.3%	0.55	0.35
		20	30.4	73.9%	12.7%	15.6%	50.2%	0.61	0.51
		30	39.7	74.4%	12.1%	6.3%	29.5%	0.68	0.61
		50	59.8	72.3%	11.1%	0.9%	6.8%	0.73	0.69
	WI	10	26.2	70.9%	14.5%	43.9%	71.5%	0.58	0.35
		20	34.1	71.6%	13.0%	26.3%	60.6%	0.63	0.50
		30	43.7	72.2%	12.2%	13.5%	43.3%	0.68	0.59
		50	64	73.3%	10.3%	3.2%	16.5%	0.71	0.66

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-8: Calendar Year 2006 Breast Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, 30% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Breast Cancer Screening	AZ	10	28.0	64.3%	17.4%	27.7%	75.9%	0.71	0.46
		20	38.0	66.6%	15.5%	16.1%	62.0%	0.76	0.62
		30	47.9	67.4%	14.8%	9.4%	46.8%	0.79	0.71
		50	69.7	69.0%	14.3%	3.1%	22.9%	0.85	0.81
	CA	10	25.4	62.0%	18.1%	28.0%	72.0%	0.69	0.47
		20	36.0	62.6%	16.6%	14.5%	55.1%	0.75	0.63
		30	46.8	62.5%	15.6%	7.7%	39.0%	0.79	0.71
		50	69.6	61.2%	14.8%	2.2%	17.4%	0.84	0.79
	IN	10	32.1	58.6%	15.9%	39.3%	85.5%	0.67	0.39
		20	39.6	59.5%	14.7%	27.6%	77.2%	0.71	0.55
		30	47.6	59.8%	14.1%	18.2%	62.5%	0.75	0.65
		50	66.1	59.5%	14.1%	6.1%	30.0%	0.82	0.77
	MA	10	26.1	69.9%	14.6%	28.9%	72.1%	0.57	0.34
		20	35.2	70.0%	13.3%	16.5%	57.5%	0.64	0.51
		30	44.9	70.9%	12.2%	8.8%	39.8%	0.68	0.59
		50	64.5	71.2%	11.1%	2.4%	15.6%	0.74	0.69
	MN	10	20.5	70.6%	15.7%	26.1%	67.8%	0.57	0.39
		20	29.0	72.6%	13.6%	11.2%	43.6%	0.63	0.54
		30	38.1	73.1%	12.6%	4.1%	21.3%	0.69	0.63
		50	57.3	71.2%	12.3%	0.4%	2.9%	0.75	0.72
	WI	10	24.2	68.7%	15.5%	38.0%	76.9%	0.59	0.37
		20	32.5	69.4%	13.8%	21.0%	59.6%	0.64	0.53
		30	41.5	70.2%	12.7%	10.2%	38.2%	0.69	0.61
		50	61.0	72.5%	11.5%	1.9%	10.5%	0.75	0.71

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-9: Calendar Year 2006 Breast Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, 50% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Breast Cancer Screening	AZ	10	25.6	63.8%	18.0%	22.5%	64.0%	0.70	0.48
		20	35.6	66.9%	15.9%	12.0%	47.6%	0.76	0.64
		30	45.9	68.3%	14.8%	6.4%	32.5%	0.79	0.71
		50	66.5	69.8%	15.4%	1.9%	14.1%	0.87	0.83
	CA	10	23.1	61.2%	18.8%	21.0%	56.3%	0.69	0.49
		20	34.0	62.3%	17.1%	9.6%	38.0%	0.76	0.65
		30	44.6	62.2%	16.3%	4.7%	24.1%	0.80	0.73
		50	67.8	59.7%	15.4%	1.1%	9.0%	0.85	0.81
	IN	10	30.0	58.4%	16.5%	32.1%	74.2%	0.68	0.42
		20	36.9	59.6%	15.3%	22.3%	63.3%	0.72	0.58
		30	45.0	60.1%	14.9%	13.6%	47.0%	0.76	0.68
		50	64.5	59.0%	15.5%	3.6%	17.8%	0.85	0.81
	MA	10	24.8	69.4%	15.0%	23.5%	62.3%	0.57	0.35
		20	33.3	70.1%	13.7%	13.1%	46.6%	0.64	0.52
		30	42.8	70.7%	12.4%	6.5%	29.7%	0.67	0.59
		50	64.4	71.8%	10.3%	1.3%	8.8%	0.71	0.65
	MN	10	18.9	71.5%	16.1%	20.1%	53.5%	0.57	0.41
		20	27.5	74.0%	13.9%	7.5%	29.1%	0.65	0.57
		30	36.8	74.1%	13.3%	2.2%	11.2%	0.72	0.68
		50	54.8	72.3%	16.0%	0.1%	1.1%	0.82	0.81
	WI	10	22.5	68.9%	15.9%	30.8%	65.2%	0.59	0.39
		20	30.7	70.0%	14.3%	15.7%	45.2%	0.66	0.56
		30	40.0	71.1%	13.3%	6.7%	25.0%	0.71	0.64
		50	60.1	73.4%	11.8%	0.8%	4.8%	0.76	0.73

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-10: Calendar Year 2006 Breast Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, Maximum Frequency Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Breast Cancer Screening	AZ	10	26.7	64.4%	17.9%	24.4%	72.2%	0.71	0.48
		20	36.4	67.1%	15.7%	13.7%	55.5%	0.76	0.63
		30	47.1	68.5%	14.9%	7.4%	38.4%	0.80	0.71
		50	68.3	70.1%	14.8%	2.3%	17.3%	0.86	0.82
	CA	10	24.1	62.0%	18.5%	24.3%	68.0%	0.69	0.48
		20	34.9	62.7%	16.8%	11.8%	47.6%	0.76	0.64
		30	45.5	62.9%	15.9%	6.0%	31.5%	0.80	0.72
		50	68.0	60.8%	15.3%	1.6%	12.6%	0.85	0.80
	IN	10	30.8	58.6%	16.0%	34.9%	82.6%	0.67	0.40
		20	37.9	59.8%	14.9%	24.4%	71.1%	0.71	0.57
		30	45.6	60.3%	14.4%	15.5%	54.5%	0.75	0.66
		50	64.9	59.2%	14.8%	4.4%	21.9%	0.83	0.79
	MA	10	25.2	69.9%	14.8%	25.3%	68.0%	0.57	0.34
		20	34.0	70.2%	13.3%	14.1%	51.4%	0.64	0.51
		30	43.4	71.2%	12.1%	7.3%	33.6%	0.67	0.58
		50	64.6	71.7%	10.5%	1.6%	10.7%	0.72	0.66
	MN	10	19.6	70.9%	16.0%	22.8%	62.8%	0.57	0.41
		20	28.4	73.2%	14.2%	8.9%	35.6%	0.66	0.58
		30	37.4	74.9%	12.8%	2.9%	15.1%	0.70	0.65
		50	55.9	72.6%	14.2%	0.2%	1.8%	0.80	0.78
	WI	10	23.3	68.7%	15.7%	33.2%	72.6%	0.59	0.39
		20	31.5	69.5%	14.1%	17.6%	51.9%	0.65	0.55
		30	40.5	70.7%	13.1%	8.0%	30.3%	0.70	0.63
		50	60.0	73.3%	11.5%	1.2%	6.6%	0.75	0.72

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-11: Calendar Year 2006 Colorectal Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, 1-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Colorectal Cancer Screening	AZ	10	154.0	48.3%	14.8%	58.2%	82.3%	0.92	0.43
		20	173.0	48.5%	14.3%	51.4%	82.1%	0.93	0.60
		30	185.0	48.7%	14.3%	47.4%	81.8%	0.93	0.70
		50	207.0	49.2%	14.2%	41.2%	80.9%	0.94	0.80
	CA	10	138.0	47.2%	15.0%	60.5%	81.5%	0.92	0.44
		20	153.0	47.2%	14.7%	54.1%	81.3%	0.92	0.61
		30	163.0	47.3%	14.6%	50.0%	81.0%	0.93	0.71
		50	183.0	47.7%	14.6%	43.1%	79.9%	0.94	0.80
	IN	10	191.0	46.4%	13.4%	66.2%	88.2%	0.92	0.39
		20	208.0	46.4%	13.2%	60.2%	88.1%	0.93	0.56
		30	219.0	46.3%	13.1%	57.1%	88.0%	0.96	0.66
		50	234.0	46.4%	13.1%	52.7%	87.7%	0.94	0.76
	MA	10	119.0	49.3%	13.4%	57.0%	79.0%	0.86	0.35
		20	135.0	49.3%	12.7%	49.6%	78.3%	0.88	0.52
		30	146.0	49.3%	12.5%	44.9%	77.5%	0.89	0.62
		50	166.0	49.2%	12.2%	37.9%	75.7%	0.90	0.73
	MN	10	108.0	49.0%	12.0%	66.5%	81.0%	0.77	0.24
		20	122.0	49.1%	11.0%	58.3%	80.6%	0.79	0.38
		30	132.0	48.9%	10.3%	52.6%	79.9%	0.80	0.47
		50	150.0	48.6%	9.7%	44.0%	78.3%	0.81	0.59
	WI	10	142.0	48.7%	12.6%	68.1%	82.6%	0.87	0.33
		20	154.0	48.8%	12.1%	62.4%	82.3%	0.88	0.49
		30	162.0	49.0%	11.7%	58.5%	81.9%	0.89	0.59
		50	176.0	49.1%	11.5%	52.7%	81.0%	0.89	0.70

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-12: Calendar Year 2006 Colorectal Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, 2-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Colorectal Cancer Screening	AZ	10	121.0	48.8%	15.1%	52.0%	76.7%	0.90	0.43
		20	137.0	49.0%	14.8%	45.5%	76.4%	0.91	0.60
		30	147.0	49.3%	14.7%	41.8%	75.9%	0.92	0.70
		50	170.0	49.6%	14.4%	34.3%	74.0%	0.93	0.79
	CA	10	105.0	47.7%	15.9%	53.8%	75.3%	0.90	0.46
		20	118.0	47.9%	15.5%	47.3%	74.9%	0.91	0.63
		30	129.0	48.2%	15.4%	42.4%	74.1%	0.92	0.72
		50	148.0	48.7%	15.3%	34.8%	71.9%	0.93	0.82
	IN	10	149.0	46.8%	14.3%	59.5%	82.7%	0.91	0.40
		20	162.0	46.6%	13.9%	54.5%	82.5%	0.92	0.57
		30	170.0	46.5%	13.7%	51.4%	82.3%	0.92	0.67
		50	184.0	46.4%	13.4%	46.3%	81.6%	0.92	0.77
	MA	10	94.2	49.6%	13.9%	48.5%	70.6%	0.83	0.34
		20	108.0	49.2%	13.1%	41.4%	69.7%	0.85	0.51
		30	119.0	49.0%	12.6%	36.7%	68.6%	0.86	0.60
		50	138.0	48.6%	12.0%	29.6%	65.7%	0.87	0.71
	MN	10	75.3	49.4%	12.9%	54.9%	71.1%	0.73	0.26
		20	88.2	49.1%	11.6%	45.4%	70.1%	0.75	0.41
		30	99.3	48.8%	10.8%	38.7%	68.5%	0.77	0.50
		50	117.0	49.0%	10.2%	29.9%	64.9%	0.79	0.62
	WI	10	103.0	49.5%	13.4%	59.6%	73.9%	0.84	0.34
		20	113.0	49.3%	12.6%	53.5%	73.4%	0.85	0.50
		30	121.0	49.3%	12.3%	49.1%	72.8%	0.86	0.60
		50	138.0	49.3%	11.9%	40.7%	70.4%	0.87	0.71

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-13: Calendar Year 2006 Colorectal Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, 30% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Colorectal Cancer Screening	AZ	10	99.9	42.2%	16.4%	46.7%	81.2%	0.89	0.45
		20	116.0	42.4%	15.3%	39.5%	80.2%	0.90	0.61
		30	129.0	42.5%	14.5%	34.4%	78.7%	0.91	0.69
		50	155.0	43.5%	14.2%	26.9%	74.7%	0.92	0.79
	CA	10	81.1	41.3%	17.1%	48.7%	79.8%	0.88	0.48
		20	93.1	41.9%	16.3%	41.3%	78.5%	0.89	0.65
		30	105.0	42.4%	15.9%	35.3%	76.4%	0.90	0.73
		50	126.0	43.1%	15.3%	26.6%	70.9%	0.92	0.81
	IN	10	123.0	40.5%	14.8%	54.4%	87.3%	0.89	0.41
		20	135.0	40.7%	13.9%	48.9%	86.9%	0.90	0.57
		30	145.0	40.6%	13.7%	45.0%	86.2%	0.91	0.67
		50	160.0	41.1%	13.4%	39.3%	84.3%	0.91	0.77
	MA	10	83.4	45.7%	14.6%	48.1%	78.2%	0.82	0.35
		20	98.0	45.4%	13.2%	39.8%	76.8%	0.83	0.51
		30	110.0	45.2%	12.5%	34.1%	74.7%	0.85	0.60
		50	130.0	45.0%	12.0%	26.7%	70.3%	0.86	0.71
	MN	10	68.8	44.2%	13.4%	50.9%	79.0%	0.75	0.30
		20	82.1	44.5%	12.2%	41.0%	76.9%	0.78	0.46
		30	93.6	45.0%	11.7%	34.2%	74.2%	0.80	0.56
		50	113.0	45.5%	11.3%	25.4%	67.9%	0.83	0.68
	WI	10	91.7	44.5%	14.3%	58.6%	81.7%	0.84	0.36
		20	103.0	44.8%	13.2%	51.2%	80.8%	0.85	0.52
		30	113.0	45.1%	12.7%	45.5%	79.4%	0.85	0.61
		50	131.0	45.7%	11.9%	36.6%	75.4%	0.86	0.71

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-14: Calendar Year 2006 Colorectal Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, 50% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Colorectal Cancer Screening	AZ	10	90.8	40.1%	16.3%	40.6%	74.3%	0.89	0.46
		20	107.0	41.0%	15.6%	33.4%	72.3%	0.90	0.63
		30	121.0	41.9%	15.3%	28.6%	69.9%	0.91	0.72
		50	145.0	43.5%	15.1%	22.2%	64.9%	0.93	0.81
	CA	10	69.9	39.6%	17.6%	42.6%	70.8%	0.87	0.50
		20	83.9	40.6%	16.7%	34.1%	68.0%	0.89	0.66
		30	95.8	41.3%	16.3%	28.4%	64.7%	0.90	0.74
		50	117.0	42.2%	15.8%	20.6%	57.5%	0.91	0.82
	IN	10	111.0	38.7%	15.2%	48.4%	80.8%	0.89	0.43
		20	124.0	39.4%	14.7%	42.6%	79.6%	0.90	0.60
		30	135.0	40.1%	14.2%	38.5%	78.1%	0.91	0.69
		50	151.0	40.9%	14.0%	33.0%	75.0%	0.92	0.78
	MA	10	79.0	44.8%	14.4%	42.2%	73.8%	0.81	0.35
		20	94.1	44.8%	13.3%	34.2%	71.2%	0.83	0.52
		30	107.0	44.9%	12.8%	29.0%	68.5%	0.85	0.61
		50	126.0	44.9%	12.2%	22.5%	62.9%	0.86	0.72
	MN	10	63.1	43.8%	14.0%	43.7%	71.6%	0.79	0.49
		20	77.7	44.7%	12.7%	33.7%	67.9%	0.81	0.59
		30	88.4	45.0%	12.3%	28.1%	64.4%	0.84	0.71
		50	107.0	46.0%	12.1%	20.5%	56.9%	0.75	0.31
	WI	10	83.6	43.5%	14.5%	52.4%	76.4%	0.82	0.36
		20	95.9	44.3%	13.2%	44.5%	74.5%	0.84	0.52
		30	106.0	44.9%	12.7%	39.0%	72.2%	0.85	0.61
		50	125.0	45.6%	12.1%	30.4%	66.3%	0.86	0.71

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-15: Calendar Year 2006 Colorectal Cancer Screening Measure Rates and Reliability, by State and Selected Denominator, Maximum Frequency Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Colorectal Cancer Screening	AZ	10	93.6	41.5%	16.0%	42.9%	80.9%	0.88	0.44
		20	110.0	42.0%	15.0%	35.5%	78.8%	0.9	0.61
		30	124.0	42.6%	14.9%	30.5%	76.4%	0.91	0.7
		50	149.0	44.0%	14.6%	23.5%	70.9%	0.92	0.8
	CA	10	74.0	41.3%	17.1%	45.4%	79.9%	0.87	0.48
		20	86.7	41.7%	16.3%	37.5%	77.2%	0.89	0.64
		30	98.9	42.4%	15.8%	31.3%	73.7%	0.9	0.73
		50	121.0	43.2%	15.5%	22.9%	65.9%	0.91	0.81
	IN	10	114.0	40.0%	14.8%	50.8%	87.3%	0.89	0.41
		20	128.0	40.4%	14.2%	44.7%	86.0%	0.9	0.58
		30	137.0	40.8%	13.9%	41.0%	84.7%	0.9	0.67
		50	154.0	41.5%	13.6%	34.9%	81.1%	0.91	0.77
	MA	10	79.7	45.3%	14.2%	43.9%	77.5%	0.8	0.34
		20	94.6	45.1%	13.1%	35.8%	75.0%	0.83	0.51
		30	107.0	45.2%	12.6%	30.3%	72.1%	0.84	0.6
		50	127.0	45.2%	12.0%	23.5%	66.1%	0.86	0.71
	MN	10	64.5	44.3%	13.8%	38.7%	78.7%	0.75	0.31
		20	78.8	45.0%	12.3%	30.1%	74.8%	0.78	0.47
		30	89.6	45.3%	12.0%	25.1%	71.0%	0.8	0.57
		50	108.0	46.1%	11.7%	18.4%	62.9%	0.83	0.7
	WI	10	85.5	44.2%	14.2%	41.1%	81.5%	0.82	0.35
		20	97.6	44.8%	13.0%	35.2%	79.6%	0.83	0.5
		30	108.0	45.3%	12.5%	30.8%	77.1%	0.84	0.6
		50	126.0	46.0%	12.0%	24.4%	71.3%	0.86	0.71

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-16: Calendar Year 2006 HbA1c Testing Measure Rates and Reliability, by State and Selected Denominator, 1-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
HbA1c Testing	AZ	10	39.4	81.5%	11.2%	43.4%	83.8%	0.58	0.26
		20	49.2	82.1%	9.4%	31.2%	79.8%	0.61	0.39
		30	60.4	82.3%	8.9%	21.5%	72.5%	0.66	0.49
		50	83.8	82.7%	8.1%	10.4%	55.8%	0.73	0.61
	CA	10	45.1	80.6%	12.3%	45.8%	87.8%	0.71	0.35
		20	56.7	80.4%	11.6%	33.3%	84.3%	0.75	0.52
		30	69.3	80.3%	11.4%	24.0%	78.7%	0.79	0.62
		50	94.3	79.9%	11.0%	13.2%	66.0%	0.84	0.74
	IN	10	60.9	80.3%	11.3%	56.4%	93.1%	0.74	0.31
		20	67.3	80.6%	10.5%	49.5%	92.2%	0.75	0.47
		30	74.5	80.7%	10.2%	42.4%	90.0%	0.77	0.57
		50	93.2	80.7%	9.6%	27.8%	80.3%	0.80	0.68
	MA	10	36.1	80.6%	11.4%	40.2%	80.0%	0.59	0.29
		20	46.1	80.6%	10.3%	27.6%	73.7%	0.65	0.44
		30	56.9	80.8%	9.8%	18.5%	65.2%	0.69	0.54
		50	80.5	80.6%	8.8%	8.2%	46.4%	0.75	0.64
	MN	10	29.3	87.4%	8.5%	46.2%	83.4%	0.29	0.12
		20	37.5	87.8%	6.8%	29.8%	76.0%	0.33	0.20
		30	47.0	87.9%	6.3%	17.3%	63.8%	0.40	0.29
		50	66.2	88.7%	6.0%	5.6%	36.0%	0.53	0.46
	WI	10	39.7	88.3%	7.6%	55.8%	87.7%	0.42	0.15
		20	47.3	88.3%	6.8%	42.9%	83.8%	0.46	0.27
		30	55.9	88.3%	6.5%	31.2%	76.4%	0.51	0.35
		50	75.3	88.6%	5.7%	14.6%	55.9%	0.57	0.47

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-17: Calendar Year 2006 HbA1c Testing Measure Rates and Reliability, by State and Selected Denominator, 2-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
HbA1c Testing	AZ	10	33.2	82.3%	11.3%	37.4%	78.7%	0.56	0.28
		20	43.5	83.2%	9.6%	24.3%	71.4%	0.61	0.42
		30	55	83.7%	9.1%	15.3%	60.4%	0.67	0.53
		50	82.9	84.5%	8.1%	5.8%	38.5%	0.75	0.64
	CA	10	36.5	81.2%	13.0%	38.0%	81.6%	0.70	0.39
		20	48.8	81.1%	12.3%	24.6%	74.7%	0.76	0.56
		30	61	81.1%	12.0%	16.4%	66.2%	0.80	0.66
		50	86.7	81.1%	11.6%	7.7%	49.3%	0.85	0.77
	IN	10	47.9	81.1%	12.0%	51.0%	89.3%	0.72	0.35
		20	54.8	81.5%	11.0%	42.3%	87.1%	0.74	0.51
		30	63.4	81.4%	11.0%	32.9%	81.7%	0.78	0.62
		50	83.1	81.4%	10.7%	18.2%	64.7%	0.82	0.74
	MA	10	30.4	81.3%	11.8%	32.0%	71.4%	0.59	0.32
		20	40.5	81.3%	10.8%	19.7%	61.9%	0.66	0.49
		30	50.5	81.6%	10.2%	12.2%	50.4%	0.70	0.58
		50	75	81.9%	9.2%	4.2%	28.9%	0.77	0.69
	MN	10	22.6	88.7%	8.5%	32.2%	72.3%	0.30	0.16
		20	31.7	89.0%	7.3%	15.6%	55.0%	0.39	0.29
		30	41.7	89.5%	6.2%	6.8%	34.8%	0.44	0.36
		50	64.3	90.9%	6.1%	1.3%	11.4%	0.67	0.61
	WI	10	30.7	89.0%	8.1%	45.4%	80.2%	0.43	0.20
		20	39.2	89.2%	7.3%	30.0%	71.4%	0.49	0.33
		30	48.4	89.3%	6.8%	18.7%	58.0%	0.54	0.43
		50	71	89.5%	6.5%	5.9%	30.2%	0.66	0.58

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-18: Calendar Year 2006 HbA1c Testing Measure Rates and Reliability, by State and Selected Denominator, 30% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
HbA1c Testing	AZ	10	28.0	80.3%	13.2%	28.9%	77.6%	0.62	0.37
		20	38.1	81.5%	11.9%	16.7%	63.0%	0.69	0.53
		30	49.2	82.1%	11.5%	9.3%	46.1%	0.74	0.64
		50	74.7	83.0%	9.7%	3.0%	23.1%	0.78	0.71
	CA	10	29.0	78.9%	15.5%	28.7%	79.9%	0.72	0.47
		20	41.3	79.5%	14.5%	15.7%	64.6%	0.79	0.64
		30	53.6	79.6%	14.3%	9.2%	50.2%	0.83	0.73
		50	78.4	79.5%	14.0%	3.6%	29.5%	0.88	0.83
	IN	10	37.6	79.3%	14.4%	42.9%	90.2%	0.75	0.44
		20	45.1	80.3%	13.2%	32.4%	84.0%	0.77	0.60
		30	53.6	80.6%	13.0%	23.1%	72.6%	0.80	0.70
		50	71.6	81.0%	13.1%	10.6%	45.9%	0.86	0.81
	MA	10	26.3	79.8%	13.4%	28.1%	73.9%	0.62	0.38
		20	35.9	80.0%	12.1%	15.7%	58.2%	0.68	0.54
		30	46.3	80.4%	11.3%	8.3%	40.5%	0.72	0.63
		50	68.6	79.6%	10.2%	2.3%	17.2%	0.77	0.71
	MN	10	20.4	87.6%	9.9%	24.2%	69.6%	0.36	0.22
		20	29.4	88.1%	8.2%	10.2%	43.9%	0.45	0.35
		30	39.5	88.6%	8.5%	3.7%	22.0%	0.60	0.53
		50	60.5	92.0%	5.7%	0.5%	4.9%	0.60	0.55
	WI	10	26.6	88.0%	9.3%	38.9%	82.2%	0.45	0.23
		20	34.9	88.5%	8.1%	23.2%	66.9%	0.50	0.36
		30	44.5	89.0%	7.1%	12.4%	46.6%	0.54	0.44
		50	65.8	88.8%	7.4%	3.0%	17.1%	0.70	0.64

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-19: Calendar Year 2006 HbA1c Testing Measure Rates and Reliability, by State and Selected Denominator, 50% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
HbA1c Testing	AZ	10	25.6	80.6%	14.0%	24.1%	66.8%	0.64	0.41
		20	35.7	81.8%	12.8%	12.8%	49.4%	0.71	0.57
		30	48	82.3%	11.8%	6.3%	32.5%	0.76	0.66
		50	74.2	83.3%	10.1%	1.8%	14.3%	0.81	0.74
	CA	10	26.3	79.2%	16.4%	22.6%	65.3%	0.73	0.51
		20	38.7	79.6%	15.4%	11.4%	48.3%	0.80	0.67
		30	50.6	79.6%	15.3%	6.3%	34.9%	0.84	0.76
		50	74.8	79.2%	15.0%	2.2%	18.3%	0.89	0.84
	IN	10	34.1	80.0%	15.0%	36.6%	79.7%	0.75	0.47
		20	41.8	81.0%	14.0%	26.3%	70.2%	0.78	0.63
		30	50.3	80.9%	13.7%	17.7%	56.9%	0.81	0.72
		50	67.2	81.0%	14.3%	7.3%	31.4%	0.87	0.83
	MA	10	24.8	80.2%	13.8%	23.9%	65.6%	0.64	0.41
		20	34.3	80.4%	12.2%	12.6%	48.0%	0.71	0.57
		30	44.1	80.5%	11.9%	6.5%	31.7%	0.76	0.66
		50	67	79.4%	10.4%	1.5%	11.4%	0.77	0.72
	MN	10	19.4	88.4%	9.5%	19.1%	56.5%	0.34	0.21
		20	28.3	88.9%	8.0%	7.3%	31.7%	0.43	0.35
		30	37.7	89.6%	8.3%	2.5%	14.3%	0.57	0.52
		50	58.3	92.6%	6.8%	0.2%	2.2%	0.72	0.68
	WI	10	25	88.4%	9.5%	32.5%	72.1%	0.45	0.25
		20	33.3	89.2%	8.2%	18.4%	54.3%	0.51	0.38
		30	42.5	89.3%	7.4%	9.3%	35.2%	0.57	0.48
		50	63.3	89.0%	8.3%	1.9%	10.6%	0.74	0.69

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-20: Calendar Year 2006 HbA1c Testing Measure Rates and Reliability, by State and Selected Denominator, Maximum Frequency Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
HbA1c Testing	AZ	10	26.7	80.8%	13.8%	25.6%	74.0%	0.64	0.40
		20	36.2	82.0%	12.4%	14.5%	56.8%	0.70	0.56
		30	48.1	82.6%	11.3%	7.4%	38.4%	0.74	0.64
		50	74.2	83.2%	10.3%	2.1%	17.1%	0.80	0.73
	CA	10	27.4	79.2%	15.9%	25.4%	76.6%	0.72	0.49
		20	39.7	79.7%	15.0%	13.4%	58.2%	0.80	0.66
		30	51.9	79.9%	14.6%	7.5%	42.9%	0.84	0.75
		50	76.7	79.6%	14.4%	2.8%	23.2%	0.89	0.83
	IN	10	35.5	79.9%	14.7%	39.1%	88.5%	0.75	0.46
		20	42.7	80.8%	13.6%	29.1%	79.3%	0.77	0.62
		30	51.4	80.9%	13.3%	19.8%	65.0%	0.81	0.71
		50	68.8	81.1%	13.4%	8.5%	37.3%	0.86	0.81
	MA	10	25.2	80.2%	13.6%	25.1%	70.1%	0.61	0.39
		20	34.8	80.4%	12.1%	13.4%	51.8%	0.68	0.55
		30	44.7	80.6%	11.6%	7.0%	34.6%	0.73	0.64
		50	67.1	79.1%	9.8%	1.7%	13.0%	0.75	0.69
	MN	10	19.7	88.0%	9.8%	21.4%	64.6%	0.34	0.21
		20	28.9	88.9%	7.9%	8.4%	37.1%	0.42	0.33
		30	38.4	89.8%	6.6%	3.0%	17.7%	0.47	0.41
		50	60.9	91.9%	5.9%	0.3%	2.9%	0.64	0.59
	WI	10	25.6	88.3%	9.4%	34.7%	78.8%	0.45	0.24
		20	34.1	89.0%	8.1%	19.9%	60.1%	0.51	0.38
		30	43.2	89.3%	7.3%	10.4%	39.9%	0.56	0.47
		50	63.8	89.0%	7.6%	2.4%	13.4%	0.70	0.65

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-21: Calendar Year 2006 LDL Testing Measure Rates and Reliability, by State and Selected Denominator, 1-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
LDL Testing	AZ	10	39.4	77.6%	13.4%	43.4%	83.8%	0.69	0.37
		20	49.2	78.0%	12.2%	31.2%	79.8%	0.73	0.53
		30	60.4	78.1%	11.7%	21.5%	72.5%	0.77	0.63
		50	83.8	78.6%	11.0%	10.4%	55.8%	0.82	0.73
	CA	10	45.1	80.6%	11.9%	45.8%	87.8%	0.69	0.33
		20	56.7	80.7%	10.9%	33.3%	84.3%	0.74	0.49
		30	69.3	81.0%	10.4%	24.0%	78.7%	0.77	0.59
		50	94.3	81.4%	9.8%	13.2%	66.0%	0.82	0.70
	IN	10	60.9	74.3%	12.5%	56.4%	93.1%	0.77	0.35
		20	67.3	74.5%	12.0%	49.5%	92.2%	0.78	0.52
		30	74.5	74.7%	11.7%	42.4%	90.0%	0.80	0.62
		50	93.2	75.0%	11.0%	27.8%	80.3%	0.82	0.72
	MA	10	36.1	74.3%	14.1%	40.2%	80.0%	0.68	0.37
		20	46.1	74.3%	12.8%	27.6%	73.7%	0.72	0.53
		30	56.9	74.5%	11.9%	18.5%	65.2%	0.75	0.61
		50	80.5	74.3%	10.3%	8.2%	46.4%	0.78	0.68
	MN	10	29.3	79.9%	11.5%	46.2%	83.4%	0.48	0.24
		20	37.5	80.0%	9.9%	29.8%	76.0%	0.52	0.37
		30	47.0	80.2%	8.9%	17.3%	63.8%	0.55	0.44
		50	66.2	81.0%	7.7%	5.6%	36.0%	0.59	0.52
	WI	10	39.7	80.3%	10.5%	55.8%	87.7%	0.56	0.24
		20	47.3	80.4%	9.7%	42.9%	83.8%	0.60	0.39
		30	55.9	80.7%	9.1%	31.2%	76.4%	0.63	0.48
		50	75.3	80.9%	8.3%	14.6%	55.9%	0.69	0.60

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-22: Calendar Year 2006 LDL Testing Measure Rates and Reliability, by State and Selected Denominator, 2-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
LDL Testing	AZ	10	33.2	78.8%	13.6%	37.4%	78.7%	0.67	0.38
		20	43.5	79.6%	12.1%	24.3%	71.4%	0.71	0.54
		30	55	80.3%	11.2%	15.3%	60.4%	0.75	0.62
		50	82.9	81.6%	10.2%	5.8%	38.5%	0.81	0.72
	CA	10	36.5	81.3%	12.5%	38.0%	81.6%	0.67	0.36
		20	48.8	81.6%	11.2%	24.6%	74.7%	0.73	0.52
		30	61	81.8%	10.7%	16.4%	66.2%	0.77	0.62
		50	86.7	82.3%	10.0%	7.7%	49.3%	0.82	0.73
	IN	10	47.9	75.3%	13.6%	51.0%	89.3%	0.76	0.40
		20	54.8	75.6%	13.1%	42.3%	87.1%	0.79	0.57
		30	63.4	76.0%	12.6%	32.9%	81.7%	0.81	0.66
		50	83.1	76.1%	12.1%	18.2%	64.7%	0.84	0.76
	MA	10	30.4	74.9%	14.7%	32.0%	71.4%	0.68	0.41
		20	40.5	75.2%	13.5%	19.7%	61.9%	0.73	0.57
		30	50.5	75.6%	12.7%	12.2%	50.4%	0.77	0.66
		50	75	75.3%	11.6%	4.2%	28.9%	0.82	0.75
	MN	10	22.6	81.1%	12.2%	32.2%	72.3%	0.49	0.29
		20	31.7	81.4%	10.7%	15.6%	55.0%	0.56	0.44
		30	41.7	81.4%	9.5%	6.8%	34.8%	0.58	0.50
		50	64.3	82.8%	8.1%	1.3%	11.4%	0.66	0.60
	WI	10	30.7	81.2%	11.7%	45.4%	80.2%	0.58	0.31
		20	39.2	81.8%	10.5%	30.0%	71.4%	0.63	0.47
		30	48.4	82.0%	10.0%	18.7%	58.0%	0.68	0.57
		50	71	82.4%	9.1%	5.9%	30.2%	0.75	0.68

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-23: Calendar Year 2006 LDL Testing Measure Rates and Reliability, by State and Selected Denominator, 30% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
LDL Testing	AZ	10	28.0	75.9%	16.2%	28.9%	77.6%	0.71	0.47
		20	38.1	77.7%	14.3%	16.7%	63.0%	0.76	0.62
		30	49.2	78.9%	13.6%	9.3%	46.1%	0.80	0.71
		50	74.7	80.2%	12.0%	3.0%	23.1%	0.84	0.78
	CA	10	29.0	78.1%	15.3%	28.7%	79.9%	0.71	0.46
		20	41.3	79.0%	13.8%	15.7%	64.6%	0.77	0.62
		30	53.6	79.4%	13.2%	9.2%	50.2%	0.81	0.71
		50	78.4	79.7%	12.3%	3.6%	29.5%	0.86	0.80
	IN	10	37.6	73.4%	15.8%	42.9%	90.2%	0.77	0.47
		20	45.1	74.3%	14.9%	32.4%	84.0%	0.80	0.64
		30	53.6	74.0%	14.2%	23.1%	72.6%	0.82	0.72
		50	71.6	74.4%	14.3%	10.6%	45.9%	0.87	0.82
	MA	10	26.3	73.4%	16.3%	28.1%	73.9%	0.70	0.47
		20	35.9	74.0%	15.0%	15.7%	58.2%	0.76	0.63
		30	46.3	74.3%	13.9%	8.3%	40.5%	0.79	0.71
		50	68.6	73.4%	13.3%	2.3%	17.2%	0.84	0.80
	MN	10	20.4	79.9%	13.8%	24.2%	69.6%	0.54	0.37
		20	29.4	81.0%	11.6%	10.2%	43.9%	0.60	0.51
		30	39.5	80.7%	11.9%	3.7%	22.0%	0.71	0.65
		50	60.5	85.0%	8.6%	0.5%	4.9%	0.73	0.69
	WI	10	26.6	79.9%	13.2%	38.9%	82.2%	0.61	0.37
		20	34.9	81.2%	11.7%	23.2%	66.9%	0.66	0.53
		30	44.5	81.6%	11.1%	12.4%	46.6%	0.72	0.63
		50	65.8	81.9%	10.6%	3.0%	17.1%	0.80	0.75

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-24: Calendar Year 2006 LDL Testing Measure Rates and Reliability, by State and Selected Denominator, 50% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
LDL Testing	AZ	10	25.7	76.2%	17.1%	24.1%	66.8%	0.72	0.51
		20	35.7	78.5%	14.4%	12.8%	49.3%	0.76	0.64
		30	47.9	79.1%	13.6%	6.3%	32.7%	0.81	0.72
		50	74	80.4%	12.2%	1.8%	14.4%	0.85	0.79
	CA	10	26.3	78.3%	16.0%	22.6%	65.3%	0.71	0.49
		20	38.6	79.3%	14.3%	11.4%	48.3%	0.78	0.64
		30	50.6	79.7%	13.6%	6.3%	34.9%	0.82	0.72
		50	74.8	79.8%	13.6%	2.2%	18.3%	0.88	0.83
	IN	10	34.1	73.7%	16.6%	36.7%	79.8%	0.77	0.50
		20	41.8	74.7%	15.6%	26.3%	70.2%	0.81	0.66
		30	50.2	74.9%	15.3%	17.8%	57.0%	0.83	0.75
		50	67.5	74.6%	15.6%	7.2%	31.2%	0.88	0.84
	MA	10	24.8	73.9%	16.7%	23.8%	65.5%	0.70	0.48
		20	34.4	74.4%	15.5%	12.6%	48.0%	0.76	0.65
		30	44.2	74.6%	14.5%	6.5%	31.7%	0.79	0.72
		50	67.3	72.8%	13.4%	1.5%	11.2%	0.85	0.80
	MN	10	19.4	81.0%	13.6%	19.0%	56.4%	0.54	0.37
		20	28.2	81.7%	11.8%	7.4%	32.0%	0.61	0.53
		30	37.8	81.6%	11.2%	2.5%	14.2%	0.68	0.63
		50	57.8	85.7%	9.1%	0.3%	2.3%	0.74	0.71
	WI	10	25.1	80.6%	13.7%	32.3%	72.0%	0.62	0.40
		20	33.4	81.8%	12.0%	18.3%	54.2%	0.68	0.56
		30	42.7	82.4%	11.3%	9.2%	34.9%	0.73	0.66
		50	63.1	82.4%	10.7%	1.9%	10.7%	0.80	0.76

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-25: Calendar Year 2006 LDL Testing Measure Rates and Reliability, by State and Selected Denominator, Maximum Frequency Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
LDL Testing	AZ	10	26.7	76.4%	16.7%	25.6%	74.0%	0.72	0.50
		20	36.2	78.5%	14.5%	14.5%	56.8%	0.76	0.64
		30	48.1	79.6%	13.5%	7.4%	38.4%	0.81	0.72
		50	74.2	80.0%	12.8%	2.1%	17.1%	0.86	0.80
	CA	10	27.4	78.5%	15.4%	25.4%	76.6%	0.71	0.47
		20	39.7	79.6%	13.8%	13.4%	58.2%	0.77	0.63
		30	51.9	80.0%	13.2%	7.5%	42.9%	0.81	0.71
		50	76.7	80.1%	12.9%	2.8%	23.2%	0.87	0.81
	IN	10	35.5	73.9%	16.1%	39.1%	88.5%	0.77	0.49
		20	42.7	74.6%	15.4%	29.1%	79.3%	0.80	0.66
		30	51.4	75.0%	14.9%	19.8%	65.0%	0.83	0.74
		50	68.8	74.3%	15.2%	8.5%	37.3%	0.88	0.84
	MA	10	25.2	73.7%	16.6%	25.1%	70.1%	0.70	0.48
		20	34.8	74.4%	15.2%	13.4%	51.8%	0.76	0.64
		30	44.7	74.6%	14.3%	7.0%	34.6%	0.79	0.72
		50	67.1	72.9%	13.0%	1.7%	13.0%	0.83	0.79
	MN	10	19.7	80.3%	14.1%	21.4%	64.6%	0.54	0.38
		20	28.9	81.7%	11.5%	8.4%	37.1%	0.60	0.51
		30	38.4	82.3%	10.5%	3.0%	17.7%	0.66	0.60
		50	60.9	84.5%	8.9%	0.3%	2.9%	0.75	0.71
	WI	10	25.6	80.3%	13.5%	34.7%	78.8%	0.62	0.39
		20	34.1	81.8%	11.8%	19.9%	60.1%	0.67	0.54
		30	43.2	82.0%	11.4%	10.4%	39.9%	0.73	0.65
		50	63.8	82.1%	10.5%	2.4%	13.4%	0.80	0.76

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-26: Calendar Year 2006 Nephropathy Testing Measure Rates and Reliability, by State and Selected Denominator, 1-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Nephropathy Testing	AZ	10	39.4	70.1%	14.9%	43.4%	83.8%	0.74	0.42
		20	49.2	70.0%	14.1%	31.2%	79.8%	0.79	0.60
		30	60.4	69.6%	14.0%	21.5%	72.5%	0.83	0.70
		50	83.8	70.0%	13.7%	10.4%	55.8%	0.88	0.81
	CA	10	45.1	69.4%	14.9%	45.8%	87.8%	0.77	0.43
		20	56.7	69.5%	14.1%	33.3%	84.3%	0.81	0.61
		30	69.3	69.7%	13.7%	24.0%	78.7%	0.85	0.70
		50	94.3	70.1%	13.3%	13.2%	66.0%	0.89	0.80
	IN	10	60.9	61.4%	15.7%	56.4%	93.1%	0.84	0.46
		20	67.3	61.2%	15.4%	49.5%	92.2%	0.85	0.63
		30	74.5	61.4%	15.1%	42.4%	90.0%	0.86	0.72
		50	93.2	61.8%	14.6%	27.8%	80.3%	0.89	0.81
	MA	10	36.1	68.3%	15.3%	40.2%	80.0%	0.72	0.42
		20	46.1	68.1%	14.2%	27.6%	73.7%	0.77	0.59
		30	56.9	68.6%	13.8%	18.5%	65.2%	0.81	0.69
		50	80.5	69.2%	13.4%	8.2%	46.4%	0.86	0.80
	MN	10	29.3	71.9%	13.9%	46.2%	83.4%	0.62	0.36
		20	37.5	71.1%	13.0%	29.8%	76.0%	0.68	0.53
		30	47.0	70.3%	13.0%	17.3%	63.8%	0.74	0.65
		50	66.2	71.4%	12.3%	5.6%	36.0%	0.81	0.76
	WI	10	39.7	69.2%	14.9%	55.8%	87.7%	0.74	0.42
		20	47.3	69.0%	14.5%	42.9%	83.8%	0.78	0.60
		30	55.9	69.1%	14.3%	31.2%	76.4%	0.81	0.69
		50	75.3	69.9%	13.6%	14.6%	55.9%	0.85	0.79

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-27: Calendar Year 2006 Nephropathy Testing Measure Rates and Reliability, by State and Selected Denominator, 2-Touch Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Nephropathy Testing	AZ	10	33.2	70.2%	15.9%	37.4%	78.7%	0.74	0.47
		20	43.5	69.8%	15.6%	24.3%	71.4%	0.81	0.66
		30	55	70.1%	15.4%	15.3%	60.4%	0.85	0.75
		50	82.9	70.4%	15.8%	5.8%	38.5%	0.91	0.85
	CA	10	36.5	69.1%	16.3%	38.0%	81.6%	0.77	0.48
		20	48.8	69.2%	15.5%	24.6%	74.7%	0.83	0.66
		30	61	69.3%	15.2%	16.4%	66.2%	0.86	0.75
		50	86.7	69.9%	14.8%	7.7%	49.3%	0.90	0.84
	IN	10	47.9	61.1%	17.2%	51.0%	89.3%	0.83	0.51
		20	54.8	60.8%	17.1%	42.3%	87.1%	0.85	0.68
		30	63.4	60.7%	17.0%	32.9%	81.7%	0.87	0.77
		50	83.1	61.3%	16.5%	18.2%	64.7%	0.90	0.85
	MA	10	30.4	68.0%	16.5%	32.0%	71.4%	0.73	0.47
		20	40.5	68.0%	15.7%	19.7%	61.9%	0.79	0.65
		30	50.5	68.5%	15.4%	12.2%	50.4%	0.83	0.74
		50	75	68.9%	14.6%	4.2%	28.9%	0.88	0.83
	MN	10	22.6	71.4%	16.0%	32.2%	72.3%	0.64	0.44
		20	31.7	70.4%	15.7%	15.6%	55.0%	0.74	0.64
		30	41.7	70.1%	15.1%	6.8%	34.8%	0.79	0.73
		50	64.3	72.8%	14.4%	1.3%	11.4%	0.86	0.82
	WI	10	30.7	69.0%	16.9%	45.4%	80.2%	0.75	0.49
		20	39.2	68.8%	16.5%	30.0%	71.4%	0.80	0.67
		30	48.4	68.6%	16.2%	18.7%	58.0%	0.83	0.75
		50	71	69.9%	16.1%	5.9%	30.2%	0.89	0.85

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-28: Calendar Year 2006 Nephropathy Testing Measure Rates and Reliability, by State and Selected Denominator, 30% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Nephropathy Testing	AZ	10	28.0	66.8%	18.5%	28.9%	77.6%	0.76	0.53
		20	38.1	66.4%	18.1%	16.7%	63.0%	0.82	0.70
		30	49.2	66.2%	17.7%	9.3%	46.1%	0.85	0.78
		50	74.7	64.7%	17.3%	3.0%	23.1%	0.90	0.85
	CA	10	29.0	64.6%	19.1%	28.7%	79.9%	0.78	0.55
		20	41.3	64.6%	18.1%	15.7%	64.6%	0.84	0.72
		30	53.6	64.7%	17.9%	9.2%	50.2%	0.87	0.80
		50	78.4	64.7%	17.7%	3.6%	29.5%	0.91	0.87
	IN	10	37.6	57.2%	19.4%	42.9%	90.2%	0.82	0.56
		20	45.1	56.9%	19.1%	32.4%	84.0%	0.85	0.72
		30	53.6	56.7%	19.0%	23.1%	72.6%	0.88	0.80
		50	71.6	56.8%	19.1%	10.6%	45.9%	0.91	0.87
	MA	10	26.3	66.2%	17.5%	28.1%	73.9%	0.71	0.49
		20	35.9	65.6%	16.7%	15.7%	58.2%	0.78	0.66
		30	46.3	66.0%	16.5%	8.3%	40.5%	0.83	0.76
		50	68.6	63.8%	15.4%	2.3%	17.2%	0.86	0.82
	MN	10	20.4	68.3%	18.0%	24.2%	69.6%	0.67	0.50
		20	29.4	66.5%	17.6%	10.2%	43.9%	0.76	0.68
		30	39.5	66.2%	17.4%	3.7%	22.0%	0.81	0.77
		50	60.5	67.8%	15.1%	0.5%	4.9%	0.84	0.82
	WI	10	26.6	66.3%	18.7%	38.9%	82.2%	0.75	0.53
		20	34.9	66.5%	18.3%	23.2%	66.9%	0.81	0.71
		30	44.5	65.6%	18.6%	12.4%	46.6%	0.85	0.79
		50	65.8	66.1%	17.4%	3.0%	17.1%	0.88	0.85

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-29: Calendar Year 2006 Nephropathy Testing Measure Rates and Reliability, by State and Selected Denominator, 50% Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Nephropathy Testing	AZ	10	25.7	65.7%	19.7%	24.1%	66.8%	0.77	0.56
		20	35.9	65.2%	19.4%	12.7%	49.1%	0.83	0.73
		30	48	64.3%	18.9%	6.3%	32.7%	0.86	0.80
		50	74	62.0%	17.1%	1.8%	14.5%	0.89	0.84
	CA	10	26.3	63.3%	20.4%	22.6%	65.3%	0.78	0.58
		20	38.6	62.8%	19.4%	11.4%	48.2%	0.84	0.74
		30	50.5	63.3%	18.8%	6.3%	35.0%	0.88	0.81
		50	74.9	62.9%	19.4%	2.2%	18.2%	0.92	0.89
	IN	10	34	55.7%	20.9%	36.7%	79.8%	0.83	0.59
		20	41.8	55.7%	20.6%	26.3%	70.2%	0.86	0.74
		30	50.2	55.5%	20.1%	17.8%	57.0%	0.88	0.81
		50	67.3	55.8%	20.4%	7.3%	31.3%	0.91	0.88
	MA	10	24.8	65.4%	18.3%	24.0%	65.7%	0.72	0.51
		20	34.4	64.7%	17.4%	12.6%	48.1%	0.79	0.68
		30	44.3	64.9%	16.9%	6.5%	31.7%	0.83	0.76
		50	67.5	63.1%	16.2%	1.5%	11.2%	0.87	0.83
	MN	10	19.4	67.5%	19.2%	19.0%	56.4%	0.69	0.53
		20	28.2	65.4%	18.5%	7.3%	31.7%	0.77	0.70
		30	37.6	64.4%	17.4%	2.5%	14.4%	0.80	0.76
		50	58.5	65.2%	15.9%	0.2%	2.2%	0.85	0.83
	WI	10	25.1	66.1%	19.8%	32.4%	72.1%	0.77	0.57
		20	33.5	65.9%	19.5%	18.2%	54.1%	0.82	0.73
		30	42.6	64.2%	19.8%	9.3%	35.2%	0.86	0.81
		50	63.2	65.0%	19.2%	1.9%	10.8%	0.89	0.87

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Table A9-30: Calendar Year 2006 Nephropathy Testing Measure Rates and Reliability, by State and Selected Denominator, Maximum Frequency Rule

Measure	State	Minimum Denominator	Average Denominator	Mean Physician Rate	Standard Deviation	Percent eligible physicians w/ a rate[1]	Percent beneficiaries in denominator w/a physician[2]	Beta-binomial reliability at average rate	Beta-binomial reliability at minimum rate
Nephropathy Testing	AZ	10	26.7	66.5%	19.1%	25.6%	74.0%	0.76	0.55
		20	36.2	66.3%	18.7%	14.5%	56.8%	0.82	0.72
		30	48.1	65.5%	18.3%	7.4%	38.4%	0.86	0.79
		50	74.2	64.0%	17.2%	2.1%	17.1%	0.89	0.85
	CA	10	27.4	64.7%	19.6%	25.4%	76.6%	0.78	0.56
		20	39.7	64.3%	18.7%	13.4%	58.2%	0.84	0.73
		30	51.9	64.5%	18.2%	7.5%	42.9%	0.87	0.80
		50	76.7	64.3%	18.5%	2.8%	23.2%	0.92	0.88
	IN	10	35.5	57.0%	20.2%	39.1%	88.5%	0.83	0.58
		20	42.7	56.7%	20.0%	29.1%	79.3%	0.86	0.74
		30	51.4	56.3%	19.6%	19.8%	65.0%	0.88	0.81
		50	68.8	56.4%	19.8%	8.5%	37.3%	0.91	0.88
	MA	10	25.2	65.9%	17.9%	25.1%	70.1%	0.71	0.50
		20	34.8	65.0%	17.3%	13.4%	51.8%	0.79	0.68
		30	44.7	65.7%	16.8%	7.0%	34.6%	0.83	0.76
		50	67.1	63.4%	15.4%	1.7%	13.0%	0.86	0.82
	MN	10	19.7	68.2%	18.4%	21.4%	64.6%	0.67	0.51
		20	28.9	66.2%	18.3%	8.4%	37.1%	0.77	0.70
		30	38.4	66.7%	16.9%	3.0%	17.7%	0.80	0.76
		50	60.9	64.2%	16.4%	0.3%	2.9%	0.86	0.84
	WI	10	25.6	66.4%	19.5%	34.7%	78.8%	0.76	0.56
		20	34.1	66.4%	19.1%	19.9%	60.1%	0.82	0.72
		30	43.2	65.4%	19.3%	10.4%	39.9%	0.86	0.80
		50	63.8	65.2%	18.4%	2.4%	13.4%	0.89	0.86

[1] Percent of eligible physicians with a rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

[2] Percent of beneficiaries in denominator with a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Appendix 6

Defining Physician Groups to Report Quality Measures

Appendix 6
Defining Physician Groups to Report Quality Measures

One of the methodological questions the Better Quality Information (BQI) Project addressed was defining methods to assign physicians-to-physician groups using Medicare Fee-for-Service (FFS) administrative data. The structural, procedural, and reporting changes in the delivery of health care, as well as multiple methodological challenges in reporting performance at an individual physician level, have created a concomitant interest in measuring and reporting physician performance at the physician group level rather than for individual physicians. The BQI Project was tasked with providing the Centers for Medicare & Medicaid Services (CMS) and other stakeholders with information regarding the technical challenges encountered in, and the technical feasibility of, reporting physician performance rates at a physician group level using Medicare FFS administrative and related data to determine the physician group.

Some of the reasons for the growing interest in measuring and reporting physician performance at the group level include:

- Physicians appear more comfortable with group-level reporting of performance measures rather than at the individual level.
- Overall responsibility and accountability for achieving high quality, cost-effective patient care, especially in patients with chronic illness, is nearly always dependant on multiple physicians especially in a group setting.
- Except for measures where the denominator events (a population with a given disease or need for screening) occur at a relatively high rate in a single physician practice, the reliability (as measured by confidence intervals or reliability coefficients) for reporting individual physician performance results as an indicator of his/her true performance is low because of small sample size. This gives rise to major problems with misclassification (labeling someone as high performing when they are actually low and vice versa) and in the stability of measurement over time. Reporting at the group level provides substantially larger sample sizes and thus more confidence in the accuracy of feedback and classification.

Clearly, any effort to identify and to codify physician grouping methods needs to consider the impact on the resulting groups from issues arising from the current content of the Medicare claims data that are available for use. These factors include:

- What type of Medicare claims data should be used to identify groups for purposes of quality reporting?
- How many years of data are needed (i.e., look-back period)?
- What physician specialty types should be included?
- What constitutes a group (i.e., is a group of one physician considered a group or is a minimum number of physicians required)?
- Should the grouping be limited to physicians licensed in a given state?
- Should physician groups whose business location is out of the state of interest be included?

Each of these factors affects the number of physicians that can be included in the grouping, the number of groups that can be formed, and the number of beneficiaries that can be included or excluded as a result.

Special consideration should also be given to what and how the data is requested; based on either where the physician or beneficiary resides. If the data is requested by where the physician resides, it includes beneficiaries from other states that received services in the state and excludes beneficiaries who received services out-of-state. If the data is requested by where the beneficiary resides, it will include their services, no matter what state the beneficiary received the services. Depending on how the data is requested, the claims that would be included or excluded can significantly affect the grouping result.

When considering success or failure in physician grouping and measuring physician performance, the most important question to consider is whether physicians perceive the grouping method as a meaningful, acceptable, and valid way to measure their collective performance. Given these challenges, the pivotal policy question is, *Is physician grouping with Medicare FFS administrative data sufficiently accurate and robust to support its use in the reporting of physician group performance measure rates?*

To answer this question, we identified three sequential, intermediate activities:

1. Identifying feasible methods for attributing physicians to physician groups using Medicare FFS claims and related CMS data:
2. Comparing the methods for attributing physicians to physician groups by analyzing the impact of the selected options on the number of groups that can be created, the number and proportions of beneficiaries and physicians attributed to groups, and the average numbers of physicians and beneficiaries per group;
3. Comparing the methods identified above against the pilot communities' groups that are validated by provider feedback and interaction.

Using 2006 Medicare data from the carrier part B claims for six states (i.e., Arizona, California, Indiana, Massachusetts, Minnesota, and Wisconsin) and three Medicare physician reference files supplied with the claims data we first identified the data elements available in the Medicare FFS administrative data that could be used for physician grouping. The administrative physician reference files supplied by Medicare with the FFS claims data were:

- *MPIER*—"Master Physician Identification and Eligibility Record" (MPIER) contains one record for each practice setting for a physician. A physician will have a different profile number for each practice setting. The original source of information for this file is the data provided by physicians on CMS forms when physicians/health care practitioners and groups apply to become Medicare providers[1].
- *Member file* —contains one record per physician and physician group pairing. A physician will have multiple records if they are in more than one group.
- *Group file*—contains one record per group and contains information about the group.

The key data elements for physician grouping that we identified are described below.

Definition of key physician grouping data elements [2]

- *Unique Physician Identification Number (UPIN6)*—a unique 6-digit alphanumeric identifier assigned to physicians and other healthcare practitioners when they enroll in Medicare. A physician should have only one UPIN number. A UPIN is required for all physicians, health care practitioners, and group practices who bill Medicare or who order services that are billed to Medicare. UPINs are assigned nationally. CMS maintains a master file of UPIN information, and the Master Physician Identification and Eligibility Record (MPIER) file. For Carrier Claims, UPINs of both the "referring" physician and the "performing" physician appear in Medicare claims submitted by physicians. The UPIN of the physician who orders the test or service (the "referring"

1 http://www.resdac.umn.edu/tools/tbs/tb-002.asp

2 The National Provider Identification Number (NPI) is another data element that could be used for physician grouping that was implemented during the BQI Project, but had not been implemented during the timeframe covered by the data used for this project.

physician) is entered directly onto the Medicare billing form.[3]

- *Group Unique Physician Identifier Number (Group UPIN6)*—a unique 6-digit alphanumeric identifier assigned to a physician group. This data element exists only in the Medicare member file and Medicare group file (physician reference files) indicated above.
- *Group Unique Physician Identifier Number (Group UPIN4)*—A unique 4-digit number which contains information about a specific geographic location within a group UPIN6. This data element also is found only in the Medicare member file and Medicare group file (physician reference files). Neither group UPIN6 nor group UPIN4 appears in Medicare claims submitted by physicians.
- *Tax ID Number (TIN)* — A unique identifier used by the Internal Revenue Service for tax purposes. This identifier is included on the billing form.
- *Referring Physician*—A physician who requests service for the beneficiary for which payment may be made under the Medicare program.
- *Performing Physician*—Physicians who actually performs the service.

Step 1: Identify feasible methods for attributing physicians to physician groups using Medicare FFS claims and related CMS data

Based on these key data elements, we conducted analyses of assigning physicians to physician groups using the three different methods noted in Table 1. These represent the most promising methods that we identified for attributing physicians to physician groups using Medicare FFS and related CMS data. These methods utilize the UPINs and Tax IDs from the Medicare FFS claims data within the performing and referring physician fields, and the group UPINs from related CMS data files. The granularity of the group UPINs (i.e., group UPIN6) can be controlled through the use of the group UPIN4, which identifies the specific geographical location or clinic practice site within a corporate group UPIN6.

Table 1. Identified Methods to Define Physician Groups Using Medicare FFS and Related CMS Data

Name	Description
UPIN to group UPIN6	Each individual performing or referring UPIN linked to group UPIN6.
UPIN to group UPIN6+4	Each individual performing or referring UPIN linked to group UPIN6 + group UPIN4. Group UPIN4 contains information about the location of the clinic practice site within a larger group UPIN6.
Tax ID	Each individual performing UPIN linked with the Tax ID on the same claim.

Figures 1 and 2 show the processes and data flows for the UPIN grouping methods (i.e., UPIN6 or UPIN6+ UPIN4) and the Tax ID grouping methods, respectively. For the UPIN methods, we started by creating a list of eligible physicians from the claims data. This was defined as any physician with a valid UPIN (first character alpha with a value of A-M, P, R-Z followed by 5 numbers) who was listed as a "performing" or "referring" UPIN on a carrier claim in 2006. The list was not limited to primary care physicians, but did not necessarily include all other physician specialties.

3 http://www.resdac.umn.edu/tools/tbs/tb-002.asp

Figure 1. UPIN Methods Data Flow Chart

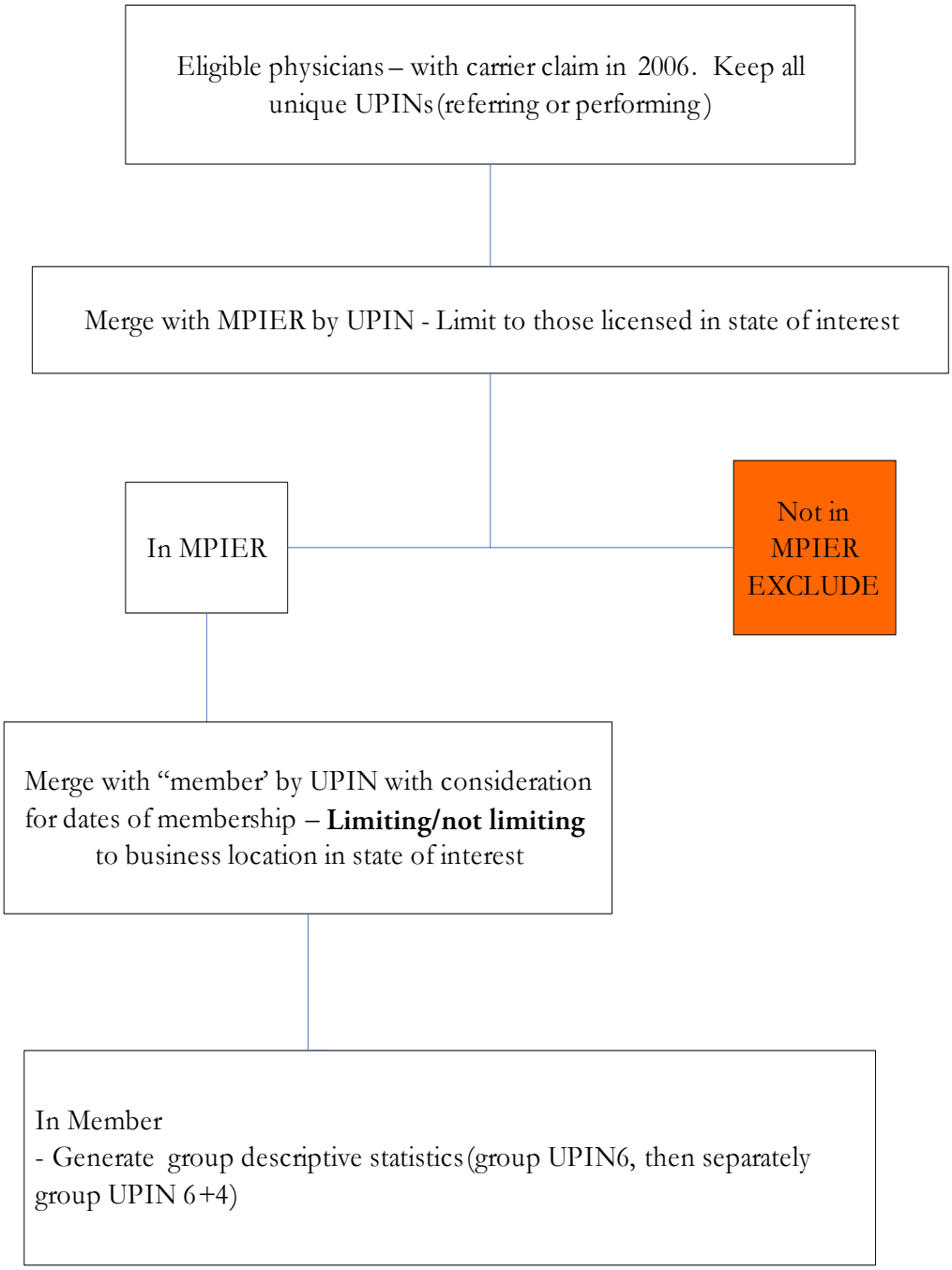
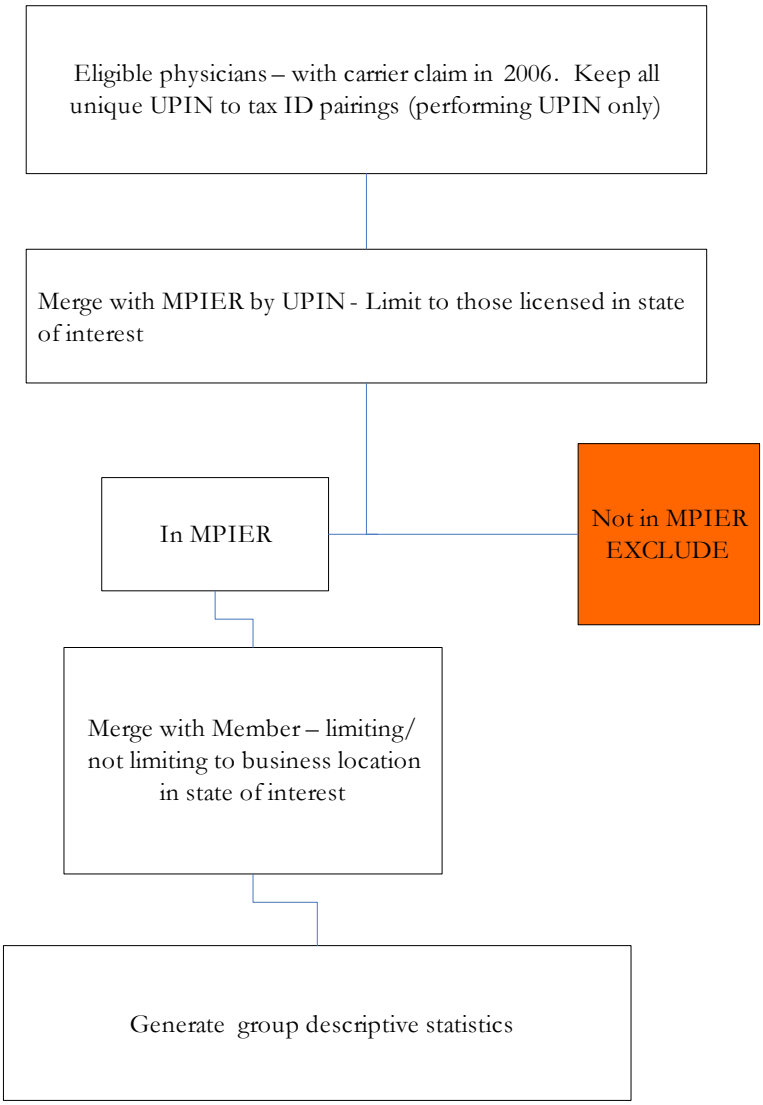


Figure 2. Tax ID Method Data Flow Chart



This list of eligible physicians was merged to the MPIER file to limit the list to *only* those licensed to practice in the state of interest. We then merged the list to the member file to pair the physicians with groups (UPIN6 and UPIN6+UPIN4 separately). The member file contains dates of group membership. For purposes of our analysis, pairs were limited to only those that were documented at any time during 2006. We then created groups using the four different options described in Table 2 below.

As noted earlier, the group UPIN6 and group UPIN4 do *not* appear in Medicare claims submitted by physicians. For UPIN grouping methods, the claims data must be merged with the member file using the physician UPIN as the common data element. If a physician belongs to multiple group UPIN6s (in the member file) and also provides services in one or more settings, there is no systematic way of knowing which patients should be assigned to a particular group UPIN6. Hence, all patients are assigned to every group UPIN6 the physician belongs to, allowing double counting of beneficiaries. If claims submitted by physicians always included both a group UPIN6 and a group UPIN4, it would effectively eliminate this problem.

For the Tax ID method, we started by creating a list of eligible physicians, with a valid UPIN (first character alpha with a value of A-M, P, R-Z followed by 5 numbers) who was listed as a "performing" UPIN on a carrier claim in 2006. The list was not limited to primary care physicians, but did not necessarily include all other physician specialties. Then this list of eligible physicians was merged to the MPIER file to limit the list to only those licensed to practice in the state of interest, and we created groups with the four options described in Table 2 below. For the options that limited the business location to the state of interest, the list was again merged with the business location contained in the member file.

Physician Grouping Options

We considered two policy-relevant eligibility conditions for attributing physicians to physician groups within the three methods for forming the overall groups discussed above. These conditions address the minimum number of physicians that can constitute a group (i.e., 1 vs. 2 or more) and an in-state limitation for the group practice (i.e., yes or no). The first of these options is important to consider because any methodology that requires at least two physicians to form a group will systematically remove solo practitioners from the analysis.

These two conditions are combined to form the four options identified in Table 2. Importantly, policy decisions on the relative merits of the choices in these options impact all three methods of group identification.

Table 2. Physician Group Attribution Options

Options	Description
Option 1	Any business location, no minimum number of physicians per group
Option 2	Any business location, minimum number of physicians per group greater than 1
Option 3	Business location restricted to state of interest, no minimum number of physicians per group
Option 4	Business location restricted to state of interest, minimum number of physicians per group greater than 1

Step 2: Comparing the methods for attributing physicians to physician groups by analyzing the impact of the selected options on the number of groups that can be created, the number and proportions of beneficiaries and physicians attributed to groups, and the average numbers of physicians and beneficiaries per group

As stated earlier, there are many factors that affect physician grouping and, consequently, physician performance results at the group level. The options considered here are among many other options that need to be considered. This section focuses on the impact of the three grouping methods and the four different grouping options that could be applied to each method on a number of issues, such as the number of groups created, the average group size, and the number of beneficiaries attributed to the groups.

Note: For each method, option 1 is used as the reference category.

Tables 3 and 4 summarize the effect of the different grouping options and grouping methods on the number of physician groups created across the six states. (For detailed results of Physician grouping and the effect of grouping options for each of the six states, and for the six states overall, refer to Tables 11 through 17, beginning on page 14.)

We reviewed the number of groups that each grouping method creates under different grouping options and found, as presented in Table 3, that regardless of the business location, group UPIN6+UPIN4 method creates more groups than any other method, followed by Tax ID and finally, group UPIN6.

Table 3. Number of Groups Created by Grouping Method (Business Location = Any State)

	UPIN to Group UPIN6 + UPIN4	Tax ID Method	UPIN to Group UPIN6
# of Physician Groups across 6 states	89,660	83,872	59,704

As shown in Table 4, when the state business location was limited to the state of interest, group UPIN6+UPIN4 still created the most groups although the difference is not as pronounced

Table 4. Number of Groups Created by Grouping Method (Business Location = State of Interest)

	UPIN to Group UPIN6 + UPIN4	Tax ID Method	UPIN to Group UPIN6
# of Physician Groups across 6 states	40,878	39,648	26,519

The group UPIN6+UPIN4 method was expected to generate the most groups because UPIN4 is a location-based subgroup of group UPIN6. Therefore, if a physician is a member of a group based on UPIN6 and practices at multiple locations but provided care at only one location, this method would attribute all patients to all locations even though the service was provided in only one location.

We also found that the change of business location from any state to the state of interest reduces the number of groups by approximately 50 percent. Setting a minimum requirement of two physicians per group, regardless of using the state of interest rule, reduces the number of groups by at least 40 percent.
Table 5 focuses on the effect of the different grouping options and grouping methods on the average number of physicians per group across the six states. Using the least restrictive option (including groups of one regardless of location) on average, the difference in number of physicians per group among the three grouping methods is minimal as shown in Table 5.

Table 5. Number of Physicians per Group using Three Grouping Methods, Non-Weighted Average (No minimum of physicians per group and Business Location = Any state)

	UPIN to Group UPIN6 + UPIN4	Tax ID Method	UPIN to Group UPIN6
Average number of physicians per group	4.1	3.6	4.8

If a minimum of two physicians is set, the average number of physicians per group increases to 10.7 for both the Tax ID method and the group UPIN6 method. For the UPIN6+UPIN4 method, the average number of physicians per group increases to 8.9. The change in this option doubles the average number of physicians per group while decreasing the number of groups by about 40%.

If business location is limited to the state of interest compared to when there is no restriction, the average number of physicians per group increases significantly. For the Tax ID method, the group size increased from 3.6 physicians per group to 9.5 physicians per group. For the group UPIN6 method, average group size increased from 4.8 physicians per group to 8.9 physicians per group. And, for the UPIN6+UPIN4 method, the average group size increased from 4.1 physicians per group to 7.2 physicians per group.

Table 6 further summarizes findings of the effect of the four physician grouping options. For simplicity and ease of presentation, the findings are based on non-weighted average results for all the six states. It is important to note that

the direction of change is the same, and overall, the magnitudes of percentage changes are largely comparable across the states. (For detailed summary by state refer to Table 11 through 17, which begin on page 17.

Table 6. Summary of Impact of Physician Grouping Options

Percentage Change in group statistics due to change in grouping options	TAX ID Percentage Change from option 1 (any business location, no minimum number of physicians per group)			GROUP UPIN6 Percentage Change from option 1 (any business location, no minimum number of physicians per group)			GROUP UPIN6+ UPIN4 Percentage Change from option 1 (any business location, no minimum number of physicians per group)		
	Option 2*	Option 3†	Option 4§	Option 2*	Option 3†	Option 4§	Option 2*	Option 3†	Option 4§
% decline in total # of medical groups	-79%	-53%	-85%	-60%	-56%	-75%	-60%	-54%	-73%
% decline in total # of physicians in all groups	-29%	-35%	-47%	-13%	-24%	-28%	-15%	-23%	-28%
% decline in total # of beneficiaries in all groups	-30%	-27%	-42%	-11%	-19%	-21%	-11%	-14%	-19%
% decline in total # of unique physicians in all groups	-28%	-38%	-47%	-8%	-12%	-16%	-10%	-12%	-17%
% decline in total # of unique beneficiaries in all groups	-7%	-6%	-11%	-1%	-1%	-2%	-1%	-1%	-2%
% Increase in average # of physicians per group	198%	163%	211%	125%	87%	202%	118%	76%	174%
% Increase in average # of beneficiaries per group	202%	44%	249%	128%	122%	224%	126%	98%	206%

* Any business location, minimum number of physicians per group greater than 1

† Business location restricted to state of interest, no minimum number of physicians per group.

§ Business location restricted to state of interest, minimum number of physicians per group greater that 1

The distinction between the total number of physicians or beneficiaries and the total number of *unique* physicians or beneficiaries is that physicians and beneficiaries may be counted multiple times depending on the number of groups to which they belong.

The small decrease in the number of beneficiaries when groups are limited to in-state business locations indicates that groups with out-of-state business locations did not account for many claims. However, there was a dramatic drop in the number of groups when limiting to in-state business locations. The likely explanation for this drop is that the data request was by the state of beneficiary; thus, there would be claims for some beneficiaries by groups who saw the beneficiary outside their state of residence.

Overall, in option 1 nearly 50 percent of the apparent groups consist of only one physician and, as a result, the group average is small. Even when the two rules are applied (minimum of two physicians and state location), the median size

of groups remains small. At the same time, with any of the three grouping methods in most states, there are few groups with hundreds of physicians. Thus, there is wide variation in distribution of the groups among the three physician grouping methods.

Compared to option 1 (the least restrictive option), the percentage decline in number of physicians is much faster than the percentage decline in unique beneficiaries as one moves across the three additional options in Table 2. In other words, the change in physician grouping option affects the number of physicians and the number of beneficiaries differently. This means although there is significant impact on the number of physicians included, those physicians excluded tend to have a smaller number of beneficiaries in their patient panel.

Discussion

The results of this analysis have important implications. If the focus is to measure the performance of physician groups, application of the more-restrictive physician grouping options will result in a relatively large loss in the number of physicians whose performance can be measured. But, if the focus is to measure the quality of services provided to the beneficiaries, the change in physician grouping options would marginally affect the number of beneficiaries for whom quality of services is measured.

It is well known that Medicare claims data was not designed or intended to measure physician performance and neither this data, nor the Tax ID number, was intended for use in assigning physician groups for quality measurement. Based on our analysis and limited work by some of the BQI pilot sites, there are indications that both the CMS data and Tax ID numbers have major limitations for use in defining physician groups.

Step 3: Comparing the methods identified above against the pilot communities' groups that are validated by provider feedback and interaction.

During the project, Delmarva requested that the pilots submit lists of their physicians groups to compare to groups created from administrative data. Although some pilots submitted physician group rosters, some of the rosters did not include any data elements that would allow comparison of the pilots' rosters against the physician groups created from the Medicare administrative data, such as Tax ID or group UPIN6. Hence, we were not able to make a thorough comparison of groups identified by the pilots opposed to those groups identified by the administrative data.

One pilot site, WCHQ, submitted a list which included Tax IDs representing 76 groups. We matched the Tax IDs of the 76 groups submitted by WHCQ (WHCQ groups) with 5,103 Tax IDs for groups created from Medicare administrative data (claims-created groups) and found that 40 of the Tax IDs matched. We then sent the list of the 40 Tax IDs and the physicians in each of the claims-created groups to WHCQ for validation. The 40 Tax ID groups represented 21 corporate entities. Of the 21 providers, 9 participated in validating the lists. The total number of physicians identified by Medicare data (UPIN) as being associated with the 9 entities was 4,218 physicians.

The WCHQ analysis indicates that only 78.7 percent of the physicians in the claims-created groups actually belonged to the WCHQ groups. Eight of the nine providers also included the number of physicians who were in the WCHQ groups but not in the claims-created groups or vice versa. According to WCHQ analysis, 343 physicians were in WCHQ groups but were not listed in claims-created groups, representing about 14 percent more physicians than had been identified in the claims-created groups. On the other hand, about 6 percent of physicians in claims-created groups were not members of the WCHQ groups.

Importantly, the 40 WCHQ groups that matched the claims-created groups represent only about 1 percent of all the groups created from the claims data, or about 27 percent of physicians in the claims-created groups for the state of Wisconsin. WCHQ groups are limited to only those provider groups with whom WCHQ works. This finding, therefore, should be interpreted with caution as it is not a representative sample of Tax ID groups in Wisconsin or of all WCHQ groups. Compared to all claims-created groups in Wisconsin, WCHQ groups are larger and have higher numbers of physicians per group.

WCHQ's analysis supported the earlier experience of another pilot site, Massachusetts Health Quality Partners (MHQP). Initially, MHQP calculated group-level results based on physician grouping determined by each of several commercial payers in their market. However, MHQP noted that the physician memberships of the medical groups did not agree across varying payers. MHQP created an interactive process by which the medical groups review the assigned physician membership of their groups and make corrections as membership changes. They found that all of the payer-created group rosters had significant inaccuracies. MHQP has found that allowing provider correction of group membership improved accuracy and garnered buy-in to the reported results. Another benefit of engaging physicians in the group identification process is that it has created a point of contact with the staff of the physician groups and allows for relationship building.

*Comparing the Physicians and B*eneficiaries *Assigned to Groups by the Differing Methods*

*Matching Individual Physicians and B*eneficiaries *in the Group UPIN Methods to the Tax ID Method*

Table 7 summarizes the number of groups created using the UPIN6 method and the number of groups created using the Tax ID method by state, and the number and percentage of physicians who are in groups created by both methods. The table also shows the percentage difference in the number of physicians in the two grouping methods. The overall objective is to identify how many physicians are in both groups (group UPIN6 method and Tax ID method), and the percentage difference in the number of physicians who are included in Tax ID and UPIN6 groups.

Table 7. Matching Individual Physicians in Groups Created Using the UPIN6 Method to Groups Created Using the Tax ID Method

STATE	# of Unique Physicians in Both Groups	# of Unique Physicians in TAX ID Group ONLY	# of Unique Physicians in GROUP UPIN6 ONLY	Total # of Unique Physicians in Group UPIN6	Total # of Unique Physicians in TAX ID Group	% in Both (of Group UPIN)	% in Both (of TAX ID)	% difference In # of physicians (of TAX ID)
Arizona	10,005	5,109	1,721	11,726	15,114	85%	66%	22%
California	45,336	31,418	10,860	56,196	76,754	81%	59%	27%
Indiana	13,529	4,717	1,984	15,513	18,246	87%	74%	15%
Massachusetts	20,549	10,811	4,084	24,633	31,360	83%	66%	21%
Minnesota	17,470	2,930	2,195	19,665	20,400	89%	86%	4%
Wisconsin	17,618	2,603	1,985	19,603	20,221	90%	87%	3%
ALL	124,507	57,588	22,829	147,336	182,095	85%	68%	19%

Across the six states, there were 182,095 unique physicians in Tax ID groups and 147,336 unique physicians in UPIN6 groups. The number of physicians in both groups was 124,507– representing 85 percent of UPIN6 groups

or 68 percent of Tax ID groups. Some variations exist among the states. Compared to Tax ID groups, the percent of physicians that are in both UPIN6 groups and Tax ID groups ranged from 59 percent for California to 87 percent for Wisconsin. Using the UPIN6 groups as the reference point, the percent of physicians that are in both UPIN6 groups and Tax ID groups ranged from 81 percent for California to 90 percent for Wisconsin.

Among the states, there is significant variation in the percentage of physicians who were in Tax ID groups but not in UPIN6 groups.

Similar to Table 7, Table 8 summarizes the number and percentage of beneficiaries who are in groups created by both the UPIN6 and the TaxID methods; and the percentage difference in the number of beneficiaries in the two grouping methods. The overall objective is to identify the number of beneficiaries in both groups and the percentage difference in the number of beneficiaries who are included in groups created by Tax ID method and groups created by the UPIN6 method. Here, we consider physician grouping methods where there is no restriction on business location and no minimum number of physicians per group (option 1).

Table 8. Matching Individual Beneficiaries in Group UPIN to Tax ID Method

STATE	# of Unique Beneficiary in both Methods	# of Unique Beneficiary in TAX ID Method ONLY	# of Unique Beneficiary in GROUP UPIN6 Method ONLY	Total # of Unique Beneficiary in Group UPIN6	Total # of Unique Beneficiary in TAX ID Method	% in both (of group UPIN6)	% in both (of TAX ID)	% difference In # Beneficiaries (of TAX ID)
Arizona	475,010	32,343	12,445	487,455	507,353	97%	94%	4%
California	2,339,028	159,497	22,225	2,361,253	2,498,525	99%	94%	5%
Indiana	777,924	35,506	2,804	780,728	813,430	100%	96%	4%
Massachusetts	717,914	3,607	24,003	721,521	741,917	100%	97%	3%
Minnesota	503,348	7,747	4,544	507,892	511,095	99%	98%	1%
Wisconsin	663,720	3,859	8,730	667,579	672,450	99%	95%	1%
ALL	5,476,944	242,559	74,751	5,526,428	5,744,770	99%	95%	4%

Overall, 5,744,770 unique beneficiaries were identified in Tax ID groups and 5,526,428 beneficiaries in UPIN6 groups. The two groups have 5,476,944 beneficiaries in common which is 99 percent of UPIN6 groups or 95 percent of Tax ID groups. While the rate of beneficiaries that are in both the UPIN6 groups and the Tax ID groups as a percentage of UPIN6 groups varies between 97 percent and 100 percent from state to state, the rate of beneficiaries that are in both the UPIN6 groups and the Tax ID groups as percentage of Tax ID groups ranges from between 94 percent to 98 percent from state to state.

The difference among the states in terms of the percentage of beneficiaries included in Tax ID groups but not UPIN6 groups is relatively low ranging from 1 to 5 percent. Overall, compared to Tax ID groups, Table 7 shows that there is a 19 percent difference in the number of physicians in Tax ID groups and UPIN6 groups. However, the 19 percent of physicians who did not match across the Tax ID groups and the UPIN6 groups and are not included in the UPIN6 groups had only a 4 percent impact in terms of the number beneficiaries who are excluded. In other words, those physicians who did not have a match also have very few beneficiaries. So, whether or not the physicians included in Tax ID groups matched with physicians included in UPIN6 groups had a big impact on the number of physicians in-

cluded in a particular group, but a much smaller impact on the number of beneficiaries included in a particular group.

Matching Physician Groups

Having stated the number of physicians and beneficiaries who are present in two different grouping methods, the final question becomes —How many of the physician groups created are identical?
Because the data elements used for grouping (Tax ID and UPIN groups) are not the same, direct matching was not possible. Therefore, we used what is referred to as a 'fuzzy match' technique to match physician groups. The process involves a bit of complicated computer programming based on having exactly the same number of physicians in both groups *and* the physicians in those two groups being identical. The result of the matching indicates how likely the two physician grouping methods are to produce the exact same groups. The groups we will be looking at are:

- Tax ID method vs. UPIN6 groups
- Tax ID method vs. UPIN6 + UPIN4 groups

Table 9 summarizes, by state, the number of UPIN6 groups and Tax ID groups, and the number and percentage of groups which are identical.

Table 9. Matching Physician Groups Based on Physician Membership: TAX ID and Group UPIN6

STATE	Total # of TAX ID Groups	Total # of UPIN6 Groups	Group UPIN6 with ≥ 1 physician in Common with TAXID Group	% of Groups with ≥ 1 Physicians in Common	# of Groups with 100% Match (identical)	% of Groups with Identical physicians in Both groups
Arizona	7,117	7,336	6,535	89%	2,529	39%
California	46,276	21,804	19,130	88%	7,849	41%
Indiana	7,329	8,435	7,235	86%	2,648	37%
Massachusetts	13,730	9,290	7,596	82%	2,250	30%
Minnesota	4,317	5,456	4,818	88%	1,668	35%
Wisconsin	5,103	7,383	6,507	88%	2,118	33%
ALL	83,872	59,704	51,821	87%	19,062	37%

There are 83,872 Tax ID groups and 59,704 UPIN6 groups among the six states. Of 59,704 UPIN6 groups, 87 percent, or 51,821, have at least 1 physician in common with Tax ID groups. Of the 51,821 physicians groups, however; only 37 percent (19,062) are identical groups. The rates are similar among the six states ranging from 30 percent for Massachusetts and 41 percent for California. However, 83 percent, or 19,062, of the groups that are identical regardless of whether the Tax ID method was used to create them or the UPIN6 method was used to create them are groups that consist of 1 physician.

Table 10 shows, by state, the number of UPIN6 + UPIN4 groups and TAX ID groups, and the number and percentage of groups which are identical.

Table 10. Matching Physician Group Based on Physician Membership: TAX ID and Group UPIN6 + UPIN4

STATE	Total # of TAX ID Groups	Total # of UPIN6+ UPIN4 Groups	Group UPIN6 + UPIN4 with ≥ 1 physician in Common with TAX ID Group	% of Groups with ≥ 1 Physicians in Common	# of Groups with 100% Match (identical)	% of Groups with Identical physicians in Both groups
Arizona	7,117	9,918	8,731	88%	3,151	36%
California	46,276	32,331	28,308	88%	9,752	34%
Indiana	7,329	12,868	11,139	87%	3,460	31%
Massachusetts	13,730	13,041	10,565	81%	2,872	27%
Minnesota	4,317	10,067	9,006	89%	2,051	23%
Wisconsin	5,103	11,435	10,135	89%	2,336	23%
ALL	83,872	89,660	77,884	87%	23,622	30%

There are 89,660 UPIN6+ UPIN4 groups among the six states. Of 89,660 UPIN4+UPIN6 groups, 87 percent or 77,884 have at least 1 physician in common with Tax ID groups. Of the 77,884 physicians groups, however; only 30 percent are identical groups

Conclusions Regarding Assigning Physicians to Groups Using Administrative Data

In response to the original question of whether physician grouping with Medicare FFS administrative data alone is sufficiently accurate and robust to support its use in the reporting of physician group performance measure rates the results of the BQI Project suggest that the answer is: not quite.

Currently, the accuracy of assigning physicians to physician groups using Medicare administrative data, or any claims-based data alone, is not accurate. To gain accuracy in reflecting actual physician group rosters, all of the pilot sites recommend validating the group rosters with the physician groups. An interactive process that allows correction, additions, and deletions over time is a preferred method.

Nonetheless, we believe that with quality improvement approaches that address the challenges surrounding accuracy of data and focused additional research, the claims data could serve as an effective method of grouping physicians for public reporting of some performance measure rates. For example, consistent use of the newly-implemented National Provider Identification (NPIs) numbers will likely improve accuracy of group attribution.

In markets dominated by very large medical groups, using the Tax ID of the corporate group to calculate group-level measures has been acceptable to those providers. The UPIN to Tax ID method is better positioned for grouping physicians than the UPIN to group UPIN methods (i.e., the UPIN6 and UPIN6+UPIN4 methods) because both the Tax ID and UPIN are included in the claims submitted by physicians, while group UPIN6 and UPIN4 are not.

Another issue of great importance is the need for involving physicians in the process of defining methods for physician grouping. Unless the methods are transparent and acceptable to physicians, progress in acceptable physician grouping and measuring physician group performance will be significantly delayed.

Future Research:

Delmarva has addressed the basic methodological questions of defining methods of physician groups using Medicare FFS administrative data. However, based on our analysis and inputs from some of the pilot sites, we believe that additional research projects, outside the scope of the current BQI Project, may include:
Validation of Grouping Result—The limited work done in collaboration with WHCQ raises serious questions regarding the quality of Medicare data in general and specifically for physician grouping purposes.
Specialty Type —The effect of limiting physicians by specialty type and deciding what type of specialty should be used for grouping purposes needs to be addressed.
Look-back Period— What is the effect of look-back time periods, if any? Is one year enough for grouping purposes— A longer look-back period will increase the number of physicians who can be assigned to a group, but it may also assign some physicians who may no longer be in the group.
Claims Data Type— What is the appropriate claims data type that should be used for physician grouping purpose (e.g., limit to carrier claims data, limit to claims for services provided in an office setting)
Acceptability to Physicians — Acceptability of using the Tax ID and UPIN methods to physicians as a way to measure their collective performance.
Analyses of Performance Measures and Reliability of Measures — Once consensus is reached on the issues raised above, analyses of performance measures and reliability should be explored at the group level.
Exploration of Alternatives— For example, could a new "quality reporting" group UPIN number be developed and implemented that would allow physicians to determine which physicians within a formal group (or in a virtual group) to be linked to for quality assessment and reporting.

These recommendations for future analysis identified in this report are just a few suggestions of the efforts that could lead to a broader understanding of the different factors that affect physician grouping results.

Limitations

Our analysis is based on Medicare administrative FFS data, and as such is a purely data-driven analysis with minimal direct validation. Due to issues around sharing of data, our effort to carry out any substantial validation work in this area was severely limited. We also believe validation work should not be limited to just providers who work with the pilot sites or similar groups. To avoid any potential bias, the work in this area should be based on a representative sample of all groups in a state (or the country), and should not be limited to just providers who work with pilots or any other similar organizations.

Table 11: Physician Grouping and Effect of Grouping Options Using 2006 Medicare FFS Data, Arizona

ARIZONA (AZ)												
Summary Statistics	Group Business Location Not Limited to AZ						Group Business Location Limited to AZ					
	UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)		UPIN to Group Tax ID (TIN)		UPIN to G roup UPIN6		UPIN to (Group UPIN6 +UPIN4)	
	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2
Distribution of groups												
Total # of physician groups	7,117	1,632	7,336	2,435	9,918	2,987	3,414	1,094	2,732	1,386	3,268	1,769
Total # of physicians attributed to a group	18,452	12,967	22,012	17,111	25,774	18,843	11,037	8,717	14,494	13,148	16,349	14,850
Total # of unique physicians attributed to a group	15,079	10,458	11,726	10,044	11,726	9,777	8,314	6,643	9,550	8,676	9,550	8,627
Average # physicians attributed per group	2.6	8.0	3.0	7.0	2.6	6.3	32.2	8.0	5.3	9.5	5.0	8.4
Min (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Max (# of physicians attributed per group)	604	604	380	380	380	380	326	326	380	380	380	380
Mode (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2

ARIZONA (AZ)												
Summary Statistics	Group Business Location Not Limited to AZ						Group Business Location Limited to AZ					
	UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)		UPIN to Group Tax ID (TIN)		UPIN to G roup UPIN6		UPIN to (Group UPIN6 +UPIN4)	
	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2
Median (# of physicians attributed per group)	1	3	1	3	1	3	1	3	2	3	2	3
Distribution of Beneficiaries												
Total # of beneficiaries attributed to a group	2,538,367	1,790,379	4,855,107	3,911,340	5,749,439	4,437,385	1,741,779	1,354,619	3,524,148	3,223,685	4,080,377	3,741,236
Total # of unique beneficiaries attributed to a group	507,320	469,654	494,068	484,507	494,068	483,316	464,768	433,647	487,455	478,885	487,455	478,489
Average # beneficiaries attributed per group	356.7	1097.1	661.8	1606.3	579.7	1485.6	510.2	1238.2	1289.0	2325.9	1248.6	2114.9
Min (# of beneficiaries attributed per group)	1	1	1	1	1	1	1	1	1	1	1	1
Max (# of beneficiaries attributed per group)	46,059	46,059	99,335	99,335	99,335	99,335	45,208	45,208	99,335	99,335	99,335	99,335
Mode (# of beneficiaries attributed per group)	1	2	1	2	1	3	1	12	1	56	1	56
Median (# of beneficiaries attributed per group)	66	376.5	107	454	89	480	125	442.5	267.5	659.5	314	735

Table 12: Physician Grouping and Effect of Grouping Options Using 2006 Medicare FFS Data, California

California (CA)												
Summary Statistics	Group Business Location Not Limited to CA						Group Business Location Limited to CA					
	UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)		UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)	
	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2
Distribution of groups												
Total # of physician groups	46,276	7,787	21,804	9,186	32,331	14,216	19,632	5,166	11,964	6,375	18,379	10,583
Total # of physicians attributed to a group	98,055	59,566	91,519	78,901	127,128	109,013	56,063	41,597	72,216	66,627	103,945	96,149
Total # of unique physicians attributed to a group	76,425	47,152	56,196	50,876	56,196	49,895	40,604	32,472	49,693	46,273	49,693	45,768
Average # physicians attributed per group	2.1	7.6	4.2	8.6	3.9	7.7	2.9	8.1	6.0	10.5	5.7	9.1
Min (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Max (# of physicians attributed per group)	923	923	1,942	1,942	347	347	632	632	1,942	1,942	347	347
Mode (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Median (# of physicians attributed per group)	1	3	1	3	1	3	1	3	2	3	2	4
Distribution of Beneficiaries												
Total # of beneficiaries attributed to a group	14,208,804	8,256,744	22,239,512	19,666,674	36,404,835	32,460,569	9,638,505	6,873,057	17,970,607	17,970,607	33,305,031	30,661,605
Total # of unique beneficiaries attributed to a group	2,498,405	2,229,183	2,361,253	2,319,173	2,361,253	2,312,321	2,302,379	2,106,651	2,346,722	2,306,382	2,346,722	2,300,283
Average # beneficiaries attributed per group	307.0	1,060.3	1,019.6	2,140.9	1,126.0	2,283.4	491.0	1,330.4	2,818.9	2,818.9	1,812.1	2,897.2
Min (# of beneficiaries attributed per group)	1	1	1	1	1	1	1	1	1	1	1	1
Max (# of beneficiaries attributed per group)	144,537	144,537	186,365	186,365	60,350	60,350	129,830	129,830	186,304	186,304	60,350	60,350
Mode (# of beneficiaries attributed per group)	1	2	1	2	1	3	1	1	6	6	1	6
Median (# of beneficiaries attributed per group)	59	314	150	635	173	742	120	471	955	955	481	1,085

Table 13: Physician Grouping and Effect of Grouping Options Using 2006 Medicare FFS Data, Indiana

INDIANA (IN)												
Summary Statistics	Group Business Location Not Limited to IN						Group Business Location Limited to IN					
	UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)		UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)	
	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥≥	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group≥ 1	Required # of Physicians Per Group ≥ 2
Distribution of groups												
Total # of physician groups	7,329	2,102	8,435	3,329	12,868	4,929	3,694	1,565	2,909	1,860	4,432	2,839
Total # of physicians attributed to a group	22,139	16,912	30,082	24,976	38,607	30,668	14,966	12,837	19,704	18,655	23,958	22,365
Total # of unique physicians attributed to a group	18,180	14,105	15,513	14,139	15,513	13,754	11,733	10,454	13,051	12,432	13,051	12,237
Average # physicians attributed per group	3.0	8.1	3.6	7.5	3.0	6.2	4.1	8.2	6.8	10.0	5.4	7.9
Min (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Max (# of physicians attributed per group)	207	207	233	233	212	212	198	198	233	233	212	212
Mode (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Median (# of physicians attributed per group)	1	3	1	3	1	3	1	4	2	4	2	4
Distribution of Beneficiaries												
Total # of beneficiaries attributed to a group	4,168,084	3,334,961	7,900,445	6,961,567	11,604,769	10,060,430	3,057,151	2,708,855	6,186,220	5,882,328	9,166,265	8,650,022
Total # of unique beneficiaries attributed to a group	813,398	781,254	780,728	774,429	780,728	772,113	768,245	746,580	774,961	768,676	774,961	766,584
Average # beneficiaries attributed per group	568.7	1586.6	936.6	2091.2	901.8	2041.1	827.6	1730.9	2126.6	3162.5	2068.2	3046.9
Min (# of beneficiaries attributed per group)	1	1	1	1	1	1	1	1	1	1	1	1
Max (# of beneficiaries attributed per group)	63,646	63,646	83,764	83,764	57,127	57,127	63,447	63,447	83,764	83,764	57,127	57,127
Mode (# of beneficiaries attributed per group)	1	2	1	4	1	2	1	3	1	54	1	75
Median (# of beneficiaries attributed per group)	90	521	106	636	108	639	148	612	541	1,048	553	1,007

Table 14: Physician Grouping and Effect of Grouping Options Using 2006 Medicare FFS Data, Massachusetts

Massachusetts (MA)												
Summary Statistics	Group Business Location Not Limited to MA						Group Business Location Limited to MA					
	UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)		UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)	
	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2
Distribution of groups												
Total # of physician groups	13,730	2,885	9,290	3,764	13,041	4,927	6,718	1,803	3,748	2,339	4,725	3,060
Total # of physicians attributed to a group	43,518	32,673	46,166	40,640	55,688	47,574	27,510	22,595	35,267	33,858	42,169	40,504
Total # of unique physicians attributed to a group	31,218	24,118	24,633	22,904	24,633	22,564	18,283	15,979	21,408	20,728	21,408	20,672
Average # physicians attributed per group	3.2	11.3	5.0	10.8	4.3	9.7	4.1	12.5	9.4	14.5	8.9	13.2
Min (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Max (# of physicians attributed per group)	1,561	1,561	2,046	2,046	1,320	1,320	1,285	1,285	2,046	2,046	1,320	1,320
Mode (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Median (# of physicians attributed per group)	1	3	1	3	1	3	1	3	2	4	2	4
Distribution of Beneficiaries												
Total # of beneficiaries attributed to a group	4,125,083	3,197,128	9,483,713	8,647,144	12,529,269	11,301,868	3,073,736	2,553,146	8,177,108	7,762,331	10,791,772	10,269,945
Total # of unique beneficiaries attributed to a group	741,888	719,969	721,521	716,764	721,521	716,366	708,932	689,142	717,485	713,517	717,485	713,400
Average # beneficiaries attributed per group	300.4	1108.2	1020.9	2297.3	960.8	2293.9	457.5	1416.1	2181.7	3318.7	2284.0	3356.2
Min (# of beneficiaries attributed per group)	1	1	1	1	1	1	1	1	1	1	1	1
Max (# of beneficiaries attributed per group)	61,848	61,848	178,502	178,502	114,560	114,560	55,306	55,306	178,502	178,502	114,560	114,560
Mode (# of beneficiaries attributed per group)	1	2	1	2	1	2	1	2	1	5	1	5
Median (# of beneficiaries attributed per group)	14	231	67	555	52	558	32	461	490	1,062	532	1,111

Table 15: Physician Grouping and Effect of Grouping Options Using 2006 Medicare FFS Data, Minnesota

Minnesota (MN)												
Summary Statistics	Group Business Location Not Limited to MN						Group Business Location Limited to MN					
	UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)		UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)	
	Required # of Physicians Per Group > 1	Required # of Physicians Per Group > 2	Required # of Physicians Per Group > 1	Required # of Physicians Per Group > 2	Required # of Physicians Per Group > 1	Required # of Physicians Per Group > 2	Required # of Physicians Per Group > 1	Required # of Physicians Per Group > 2	Required # of Physicians Per Group > 1	Required # of Physicians Per Group > 2	Required # of Physicians Per Group > 1	Required # of Physicians Per Group > 2
Distribution of groups												
Total # of physician groups	4,317	1,304	5,456	1,976	10,067	3,965	2,810	1,050	2,148	1,216	5,289	2,931
Total # of physicians attributed to a group	24,618	21,605	40,871	37,391	61,159	55,057	20,361	18,601	34,238	33,306	51,924	49,566
Total # of unique physicians attributed to a group	20,356	17,832	19,665	18,710	19,665	18,535	16,473	15,135	18,500	17,903	18,500	17,800
Average # physicians attributed per group	5.7	16.6	7.5	18.9	6.1	13.9	7.3	17.7	15.9	27.4	9.8	16.9
Min (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Max (# of physicians attributed per group)	2,671	2,671	3,440	3,440	3,425	3,425	2,304	2,304	3,440	3,440	3,425	3,425
Mode (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Median (# of physicians attributed per group)	1	3	1	3	1	4	1	3	2	4	2	4
Distribution of Beneficiaries												
Total # of beneficiaries attributed to a group	1,774,701	1,615,799	5,520,345	5,128,326	10,034,745	9,244,278	1,570,653	1,476,033	4,675,949	4,571,120	8,538,076	8,172,311
Total # of unique beneficiaries attributed to a group	511,074	501,523	507,892	504,695	507,892	504,078	499,124	492,576	505,838	502,749	505,838	502,162
Average # beneficiaries attributed per group	411.1	1239.1	1011.8	2595.3	996.8	2331.5	559.0	1405.8	2176.9	3759.1	1614.3	2788.2
Min (# of beneficiaries attributed per group)	1	1	1	1	1	1	1	1	1	1	1	1
Max (# of beneficiaries attributed per group)	48,958	48,958	138,001	138,001	137,840	137,840	47,973	47,973	138,001	138,001	137,840	137,840
Mode (# of beneficiaries attributed per group)	1	2	1	2	1	3	1	2	1	23	1	69
Median (# of beneficiaries attributed per group)	24	179	59	395	99	504	30	229	153	629	219	658

Table 16: Physician Grouping and Effect of Grouping Options Using 2006 Medicare FFS Data, Wisconsin

Wisconsin (WI)												
Summary Statistics	Group Business Location Not Limited to WI						Group Business Location Limited to WI					
	UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)		UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)	
	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2
Distribution of groups												
Total # of physician groups	5,103	1,725	7,383	3,067	11,435	4,754	3,380	1,509	3,018	1,977	4,785	3,232
Total # of physicians attributed to a group	25,557	22,179	38,985	34,669	53,447	46,766	21,267	19,396	29,319	28,278	39,938	38,385
Total # of unique physicians attributed to a group	20,375	17,785	19,803	18,673	19,803	18,384	16,540	15,345	18,078	17,495	18,078	17,361
Average # physicians attributed per group	5.0	12.9	5.3	11.3	4.7	9.8	6.3	12.9	9.7	14.3	8.4	11.9
Min (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Max (# of physicians attributed per group)	1,155	1,155	1,455	1,455	1,140	1,140	1,036	1,036	1,454	1,454	1,140	1,140
Mode (# of physicians attributed per group)	1	2	1	2	1	2	1	2	1	2	1	2
Median (# of physicians attributed per group)	1	3	1	3	1	3	1	4	2	4	2	4
Distribution of Beneficiaries												
Total # of beneficiaries attributed to a group	2,588,750	2,252,965	7,841,667	7,110,691	12,646,050	11,436,438	2,349,432	2,152,217	6,332,684	6,156,615	10,421,716	10,125,979
Total # of unique beneficiaries attributed to a group	672,423	660,709	667,579	665,053	667,579	664,585	659,817	652,866	664,926	662,693	664,926	662,367
Average # beneficiaries attributed per group	507.3	1,306.1	1,062.1	2,318.5	1,105.9	2,405.7	695.1	1,426.3	2,098.3	3,114.1	2,177.0	3,133.0
Min (# of beneficiaries attributed per group)	1	1	1	1	1	1	1	1	1	2	1	1
Max (# of beneficiaries attributed per group)	95,051	95,051	209,108	209,108	96,573	96,573	91,834	91,834	209,012	209,012	96,573	96,573
Mode (# of beneficiaries attributed per group)	1	2	1	4	1	4	1	3	1	10	1	10
Median (# of beneficiaries attributed per group)	52	217	108	463	120	542.5	77	254	273	632	331	715

Table 17: Physician Grouping and Effect of Grouping Options Using 2006 Medicare FFS Data, All Six States Overall

Overall Results (All 6 States Combined)												
Summary Statistics	Group Business Location Not Limited to the states						Group Business Location Limited to the states					
	UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)		UPIN to Group Tax ID (TIN)		UPIN to Group UPIN6		UPIN to (Group UPIN6 +UPIN4)	
	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2	Required # of Physicians Per Group ≥ 1	Required # of Physicians Per Group ≥ 2
Total # of physician groups	83,872	17,435	59,704	23,757	89,660	35,778	39,648	12,187	26,519	15,153	40,878	24,414
Total # of physicians attributed to a group	232,339	165,902	269,635	233,688	361,803	307,921	151,204	123,743	205,238	193,872	278,283	261,819
Total # of unique physicians attributed to a group	181,633.0	131,450.0	147,536.0	135,346.0	147,536.0	132,909.0	111,947.0	96,028.0	130,280.0	123,507.0	130,280.0	122,465.0
Average # physicians attributed per group	3.6	10.7	4.8	10.7	4.1	8.9	9.5	11.2	8.9	14.4	7.2	11.2
Min (# of physicians attributed per group)	29,403,789	20,447,976	57,840,789	51,425,742	88,969,107	78,940,968	21,431,256	17,117,927	46,866,716	45,566,686	76,303,237	71,621,098
Max (# of physicians attributed per group)	5,744,508	5,362,292	5,533,041	5,464,621	5,533,041	5,452,779	5,403,265	5,121,462	5,497,387	5,432,902	5,497,387	5,423,285
Mode (# of physicians attributed per group)	408.5	1,232.9	952.1	2,174.9	945.2	2,140.2	590.1	1,424.6	2,115.2	3,083.2	1,867.4	2,889.4

Appendix 7

Reliability of Physician Performance Measures

Appendix 7
Reliability of Physician Performance Measures

Rationale for Considering Measure Reliability

To spur quality improvement, both private and public entities are demanding physician accountability and rewarding those that meet stated quality goals. Public and private purchasers are also demanding more information about quality of health care services provided by physicians and hospitals (Baker and Carter, 2005; Galvin and Milstein, 2002). The purpose of this appendix is to understand the reliability of available information for making decisions about physician performance.

There are many issues and challenges in developing physician level profiles (Hofer, 1999). To support fair and accurate comparisons among physicians, a number of factors must be considered including measures that are evidence-based, data sources that are complete and accurate, and standardized methods of data collection. Other factors that affect measurement include the type of measure, the approach for scoring, the use of case- mix adjustment, and how care is attributed to physicians.

Physician-level reliability of a measure is another key consideration in this measurement. Physician-level reliability refers to the ability of a measure to distinguish an individual physician's performance from performance of physicians overall. Good physician-level reliability requires:

1. a sufficient number of patients eligible for a given measure; and
2. performance variation across physicians on that measure (Greenfield S., Kaplan S.H., Kahn R., Ninomiya J., Griffith J.L. 2002; Tucker J. L. 2000; Krein S. L., Hofer T. P. Kerr E. A., Hayward R. A. 2002).

The greater the number of a physicians' patients who are eligible for a measure, the more precise the estimate of the physician's performance will be. When performance variation for a given measure across physicians is limited, the likelihood that a physician's performance is statistically significantly different from his/her peers is also decreased. Hofer et al. showed that not controlling for a measure's physician-level reliability significantly misrepresented performance differences across physicians. Adjusting performance profiles in such a manner, however, is not common place across the health care industry today.

Ensuring that measurement results are both valid and reliable is important when purchasers and plans (and potentially consumers) use the data to make decisions about which physicians get financial rewards or other benefits. But the stakes are particularly high when profiling results are used for public reporting or eligibility for participation in a health plan network. Often, these efforts separate physicians into discrete categories (e.g., high performers, average). Increasing reliability reduces the risk that physicians will be misclassified. Thus, paying attention to the validity and reliability of data will help to ensure that these decisions are made on real differences in performance among physicians rather than any shortcomings of the poor measurement. A recent study using administrative data from nine health plans found that, in typical health plan administrative data, most physicians do not have adequate numbers of quality events to support reliable quality measurement (Scholle et al, 2008). Only 10 of 27 measures tested had reliability estimates above 0.70 at a minimum of 50 events. For other measures, reliability was low even when physicians had 50 events. The largest proportion of physicians who could be reliably evaluated on a single measure was 8 percent for colorectal cancer screening and 2 percent for nephropathy screening for diabetics.

Reliability in the BQI Study

Only a few studies have reported on the reliability of physician-level performance scores derived from public or private sector projects in quality measurement. The Better Quality Information to Improve Care for Medicare Beneficiaries (BQI) Project utilized Medicare Fee-for-Service (FFS) claims data for beneficiaries to calculate measure reliability for five measures that were consistent across all six pilot sites. The five measures used in these analyses are breast cancer screening, colorectal cancer screening, glycosulated hemoglobin (HbA1c) screening for diabetics, low-density lipoprotein (LDL) cholesterol screening for diabetes, and medical attention for intercapillary glomerulonephritis (diabetic nephropathy). This section addresses reliability issues related to these five measures.

The two research questions addressed in these analyses are:

1. How does reliability vary by measure and by different sample size categories?
2. *How does reliability vary by attribution rule?*

Calculating Reliability. The five measures used in the study are discrete variables assumed to be generated from a binomially distributed population.

A variable is binomially distributed if it is the sum of the successes of a series of *t* trials. That is, if a given event with a dichotomous outcome is repeated *t* times and the probability of a success remains constant at *p* (i.e., any trial will yield a success), then the number of successful events is distributed binomially. The assumption that *p* is constant across all trials can be factor limiting the use of the binomial distribution model. This assumption can be relaxed and instead assume that the *p* is beta distributed. This assumption of constant *p* by assuming that *p* is beta distributed results in an extended beta-binomial distribution for the final variable. This process involves first generating a beta distributed value of *p* for each trial, with parameters α and β, then constructing a binomial distribution using this probability (Mooney, 1997). The mean of beta may also be used as the *p* value:

$$\textit{Mean of beta} \quad \mu = \frac{\alpha}{(\alpha + \beta)}$$

The α and β, are used to calculate the physician variance.

$$\sigma^2_{provider-to-provider} = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}$$

The next step is to calculate the binomial variance for the error.

$$\sigma^2_{error} = \frac{p(1-p)}{n}$$

Finally, the physician and error variances are used to calculate the beta-binomial reliability scores. To calculate the α and β parameters we used a SAS macro program. The macro uses the SAS procedure NLMIXED and its facility to perform a maximum likelihood estimation on a general function that is defined by programming statements. The macro output includes estimates of the proportion of successes (with standard error), and a test of difference of this proportion from the expected under the null hypothesis, for both Binomial and Beta Binomial Models. For the later estimates of the parameters gamma and theta are provided which can be used to diagnose the amount of extra-binomial variation that the subjects introduce. Gamma is a useful measure of departure from the binomial model, it is constrained to the interval 0 – 1 and takes a value of 0 when no extra-binomial variation is present (i.e., when the

observed proportions for all subjects are equal).

For each measure, the rate of performance was computed for a physician by dividing the number of successes by the total number of eligible patients attributed to that patient.

In preliminary analyses, we observed very high variation in performance rates among physicians with less than 10 attributed quality events for a measure, so we excluded these physicians from the analysis. For each measure, we determined the number of physicians meeting four different thresholds for the minimum sample size: 10 or more events, 20 or more, 30 or more, and 50 or more. The sample size categories are not mutually exclusive; a physician with 25 events for a measure will be included in the 10 or more category and in the 20 or more category.

To estimate reliability, we used the beta-binomial reliability coefficient. This estimate is computed by dividing the physician-to-physician variance in performance rate by the sum of physician-to-physician variation and the measurement variance (or within-physician variance).

The reliability estimate is a value that is bounded by 0 at the low end and 1 on the high end. Higher reliability increases the likelihood that a physician is assigned to the "right" group minimizing the chances of misclassification. The level of reliability considered "high" is quite subjective. The choice of level raises questions about the tradeoff between feasibility and scientific soundness. A reliability of 0.70 (70 percent) or higher is typically considered acceptable for physician performance measurement purposes (Landis and Koch, 1977; Carmines and Zeller, 1979; Scholle et al, 2008; Nunnelly and Bernstein, 1994). For this analysis, we use 0.70 as a threshold for adequate reliability.

Once we had the estimates of physician-to-physician variance from the beta-binomial model, we estimated what the reliability would be for a particular measure at a particular sample size using the Spearman-Brown formula:

$$reliability = \sigma^2_{physician\text{-}to\text{-}physician} / (\sigma^2_{physician\text{-}to\text{-}physician} + \sigma^2_{error} / n)$$

where the error variance is the binomial variation and n is the sample size (Shrout, P.E. and Fleiss, J.L., 1979). The error variance depends on the performance rate for the physician. We use an average performance rate for each category. Thus, we were able to calculate reliability estimates for each measure, using different minimum sample size requirements and different attribution rules. We calculated reliability separately for the states associated with the six BQI pilot sites.

To examine how reliability varies by measure, we present the reliability estimates based on the one-touch attribution rule for each measure, by minimum sample size category and by state. To examine how reliability varies by attribution rule, we identified the minimum sample size category for each measure which yielded an adequate reliability estimate (based on the threshold of 0.70). These data are presented for each state and each measure across the five different types of physician attribution rules (one-touch, two-touch, 30%, 50%, and maximum frequency).

Reliability Results.

How does reliability vary by measure and by different sample size categories?

Table 1 shows the reliability estimates based on the one-touch rule—this is the least restrictive attribution rule where a patient quality event is attributed to any physician who had at least one health care encounter. For example, the reli-

ability estimate for the breast cancer screening measure with a minimum sample size of 50 events per physician ranges from 0.63 in Minnesota and Wisconsin to 0.79 in California. Thus, only two states meet the threshold for adequate reliability with a minimum sample size of 50 per physician, and no states have adequate reliability for this measure with less than 50 events per physician.

Two measures, colorectal cancer screening and nephropathy screening, reached the recommended reliability level of 0.70 in at least five states. For colorectal cancer screening, reliability was adequate with a minimum sample size of 50 per physicians in 5 states and with 30 observations per physician in 2 states, but not in any states with lower sample size groups. For nephropathy screening, reliability was adequate with a minimum sample size of 50 per physician in 5 states and with 30 observations per physician in 2 states, but not in any states with lower sample size. For the other measures, reliability was adequate with a minimum sample size of 50 for three states for LDL screening, two states for breast cancer screening, and only one state for HbA1c screening. California had adequate reliability with a minimum sample size of 50 for all 5 measures; Arizona for 4 measures, Indiana for 3 measures, Massachusetts and Wisconsin for 2 measures each and Minnesota for only 1 measure.

Table 1. Reliability Under One-Touch Attribution Rule at Different Minimum Observations Categories, By Measure and State

Measure	Minimum Observations Category	Arizona	California	Indiana	Massachusetts	Minnesota	Wisconsin
Breast Cancer Screening	10	0.41	0.42	0.33	0.30	0.30	0.31
	20	0.56	0.59	0.49	0.46	0.46	0.46
	30	0.66	0.69	0.58	0.55	0.54	0.55
	50	0.77	0.79	0.68	0.65	0.63	0.63
Colorectal Cancer Screening	10	0.43	0.44	0.39	0.35	0.24	0.33
	20	0.60	0.61	0.56	0.52	0.38	0.49
	30	0.70	0.71	0.66	0.62	0.47	0.59
	50	0.80	0.80	0.76	0.73	0.59	0.70
Comprehensive Diabetes Care- HbA1c Screening	10	0.26	0.35	0.31	0.29	0.12	0.15
	20	0.39	0.52	0.47	0.44	0.20	0.27
	30	0.49	0.62	0.57	0.54	0.29	0.35
	50	0.61	0.74	0.68	0.64	0.46	0.47
Comprehensive Diabetes Care- LDL Screening	10	0.37	0.33	0.35	0.37	0.24	0.24
	20	0.53	0.49	0.52	0.53	0.37	0.39
	30	0.63	0.59	0.62	0.61	0.44	0.48
	50	0.73	0.70	0.72	0.68	0.52	0.60
Comprehensive Diabetes Care- Medical Attention for Nephropathy	10	0.42	0.43	0.46	0.42	0.36	0.42
	20	0.60	0.61	0.63	0.59	0.53	0.60
	30	0.70	0.70	0.72	0.69	0.65	0.69
	50	0.81	0.80	0.81	0.80	0.76	0.79

Highlighted cells indicate reliability at or above recommended level of 0.70.
HbA1c: glycosulated hemoglobin
LDL: low-density lipoprotein

How does reliability vary by attribution rule?

Table 2 summarizes the reliability results across the different attribution rules. This shows that for breast cancer screening, only two states have adequate reliability with a minimum sample size of 50 using the least restrictive

attribution rule which looks for a single visit with the physician (i.e., one-touch rule). Under more stringent attribution rules such as the maximum frequency rule, reliability is improved. For three states, reliability meets the 0.70 standard at a minimum sample size of 50 and 2 states at a minimum sample size of 30. The same pattern appears for other measures. For example, under the 50 % rule—where events are attributed to the physician who has at least 50% of the patient's visits—all states have adequate reliability estimates for the nephropathy, colorectal cancer screening and LDL screening measures with minimum sample sizes ranging from 20 to 50. In addition, five states achieve adequate reliability for the breast cancer screening measure and four for the HbA1c screening measure (all at minimum sample sizes ranging from 30-50). Five states achieve the reliability of 0.70 with a minimum sample size of 20 on the nephropathy measure.

Table 2. Minimum Number of Observations Needed to Achieve Reliability of 0.70, By Measure, State, and Attribution Rule

Measure	State	Attribution Rule				
		One-Touch Rule	Two-Touch Rule	30% Rule	50% Rule	Maximum Frequency
Breast Cancer Screening	AZ	50	50	30	30	30
	CA	50	30	30	30	30
	IN	-	50	50	50	50
	MA	-	-	-	-	-
	MN	-	-	50	50	50
	WI	-	-	50	50	50
Colorectal Cancer Screening	AZ	30	30	50	30	30
	CA	30	30	30	30	30
	IN	50	50	50	50	50
	MA	50	50	50	50	50
	MN	-	-	-	30	50
	WI	50	50	50	50	50
Comprehensive Diabetes Care- HbA1c Screening	AZ	-	-	50	50	50
	CA	50	50	30	30	30
	IN	-	50	30	30	30
	MA	-	-	50	50	-
	MN	-	-	-	-	-
	WI	-	-	-	-	-
Comprehensive Diabetes Care- LDL Screening	AZ	50	50	30	30	30
	CA	50	50	30	30	30
	IN	50	50	30	30	30
	MA	-	50	30	30	30
	MN	-	-	-	50	50
	WI	-	-	50	50	50
Comprehensive Diabetes Care- Medical Attention for Nephropathy	AZ	30	30	20	20	20
	CA	30	30	20	20	20
	IN	30	30	20	20	20
	MA	50	30	30	30	30
	MN	50	30	30	20	20
	WI	50	30	20	20	20

AZ- Arizona, CA- California, IN- Indiana, MA- Massachusetts, MN- Minnesota, WI- Wisconsin

However, requiring higher sample size reduces the percentage of patients and physicians who will have reportable results as indicated in Tables 3 and 4.

Table 3. Average Percentage of Patients Represented Under Different Minimum Observations Categories and Different Attribution Rules, By Measure

Measure	Minimum Observations Category	Average Percentage of Patients, By Attribution Rule				
		One-Touch Rule	Two-Touch Rule	30% Rule	50% Rule	Maximum Frequency
Breast Cancer Screening	10	82	72	75	63	71
	20	78	63	59	45	52
	30	71	50	41	28	34
	50	51	26	17	9	12
Colorectal Cancer Screening	10	82	75	81	75	81
	20	82	75	80	72	79
	30	82	74	78	70	76
	50	81	71	74	64	70
Comprehensive Diabetes Care- (all three measures)	10	86	79	79	68	75
	20	82	70	63	50	57
	30	74	59	46	34	40
	50	57	37	23	15	18

Table 4. Average Percentage of Physicians Included Under Different Minimum Observations Categories and Different Attribution Rules, By Measure

Measure	Minimum Observations Category	Average Percentage of Physicians, By Attribution Rule				
		One-Touch Rule	Two-Touch Rule	30% Rule	50% Rule	Maximum Frequency
Breast Cancer Screening	10	49	38	31	25	27
	20	36	24	18	13	15
	30	26	14	10	7	8
	50	12	5	3	1	2
Colorectal Cancer Screening	10	63	55	51	45	44
	20	56	48	44	37	36
	30	52	43	38	32	32
	50	45	36	30	25	25
Comprehensive Diabetes Care- (all three measures)	10	48	39	32	26	29
	20	36	26	19	15	16
	30	26	17	11	8	9
	50	13	7	4	3	3

Conclusions

Where performance data are used for profiling physicians, measuring reliability is an important way to determine whether a physician's performance is meaningfully different. However, this report shows that efforts to ensure an adequate level of reliability are fraught with challenges. Reliability varies by measure, by the difference in physician's performance rates, and by the magnitude of the result.

Colorectal cancer screening and nephropathy screening are the most reliable performance measures of the set tested here. Colorectal cancer screening and nephropathy met the adequate sample size requirement with minimum sample sizes of 30 or 50 patients in most states. Increasing minimum sample size requirements and using more stringent attribution rules may help to increase the reliability of measures. For most measures, minimum observations of 30-50 for a physician allows making a reliable distinction in physician ranking. For a few measures, the minimum number of observations needed to achieve a reliability of at least 0.70 may be much higher than 50. However, it must also be acknowledged that requiring a minimum of 30-50 observations for an individual physician may reduce significantly the proportion of physicians who can be evaluated and the proportion of patients who are included in the analysis.

References

Adams, John. The Reliability of Provider Profiling: A Tutorial. Prepared for National Committee for Quality Assurance. RAND Health. September 2008.

Baker G. and B. Carter. Physician Pay-for-Performance Incentive Programs: 2004 National Study Results. San Francisco: Med-Vantage, Inc., 2005.

Carmines EG, Zeller RA. Reliability and Validity Assessment. Beverly Hills, CA: Sage Publications. 1979.

Hofer T. P., Hayward R. A., Greenfield S, Wagner E. H., Kaplan S.H., and Manning W.G. The unreliability of individual physician "report cards" for assessing the costs and quality of care and quality of care of a chronic disease. *JAMA* 1999; 281 (22):2098-105.

Galvin R. and A. Milstein. *Large Employers' New Strategies in Health Care*. New England Journal of Medicine. 2002: 347:939-942.

Greenfield S., Kaplan S.H., Kahn R., Ninomiya J., Griffith J.L. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. Ann Intern Medicine 2002; 136:111-121.

Krein S. L., Hofer T. P., Kerr E. A., Hayward R. A. Whom Should We Profile? Examining Diabetes Care Practice Variation Among Primary Care Physicians, Physicians Group, and Health Care Facilities. *Health Services Research*. 2002;37:1159-1180.

Landis, J. R. and Koch, G. G. (1977) "The measurement of observer agreement for categorical data" in *Biometrics*. Vol. 33, pp. 159–174 1977.

Mooney, Christopher Z. Monte Carlo Simulation. Sage Publications. 1997.

Jum C. Nunnally and Ira Bernstein. Psychometric Theory. 3rd Edition. New York: McGraw-Hill. 1994.

Sarah Hudson Scholle, Joachim Roski, John Adams, Daniel Dunn, Eve A. Kerr, Donna Dugan and Roxanne Jensen. "Reliability of Individual and Composite Measures for Profiling Physician Performance." *American Journal of Managed Care.* 2008 (in press).

Shrout, P.E. and Fleiss, J. L. (1979) "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin*, 2, 420-428.

Tucker J. L. 3rd. The theory and methodology of physician profiling. *International Journal of Health Care Quality Assurance Incorporating Leadership in Health Services* 2000;13:316-321.

Appendix 8

Final Cycle 1 Data Aggregation Report

Centers for Medicare &
Medicaid Services

Better Quality Information to Improve Care for Medicare Beneficiaries

HHSM-500-2006-MD002C
Modification No. MD0008

Delmarva Foundation
For Medical Care, Inc.
9240 Centreville Road
Easton, MD 21601

February 11, 2008

Better Quality Information to Improve Care for Medicare Beneficiaries (BQI) Project

Final Cycle 1 Data Aggregation Report

This final Cycle 1 Data Aggregation report has been prepared by Delmarva for the purpose of providing a comprehensive picture of the activities, accomplishments, and challenges associated with the first round of data aggregation. This report describes the experiences of all six pilots, and contains the following sections:

Executive Summary....3
Summary and Common Themes....5
Measure Selection Challenges and Lessons Learned....7
Working with the Medicare Data Challenges and Lessons Learned....10
Data Aggregation and Analysis Challenges and Lessons Learned....14
Preliminary, Unvalidated Cycle 1 Measure Results....19
Comparing Medicare Results to Other Payers....22
Analyses to Address Methodological Questions....24
Attribution Analyses Preliminary Results....27
Reliability and Minimum Denominator Size....28
Conclusion....29
Appendix A – BQI Project Measures....30
Appendix B – Cycle 1 Measures, by Pilot....32
Appendix C – List of Evaluation and Management codes used to identify physicians eligible to be attributed to beneficiaries....36
Appendix D – Preliminary Analysis Results, Cycle 1....37
Appendix E – Summary Statistics Figures....45

Executive Summary

The Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project is a special study funded by the Centers for Medicare & Medicaid Services (CMS) under the Quality Improvement Organization (QIO) 8th Statement of Work. The BQI Project will analyze, evaluate and report the advantages and challenges of six separate regional pilots ranking the quality of outpatient ambulatory office medical care provided in their geographic regions.

A national system to measure the quality and value of health care would give patients and consumers the information they need to make informed decisions, as well as give care providers the measurements they need to manage and improve their performance. In addition to nationally recognized standard measures of quality, such a reporting system would require accurate and reliable collection of useful information. Experience in collecting and using medical quality information in the outpatient office setting is scarce.

Leaders from the AQA (formerly the Ambulatory Quality Association), the CMS and the Agency for Healthcare Research and Quality (AHRQ) noted by the spring of 2006 that leading organizations in several regions around the United States were already independently developing quality reporting schemes for their markets. The concept of the BQI project grew out of the proposal that the existing experience of these pilot sites could be leveraged into a valuable, transferable knowledge base by:
1) Providing each regional site access to Medicare Fee-for-Service (FFS) data, allowing aggregation of multiple public and private sources of data regarding the care being delivered by physicians to Medicare beneficiaries in those communities, and
2) Evaluating and comparing the results, reliability, advantages and challenges of each existing method of quality measurement and reporting.

CMS assumed the funding and oversight responsibility, and the BQI project began in October, 2006 through CMS' QIO 8th Statement of Work. The BQI project will be completed by October 31, 2008. The Delmarva Foundation for Medical Care (Delmarva), an established QIO for Maryland and the District of Columbia, was awarded the CMS contract for overall BQI project administration and management and the evaluation and reporting of results. Subcontracts were awarded to six regional sites to receive Medicare FFS data, aggregate that data with their existing databases, and report and evaluate the results. Those six sites are:

- Arizona- Center for Health Information and Research, Arizona State University (CHIR)
- California Cooperative for Healthcare Reporting Initiative (CCHRI)
- Indiana Health Information Exchange (IHIE)
- Massachusetts Health Quality Partners (MHQP)
- Minnesota Community Measurement (MNCM)
- Wisconsin Collaborative for Healthcare Quality (WCHQ)

During the BQI project the six pilot sites will each report on twelve quality measures, five of which are consistent across all six sites. The sites use differing techniques for collecting data, various regional data bases, and independent self-designed ranking methodologies which pre-existed the BQI project. In spite of these differences, it is the goal of the BQI project to distill valuable lessons learned from the experiences of these sites, to leverage this experience by combining information from multiple public and private sources, and to transfer this knowledge to the next generation of medical quality measurement and reporting.

By completion of the BQI project on October 31, 2008, each pilot site will have performed data aggregation with Medicare FFS and other data through at least three cycles. The first cycle will report at least five measures from each site, and use calendar year 2005 data. Cycle 1 was completed in November 2007. Cycle 2 of the BQI project will be completed in January 2008 and each site will report results for twelve measures using CY 2006 data. The last cycle will be completed in the late summer of 2008, and will again report twelve measures from each site, this time using CY 2007 data.

This final Cycle 1 Data Aggregation Report has been prepared by Delmarva to provide a comprehensive picture of the activities, accomplishments, challenges and lessons learned from the first of at least three rounds of data aggregation. This report describes the experiences of all six pilots, and contains the following sections:

- Measure Selection
- Data Acquisition & Processing
- Data Aggregation & Analysis
- Preliminary Results
- Analyses to Address Methodological Questions

Summary and Common Themes

Measure Selection

At the inception of the project, it was anticipated that the 26-measure AQA Starter Set of Measures would be used in the project. As the project evolved into the BQI project, with the inclusion of the Medicare data, the number of measures that could be selected from the starter set was significantly reduced. Of the original 26 measures, only 5 were selected for use in Cycle 1. Please see "Measure Selection Challenges and Lessons Learned" for a more detailed discussion of the measure selection process.

There were two principal factors that affected a measure's ability to be used in the BQI project: the availability of the data required to calculate the measure, and a measure's applicability to the Medicare population.

For Cycle 1 data aggregation, eligible measures were limited to those for which the Medicare Fee-for-Service (FFS) datasets were available. Cycle 1 Data Aggregation was based on the 2005 measurement period, during which time there was no Part D benefit for Medicare. Therefore, no pharmacy claims were available for the first data aggregation cycle. This criterion (availability of Medicare FFS datasets), therefore, eliminated any measures that required a hybrid method of data collection (chart review), as well as any measures that required pharmacy claims or laboratory test results. In addition, the BQI project has had to consider the applicability of a measure to the Medicare population, thus measures with an upper age limit of less than 65 years of age, such as Cervical Cancer Screening, were not considered.

Data Acquisition & Processing

Working with the Medicare FFS data was a challenge during Cycle 1 data aggregation. These data were new to all the pilots involved in this project, although some of their aggregators had worked with the Medicare data previously. The datasets provided were large and cumbersome to work with, and the documentation that was available was complicated, and, in some cases, less than complete.

Delmarva and the pilots worked together as much as possible to understand the data and share documentation, tips, and reference information. This effort proved challenging as each of the pilots approached data processing and aggregation slightly differently, and at different points in time, since contract execution and subsequent data release took place over several months instead of all at once.

The challenges we have experienced working with the Medicare data are typical of new datasets. We are encouraged that the pilots were able to make such significant progress in working with the Medicare data, in the relatively short Cycle 1 timeframe. As we continue to work with the Medicare data in data aggregation

Cycles 2, 3, and 4, we anticipate that there will be more lessons learned, which we will continue to share in future reports.

Data Aggregation and Analysis

One of the major goals of the BQI project is to provide recommendations on the most effective methods to aggregate Medicare claims data with data from other payers in order to produce the most accurate and comprehensive measures of the quality of services being provided by physicians to Medicare beneficiaries.

The pilot sites on this project approached data aggregation, that is, combining the Medicare data with other datasets, in two different ways. One site chose to combine the claims data at the patient level (IHIE), and then calculated physician / physician group-level measures based on all available data. The other five sites have calculated the measures at the physician / physician group level for each payer separately, and then combined the numerators and denominators.

Regardless of the approach, one of the largest challenges with aggregation has been in identifying the physicians and defining the groups, using the Medicare data provided. Here too, the approaches have varied, but we have found that the most dependable way to identify physicians and associate them with groups has been to use community-specific provider databases or registries, that have been developed and are maintained by each pilot site. At this point in the project there is considerable concern regarding the use of the Medicare data alone to identify physicians and their group affiliations.

The ultimate goal is for Delmarva and the pilots to recommend standard methodologies that can be used nationally, while still keeping in mind that the pilots' other payer data come from local data sources, such as health plans, that have unique uses and formats.

Preliminary Pilot Aggregation Results

The preliminary and not fully validated, aggregated measure results from the six BQI pilot communities indicate that there are wide variations in results, across pilots as well as across payers. The Medicare FFS data have had a large impact on many of the measure results. Where possible, we have provided some preliminary analysis of results by payer as well as combined, to demonstrate the impact. Further discussion on validation steps can be found in the "Preliminary Results" section.

Analysis to Address Methodological Questions

During the course of cycle 1, Delmarva Foundation and the pilot sites have identified several methodological questions that should be addressed prior to calculating measures at the physician group and individual physician levels. In this report we identify the questions that we hope to address within the scope

of this project, and we provide some preliminary results regarding the selection of an attribution method and the reliability of individual physician level data. At this point, we cannot identify a 'best' attribution method. The preliminary results on reliability indicate that very few physicians have a large enough denominator size to confidently report their results.

Measure Selection Challenges and Lessons Learned

Measure Selection

At the beginning of the BQI project, it was anticipated that the measures to be calculated would be taken from the list of *AQA Starter Set of 26 Measures*[1]. This list represents measures that aligned with agreed-upon parameters and addressed agreed-upon specific conditions and areas[2]. However, while these measures can be used to measure ambulatory care performance in general, the BQI project discovered that not all of them were applicable to the project. The Data Conquerors (a workgroup comprising the technical and analytical representatives from the pilot sites and Delmarva) reviewed each measure in detail for applicability to the project. In evaluating the measures, the BQI project had to consider the availability of data as well as the applicability of the measure to the Medicare population. Once these parameters were applied to the AQA starter set of measures, the following measures were eliminated from consideration:

- Cervical Cancer Screening (not relevant to the Medicare population)
- Tobacco Use (Source data is survey, not administrative / claims)
- Advising Smokers to Quit (Source data is survey, not administrative / claims)
- Influenza Vaccination (Not relevant to the Medicare population as specified; Source data is survey, not administrative / claims)
- Pneumonia Vaccination (Look-back period is indefinite and requires chart review to determine exclusions)
- Use of Appropriate Medications for People w/ Asthma (Measure not relevant to Medicare population)
- Asthma: Pharmacologic Therapy (Measure not relevant to Medicare population)
- Screening for Human Immuno-deficiency Virus (Prenatal care, measure not relevant to Medicare population)
- Anti-D Immune Globulin (Prenatal care, measure not relevant to Medicare population)
- Appropriate Treatment for Children with Upper Respiratory Infection (Measure not relevant to Medicare population)
- Appropriate Testing for Children with Pharyngitis (Measure not relevant to Medicare population)

After it became clear that the AQA starter set of 26 measures would not provide sufficient measures for the project (each pilot site was required to select 12 for the entire project), the pilot sites and Delmarva worked

[1] http://www.aqaalliance.org/files/RevisedStarterSetApril2006.doc
[2] http://www.aqaalliance.org/files/AQAParametersforSelectingAmbulatoryCare.doc

together to identify additional measure sets that could be reviewed for applicability to the BQI project. The measure sets reviewed include the National Committee on Quality Assurance (NCQA) Health Plan Employer Data and Information Set (HEDIS®) measurement set (for which physician-level specifications were in draft) as well as other National Quality Foundation (NQF) endorsed measures, such as those from the Physician Consortium for Performance Improvement (PCPI). In addition, several pilot sites (MNCM, WCHQ) opted to include their measure specifications, which were largely based on existing measure sets such as NCQA HEDIS.

The BQI pilot sites were active participants throughout the measure selection process. Cycle 1 includes a total of seven measures (five per pilot), across all six pilots. The selection of measures varies by pilot; however, three measures are being calculated across all six pilot sites: Breast Cancer Screening, LDL Testing for Diabetics, and HbA1c Testing for Diabetics. Early in the project, the BQI sites were prepared to calculate the Cervical Cancer Screening measure, however, it was ultimately removed from consideration since, with an upper age limit of 64 years of age, CMS deemed it to not be relevant to the Medicare population.

In addition to the three measures that are consistent across all pilots, each community selected an additional two measures for Cycle 1 reporting. The selection of the 4th and 5th measure for each pilot was based not only on data availability, but also relevance to and acceptance by the local community. For example, some pilots have access to laboratory results data either from the labs themselves (IHIE) or through their data collection methodology (WCHQ and MNCM), and are therefore able to include disease control measures in their list. In Cycle 2, pilots will calculate seven additional measures, and "refresh" the five measures from Cycle 1. By the end of the project, five measures will be consistent across pilot sites. The full list of measures that will be calculated throughout the project can be found in Appendix A.

While the measures across the pilots appear to be the same, there are differences across pilots in the detailed measure specifications. These variations are important to note, and must be taken into consideration when conducting any analyses on all the data, as these variations severely limit any comparisons across the pilot sites.

Measures were chosen for inclusion in BQI based not only on those that would be consistent across sites, but measures that were applicable to the local community. Since the pilots are aggregating the Medicare FFS data along with other payer data to which they have access, they must apply the same measure specifications that are used for the other payer data to the Medicare data. In addition, since Cycle 1 aggregation was the first cycle, it was understood and accepted by all participants that there might be differences across the pilots. Each pilot's selection of measures and specifications was discussed on the Data Conqueror's calls, and the group accepted these differences, acknowledging that it was important to address local community issues throughout the project.

The specification differences can be summarized as follows. For additional details, please see Appendix B.

- Performance measure calculation method:

  Four out of the six Pilots are using administrative (e.g., claims) data only for all Cycle 1 measures (CHIR, CCHRI, IHIE, and MHQP). Two are using a combination of administrative data and hybrid data collection (e.g., chart review) (MNCM, WCHQ).

- Performance measure specifications:

  Four out of the six Pilots (CCHRI, IHIE, MHQP, MNCM) are using HEDIS 2006 based specifications for Cycle 1 measure calculations, except for the LDL Screening for Cardiovascular Disease measure, which is following the HEDIS 2007 specifications, because of a change in the exclusion diagnosis criteria. One pilot (CHIR) uses a specific software package (EBM Connect, from Symmetry) as well as its own primary care physician imputation method. The software package specifications are fairly close to HEDIS 2006 specifications. The sixth pilot, WCHQ, has a different approach to aggregation and attribution; however, its actual measurement specifications are close to HEDIS 2006.

- Populations covered:

  All Pilots are incorporating the Medicare FFS data into their results. Four out of six pilots include some Medicaid data (FFS and/or Managed Care) in their results (CHIR, IHIE, MNCM, and WCHQ). Three pilots include commercial, such as Health Maintenance Organization (HMO), Point of Service (POS), and Preferred Provider Organization (PPO) data (CCHRI, MNCM, MHQP). One has commercial fee-for-service and managed care data (IHIE), and another includes all payers (WCHQ). See Appendix B for details by pilot.

- Type of data:

  The pilots vary in the types of data they include in the measure calculations. Some have access to point of care data, while others have access to only claims data. See Appendix B for details by pilot.

- Population age ranges:

  While the age ranges covered by the measures are fairly consistent, the range for the Breast Cancer screening measure does vary widely.

- Performance measure measurement periods:

  The measurement periods are mostly consistent across the Pilot sites for all measures. One pilot (CHIR) is using a 27-month time period for identifying diabetics (versus 2 years in the other pilots). For Breast Cancer Screening, one pilot (WCHQ) is using 3 years' worth of data to define the denominator, whereas the others are using a 2-year period.

- Medical professionals and specialties:

  There are variations in the medical professionals to whom care can be attributed. In addition, the specialties included in the measure specifications vary. For the most part, pilots are including Doctors of Medicine and Osteopathy in their attribution methodology. Some also include Physician Assistants and Nurse Practitioners. Others only attribute care to Doctors of Medicine and Osteopathy, but give these medical professionals "credit" for care provided by other medical professionals.

While it was challenging to select measures that were applicable and appropriate for this project, we are encouraged by both the consistency and the variation. Having consistent measures will allow us to test the effects of minor specifications differences on the measures. At the same time, having variation in the measures will allow us to provide CMS with a clearer picture of which measures are most appropriate for future work.

Working with the Medicare Data Challenges and Lessons Learned

Data Acquisition

The process for acquiring the Medicare FFS data actually began in December 2005, when Nancy Wilson started a dialog with CMS Office of Information Services (OIS) regarding getting data to the pilot communities. Delmarva became involved in May 2006 to clarify the data use agreement (DUA) process. At that time, claims data were expected to be distributed on a "rolling" basis to some of the sites and to be distributed as soon as they were available, and analysis and measurement would be performed on a more frequent basis than annually or semi-annually. Arranging for some sites to receive the monthly TAP file (and be responsible for processing adjustments) was envisioned to work around the 3-6 month lag that was required to have 95% complete final action paid claims. In addition, at the beginning of the project, efficiency measures were under consideration, which would require detailed financial information.

Throughout the process of specifying the data, there was considerable discussion regarding the execution of Data Use Agreements (DUAs) and data flows. As such, CMS and Delmarva determined that the most efficient way to receive the data would be to request it directly from OIS. Therefore, on October 5, 2006, Delmarva submitted the initial data request to CMS, via the government task leader. Delmarva also worked to get the file specifications for distribution to the pilot sites in advance of the datasets themselves, so that pilots could do as much as possible to be prepared to process the files upon receipt. By mid-December 2006, Delmarva had prepared and submitted the specifications for the beneficiary information required by the project. CMS prepared the claims files first, but would not release them until contracts had been signed.

In mid-December, Delmarva became aware of a newly released CMS policy for encryption of portable media (including tape cartridges), effective January 1, 2007. This new requirement would result in additional data acquisition delay (all project timelines had been based on an initial data receipt date of October 1, 2006); therefore Delmarva requested that all the pilot sites' data be sent to Delmarva for subsequent release to the pilots. Delmarva received, catalogued, and secured the data in late December 2006. In addition, Delmarva pulled the data off the tapes prior to release for its own analyses.

After subcontracts were executed, Delmarva was able to release the Medicare data to each pilot site or its data aggregator. These release dates varied from late December 2006 to late April 2007, as shown in Table 1. Unless otherwise noted, the files were created by CMS and provided on 3490E tape cartridges. The datasets included:

- Standard Analytic Files (SAFs) for Inpatient, Outpatient, and Carrier (Physician/Supplier, Part B) claims, for calendar years 2003, 2004, and 2005, based on provider state of record
- Master Physician file, all physicians
- Physician group membership file, all physician groups
- Group UPIN file, all physician groups
- Provider of services file, all non-physician providers (provided on CD)
- Clinical laboratory file, all laboratories (provided on CD)
- Beneficiary denominator files, for 2003, 2004, 2005, based on beneficiary state of residence (provided on CD)

This initial data delivery did not include the additional beneficiary information requested, such as full-length name fields, Social Security Number, and other key identifiers, which several pilots needed to perform aggregation. As the Quality Improvement Organization (QIO) for Maryland and the District of Columbia, Delmarva has access to the CMS Medicare Beneficiary Database; therefore, Delmarva extracted the beneficiary supplemental information for each pilot site. Subsequently, it was discovered that by requesting claims based on provider state, the claims files had not been cross-referenced to be able to track beneficiaries. While a beneficiary's Health Insurance Claim (HIC) number is unique to an individual, an individual's HIC may change over time. The process used to identify all the HICs a beneficiary may have had over time is referred to as the HIC cross-reference procedure and the resultant file is the HIC Cross Reference File. Therefore, Delmarva also extracted the cross-referencing information from the Medicare Beneficiary Database. Note that in subsequent data aggregation cycles, claims have been requested based on a specific set of Medicare beneficiaries, and the cross-referencing process occurs when the claims are extracted, therefore this extra dataset is no longer needed.

Table 1: Subcontract Execution and Data Release Timeframes

Pilot	Subcontract Execution	Data Release Date
Wisconsin Collaborative for Health Care Quality (WCHQ)	12/29/2006	12/29/2006
Minnesota Community Measurement (MNCM)	01/02/2007	01/03/2007
Indiana Health Information Exchange (IHIE)	01/31/2007	02/02/2007
Arizona State University, Center for Health Information & Research (CHIR)	03/09/2007	03/12/2007
*California Cooperative Healthcare Reporting Initiative (CCHRI)	03/03/2007	03/28/2007
*Massachusetts Health Quality Partners (MHQP)	02/05/2007	04/30/2007

Note: Four sites contracted with an external data aggregator; two of these sites (CA and MA) requested data be shipped directly to their data aggregator, which Delmarva delayed until subcontracts and DUAs were executed between the site and its data aggregator.

The files listed above have presented the following significant challenges:

- The SAFs contain much more data than is required to calculate measures, and are therefore very large and cumbersome. The pilots have effectively determined which fields are required. However, after much discussion on the Data Conquerors calls, only one pilot (MNCM) opted for a different file that contained less data elements, for subsequent cycles. The remaining five pilots determined that it would be less efficient and cost-effective to change data formats in the middle of the project.
- The SAFs, master physician file, and physician group membership files were delivered on 3490E tape cartridges. Only one pilot site was able to process all the data using the tape cartridges with no additional support from the QIO (CHIR). Two sites could not handle the tapes but could process the files in their "native" format (WCHQ, MNCM). Two sites (MHQP, IHIE) required conversion of all files. The sixth site's subcontractor could handle the tapes but had to request three replacement files from Delmarva because of tape failures or corrupted files (CCHRI). It is important to note that in the second data aggregation cycle, encrypted files were delivered on DVD, which helped. However, in the third cycle the files were delivered on encrypted tape cartridges.
- Using the cross-referencing information provided by Delmarva in conjunction with the other files provided by CMS has posed an analytic challenge. Delmarva and the pilots worked together to identify questions and discuss approaches to linking all the datasets together efficiently and effectively. Please note that this challenge has been overcome in subsequent cycles, as the claims data are extracted based on a specific set of HICs, and the cross-referencing process occurs at the time the claims were extracted.

Specific comments and challenges:

The following are specific comments as submitted by the pilots in their Cycle 1 Data Aggregation reports:

- The data file they received contained "too much data," which caused the aggregator to "spend extended time trying to understand the data and all the variables," and "spend extended time waiting for SAS code to run on the large dataset." In addition, the data aggregator indicated that the "data

dictionary [provided by CMS with the files] was not specific to this project," and there was "no clear logic on how to link all the data." (MNCM)

- "We had difficulty matching the beneficiary master [denominator] file to the claim files." The records needed by the aggregator require specific beneficiary identifiable information to be associated with the claims. (IHIE)
- "Sometimes the metadata [file documentation] was clear and copious, other times obscure and scanty… sometimes the metadata was inconsistent about how it described field types from one file to another; [we had to work out] by discovery, since no linking schema was included, the actual relational links between all of the files." (WCHQ)
- There were "differences between the logical file structures: Carrier files contain one procedure per record, whereas the [inpatient and outpatient claims] files contained an entire stay in one record." (WCHQ)
- There were "a variety of date field structures which required resolution and fixing – e.g. the same date might be supplied in different files as different types (some text, some numeric)… Fortunately by examining the data itself we recognized the problem and devised ways of restoring the dates to make them usable." (WCHQ, IHIE)
- "It took some time to fully grasp the coding identification scheme(s) [for caregiver group identification. For example,] upin6+upin4 codes are a local practice site, whereas the upin6 code alone refers to the billing site which might include a variety of practice sites. Similarly, [in the member file,] mem_upin6 refers to an individual physician, but each time a physician joins a different or another practice site s/he receives a new mem_upin4 code to go along with the static upin6 code. In other words, a physician who is initially assigned a [state] upin6+upin4 code will continue that upin6 code but receive a new upin4 postfix and another record in the data file each time they join a new practice site, including if they join several sites at the same time rather than serially over time." (WCHQ)
- "The original CMS data was selected based on all claims for [state] physicians/caregivers… consequently… where [state] physicians saw beneficiaries from other state, [we] were unable to make a "match" with that patient in the denominator files. Following further analysis [we] determined that approximately 12% of the claims for any given year fell into this unidentified patient category. This issue is an issue for all BQI sites… in the short term [we have] removed the 'unidentifiable' claims… [and the] quality measures analysis continues on identifiable claims." (WCHQ)
- "The primary difficulty was in properly structuring the provider tables and relating it to the provider information in the carrier and facility tables. There was confusion relating to whether to use a Tax ID Number (TIN) or the Universal Physician ID Number (UPIN). For our proposed data aggregation methodology, we also needed to have a single record of information (e.g., specialty) for each provider, which required additional processing of the provider tables as they allowed for multiple specialties at different locations." (CHIR)

Data Aggregation and Analysis Challenges and Lessons Learned

Aggregation Approach

The pilot sites have approached data aggregation; that is, combining the Medicare data with other datasets, in two distinct ways.

One site (IHIE) chose to combine the claims data at the patient level, and then produced provider reports based on all available data. All the other pilot sites calculate the measures at the physician group / individual physician level for each data source and/or payer separately, and then add the numerators and denominators to get an aggregated measure result. In some cases, pilots chose this approach because of the known lack of a unique patient identifier across payers. One pilot (CHIR) chose to take this approach to address the data use restrictions within the contract, which did not permit use of the CMS data outside the scope of BQI. Therefore, this pilot opted not to include the Medicare data in its existing multi-payer database.

In this latter method of aggregation, patients may be double-counted if they are in the denominator for multiple payers. This method may also interfere with the attribution methods that require identifying all of the E&M claims for a beneficiary. For example, if a beneficiary has 3 claims from one physician from one payer and another 3 claims from a different physician and a different payer, we will not be able to appropriately identify when those beneficiary-physician links reach the 30% or 50% thresholds needed for 2 of the attribution rules. It is not yet known the extent that patients may be double-counted, however, we intend to present any findings by the end of the BQI project. Currently, the only identified pitfall or disadvantage to the IHIE approach to data aggregation (patient level) is that the approach could potentially require larger data processing resources.

Minnesota Community Measurement (MNCM), Indiana Health Information Exchange (IHIE), Wisconsin Collaborative for Healthcare Quality (WCHQ), and the California Cooperative Healthcare Reporting Initiative (CCHRI), each subcontracted parts of data aggregation. MNCM subcontracted with Stratis Health, the Minnesota Quality Improvement Organization (QIO), to process the Medicare FFS data, extract the relevant variables, and calculate Medicare FFS measure rates at the physician group level. IHIE processed the Medicare FFS data at the patient level, mapped it to a standard, synchronized format internally, then converted the data to a standard HL7 (Health Level 7) format that its aggregator, Regenstrief Institute (RI) requires. The data are then included in the Indiana Network for Patient Care (INPC) database system. WCHQ contracted with the Center for Health Systems Research & Analysis (CHSRA) at the University of Wisconsin-Madison to process the Medicare FFS data and apply the WCHQ measurement methodology to the FFS data. CCHRI contracted with Thomson Healthcare to perform its data aggregation and analysis. MHQP contracted with ViPS to perform data aggregation and calculate measure results.

Analytic Considerations

Calculating Measures

Another challenge we have experienced has been working through the measure calculation process. We have worked through several analytic issues specific to the Medicare beneficiaries:

Claims files specifications:

The initial claims files were pulled by the provider state of residence. Pulling the data in this fashion has introduced what we have termed the "snowbird/robin/border doctor" issue, as the claims files may be missing any or all of the following claims:

- For any beneficiaries serviced by in-state providers, claims for services they received from an out-of-state provider. ("snowbird" issue)
- For any in-state resident beneficiaries, claims for services they received from an out-of-state provider. ("snowbird" issue)
- For any providers who are part of the pilot communities whose state of residence is across the border from the pilot state, claims for services billed by providers in the border state.

We have not determined the extent of this issue nor the exact impact on the measure rates, but are investigating approaches to analyzing this issue further, perhaps using Wisconsin and Minnesota data. For Cycle 2, the pilots and Delmarva jointly decided to use a "beneficiary-centered" approach to pulling the data.

Beneficiaries under 65:

Delmarva found beneficiaries in the 2005 denominator files that were under the age of 65. Since we have always limited our measure selection to measures whose age bands went above the age of 65, we questioned whether we should reconsider that decision. After consultation with CMS management, it was decided that beneficiaries who are under the age of 65 should be included in all measures that are applicable to the >=65 population, if the age range for the measure begins at less than 65.

Institutionalized beneficiaries:

There was considerable discussion regarding the inclusion or exclusion of institutionalized beneficiaries in the measure denominators. There was concern that a primary care provider for a beneficiary in an institutionalized setting may no longer be able to deliver the care once in the setting, but would still be held accountable. The pilots and Delmarva decided that they would investigate ways that institutionalized beneficiaries might be excluded from the Medicare FFS data; however, the group felt it might be

challenging. This issue will be further explored during Cycle 2 aggregation. Delmarva requested the Skilled Nursing Facility (SNF) claims for Cycle 2, to aid in the investigation of this issue.

Continuous Enrollment:

The pilots and Delmarva are using the Healthcare Effectiveness Data and Information Set (HEDIS) specifications from the National Committee on Quality Assurance (NCQA) as the basis for measure calculation. However, in some cases several refinements have been made to adapt the continuous enrollment criteria to what is available in the Medicare data. For example, CCHRI is applying the criteria as follows:

- Beneficiaries were required to have 11 months enrollment in Part A and Part B for both 2004 and/or 2005, wholly dependent upon whether Cycle 1 measure needs one or two years of continuous enrollment.
- Beneficiaries had to be enrolled in the anchor month – December 2005 (however, Medicare Health Maintenance Organization (HMO) enrollment in the anchor month did NOT count as enrollment). Also excluded were any beneficiaries that died in December 2005.
- The one-month gap in either year could be due to either no Medicare enrollment, or enrollment in Medicare HMO for that single month.

Look-back periods and claims availability:

For the Colorectal Cancer Screening measure, pilots have expressed some concern regarding the long look-back period (10 years) suggested by the HEDIS specifications. The NCQA has conducted some field testing to see the effect of enrollment on the measures. The results of this testing are shown in Table 2.

Table 2: Any Screening Test for Colorectal Cancer, Based on Administrative Data, for Varying Lengths of Enrollment

Continuous Enrollment Period	Plan A	Plan B	Plan C	Plan D
2 year	35.9	47.2	27.3	44.4
3 year	36.4	48.7	27.3	45.9
4 year	37.3	49.8	28.7	47.9
5 year	38.9	50.7	30.3	48.8
7 year	38.2	51.4	31.4	51.2
10 year	37.2	NA	NA	49.1

Note: Data provided by NCQA - 2003 HEDIS measure field tests. Insufficient sample size for Plans B and C at 10 year continuous enrollment interval to calculate rates.

These data suggest that increasing the look-back periods increases the colorectal screening rates. Further investigation of this issue is warranted and will be conducted as possible in subsequent data aggregation cycles

Identifying physicians and groups:

One of the largest challenges with aggregation has been in identifying the physicians and defining the groups, using the Medicare data provided since the pilots are expected to generate multi-payer results at the individual physician, physician group level, or both. Regardless of the aggregation methodology chosen, it is critical that physicians and physician groups be accurately identified across all data sources, to ensure that measure results are correct and credible when reported at the physician group level.

Two of our pilots have attempted to use the combination of the Master Physician, Group, and Member files provided by CMS to identify and define physician groups (CHIR, WCHQ). Both sites have reported challenges with this approach. For example, one site found that the member file, which is supposed to link individual physicians with a group, included members who were no longer affiliated with a group. They also found that many in-state providers were being included in out-of-state provider groups.

One pilot that has a robust provider database added the Medicare information by matching on last name + UPIN, State license, DOB, or other identifiers (MHQP). They found that ~50% of MDs did not match to their database. None of a sample of non-matching MDs was found listed in the state Board of Registration in Medicine. Another pilot has also developed, in conjunction with its participating physician practices, a provider database that contains a single global identifier and all other possible national identifiers, such as UPIN, National Provider Identifier (NPI), Employer Identification Number (EIN), and Drug Enforcement Agency (DEA) number (IHIE).

In another pilot, we discovered that the groups, as defined by the pilot (and confirmed by the groups themselves) could not necessarily provide the correct UPIN for identification. By also incorporating the TaxID into its analysis, this pilot has gotten closer to the correct composition (WCHQ).

Yet another site used a "pre-determined" set of TaxIDs and only used the claims that included those TaxIDs in its analysis (MNCM). This pilot was comfortable with the measure results using this methodology.

Specific comments:

The following are specific comments as submitted by the pilots in their Cycle 1 Data Aggregation reports:

- Linking large volumes of claims data, where the data do not contain a single consistent and unique identifier for each physician, is significantly challenging. Inconsistencies and incompleteness of the physician data (e.g., individual physician names and groups entered into the same field, non-

populated unique identifier fields such as SSN, DEA, and UPIN) further complicate this process. There is concern about how well the claim's rendering physician identifiers from all data sources will link to their respective provider files and among all files entering linkage. Reporting linkage results between all professional claims received by a data supplier and their provider files could be misleading. Consequently, the physician linking process is extremely time-intensive, both in terms of computer processing time as well as human processing time. Further, the process is sensitive to anomalies in the data, which may impact the linking outcomes. A preliminary Master Physician List (MPL) was created based on the automated and first round of manual review. However, due to the issues noted above in the data, a number of physician records were still not able to be linked, and some group records were still flagged as physician records. A second round of manual reconciliation was required; these results will be incorporated into the creation of the final MPL of all unique physicians identified across the four data sources. (CCHRI)

- Issues were encountered when attempting to roll individual providers to their provider group. It appears that the "member" table which links individual providers to a group does indicate members who are no longer affiliated with the group, as many Phoenix-area providers were being included in out-of-state provider groups, often in areas where the provider performed their residency. This issue has not yet been resolved. Linking the providers across the two databases also posed some challenges. Because provider UPINs were not available in AZHQ data, it was necessary to link providers using other available information, including license number, name, and location. This involved a multi-step process with several manual reviews of potential matches. (CHIR)
- It took some time to fully grasp the coding identification scheme(s) [for caregiver group identification. For example,] upin6+upin4 codes are a local practice site, whereas the upin6 code alone refers to the billing site which might include a variety of practice sites. Similarly, [in the member file,] mem_upin6 refers to an individual physician, but each time a physician joins a different or another practice site s/he receives a new mem_upin4 code to go along with the static upin6 code. In other words, a physician who is initially assigned a [state] upin6+upin4 code will continue that upin6 code but receive a new upin4 postfix and another record in the data file each time they join a new practice site, including if they join several sites at the same time rather than serially over time. (WCHQ)

The challenges reported by the pilots in accurately identifying physicians and physician groups, using the CMS supplied data alone, suggest that a national implementation with just Medicare data may generate results that are not credible in local healthcare communities. The implementation of the National Provider Identifier may ease some of the concerns, however, full and exclusive use of this identifier will not occur in the timeframe of the BQI project. Additional detailed investigation into the accuracy of the Medicare provider data is ongoing.

Preliminary, Unvalidated Cycle 1 Measure Results

Despite the limitations and challenges encountered, the BQI pilot sites have been able to generate measure results, at either the physician group or individual physician level, and have provided de-identified results to Delmarva. Descriptive statistics for these results are presented in Tables 3a – 3h. Unless otherwise noted, the results presented are aggregated, multi-payer results. Please see Appendix E.

Table 3a: Indiana Health Information Exchange, Cycle 1 Descriptive Statistics

Indiana Health Information Exchange (Physician Group Level) Timeframe: Current through May 31, 2007					
Measure	Average Rate	Standard Deviation	Rate Range	Denominator Range	Number of Groups Reporting
Colorectal Cancer Screening	25.83%	16.61%	0.00 - 82.68%	30 - 332	321
HbA1C Testing for Diabetics	20.88%	16.78%	0.00 - 63.29%	30 - 304	134
LDL Testing for Diabetics	40.30%	24.67%	1.39 - 86.08%	30 - 304	134
Nephropathy Testing for Diabetics	50.00%	14.46%	24.05 - 92.31%	30 - 304	134
Mammography Screening	43.07%	18.24%	0.00 - 86.15%	30 - 216	229

Note: Medicare FFS data through December 31, 2005 only. Minimum denominator of 30 for reporting.

Table 3b: Massachusetts Health Quality Partners, Cycle 1 Descriptive Statistics

Massachusetts Health Quality Partners (Physician Group Level) Timeframe: Calendar 2005					
Measure	Average Rate	Standard Deviation	Rate Range	Denominator Range	Number of Groups Reporting
Colorectal Cancer Screening	46.16%	12.75%	16.4% - 79.0%	10 - 19,988	154
HbA1c Testing for Diabetics	84.13%	7.46%	49.1% - 96.1%	12 - 3,471	147
Eye Exams for Diabetics	58.76%	7.38%	35.4% - 83.8%	12 - 3,471	147
LDL Testing for Diabetics	87.96%	8.26%	47.2% - 100.0%	12 - 3,471	147
Mammography Screening	76.84%	9.87%	32.3% - 92.3%	12 - 5,647	150

Table 3c: Minnesota Community Measurement, Cycle 1 Descriptive Statistics

Minnesota Community Measurement (Physician Group Level) Timeframe: Calendar 2005					
Measure	Average Rate	Standard Deviation	Rate Range	Denominator Range	Number of Groups Reporting
HbA1C Testing for Diabetics	80.87%	11.96%	41.85 – 98.24%	59 – 3,577	56
LDL Testing for Diabetics	83.77%	9.95	52.01 - 97.71%	49 – 3,819	56
Eye Exam for Diabetics	64.83%	8.85%	37.49 - 88.90%	43 – 2,794	56
Nephropathy Testing for Diabetics	68.84%	11.91%	25.40 - 98.21%	48 – 3,246	56
Mammography Screening	77.36%	10.69%	41.85 - 97.71%	18 – 18,857	112

Table 3d: Wisconsin Collaborative for Healthcare Quality, Cycle 1 Descriptive Statistics

Wisconsin Collaborative for Healthcare Quality (Physician Group Level) – WCHQ Results Timeframe: Calendar 2005					
Measure	Average Rate	Standard Deviation	Rate Range	Denominator Range	Number of Groups Reporting
Colorectal Cancer Screening	4.01%	11.57%	0.0% - 100.0%	1 – 63,606	2,365
HbA1c Testing for Diabetics	19.46%	28.61%	0.0% - 100.0%	1 – 7,625	805
HbA1c Control for Diabetics (note)	48.26%	7.22%	35.1% - 63.1%	370 – 25,374	18
LDL Testing for Diabetics	24.89%	32.74%	0.0% - 100.0%	1 – 7,625	805
Mammography Screening	6.56%	14.44%	0.0% - 100.0%	1 – 16,094	2,092

Note: HbA1c control is based on WCHQ reporting groups (not CMS-data-defined groups), contains all patients regardless of payer, and is the most recent 12 months ending June 30, 2006. All other measures are based on the Medicare FFS data only.

Table 3e: Center for Health Information & Research, Cycle 1 Descriptive Statistics

Center for Health Information & Research - (Physician Level) Timeframe: Calendar 2005					
Measure	Average Rate	Standard Deviation	Rate Range	Denominator Range	Number of Physicians Reporting
Colorectal Cancer Screening	37.75%	15.69%	0.0% - 98.50%	8 - 976	1,315
HbA1c Testing for Diabetics	69.02%	24.88%	0.0% - 100.0%	1 - 376	1,291
Eye Exam for Diabetics	41.20%	19.69%	0.0% - 100.0%	1 - 329	1,297
LDL Testing for Diabetics	78.44%	20.92%	0.0% - 100.0%	1 - 376	1,291
Mammography Screening	59.07%	22.41%	0.0% - 100.0%	1 - 181	1,315

Note: Minimum denominator of 1 for inclusion in statistics.

Table 3f: CCHRI, Cycle 1 Descriptive Statistics, Attribution Method 1

California Cooperative Healthcare Reporting Initiative - (Physician Level)
Attribution Method 1: All physicians of any specialty – one E & M visit
Timeframe: Calendar 2005

Measure	Average Rate	Standard Deviation	Rate Range	Denominator Range	Number of Physicians Reporting
Colorectal Cancer Screening	36.29%	25.71%	0.0% - 100.0%	1 - 2,075	63,578
HbA1c Testing for Diabetics	69.76%	29.73%	0.0% - 100.0%	1 - 561	47,811
LDL Testing for Diabetics	78.60%	27.84%	0.0% - 100.0%	1 - 561	47,811
Cardiovascular LDL Testing	72.40%	31.87%	0.0% - 100.0%	1 - 326	35,598
Mammography Screening	69.25%	27.95%	0.0% - 100.0%	1 - 691	54,606

Table 3g: CCHRI, Cycle 1 Descriptive Statistics, Attribution Method 2

California Cooperative Healthcare Reporting Initiative - (Physician Level)
Attribution Method 2: Any physician of relevant specialty – one E & M visit, specific specialties
Timeframe: Calendar 2005

Measure	Average Rate	Standard Deviation	Rate Range	Denominator Range	Number of Physicians Reporting
Colorectal Cancer Screening	35.56%	22.24%	0.0% - 100.0%	1 - 1,248	26,527
HbA1c Testing for Diabetics	70.63%	28.17%	0.0% - 100.0%	1 - 561	21,831
LDL Testing for Diabetics	79.78%	25.96%	0.0% - 100.0%	1 - 561	22,378
Cardiovascular LDL Testing	72.85%	31.05%	0.0% - 100.0%	1 - 326	16,850
Mammography Screening	69.53%	25.81%	0.0% - 100.0%	1 - 691	26,154

Table 3h: CCHRI, Cycle 1 Descriptive Statistics, Attribution Method 3

California Cooperative Healthcare Reporting Initiative - (Physician Level)
Attribution Method 3: Single physician of relevant specialty, largest number of E & M visits, specific specialties
Timeframe: Calendar 2005

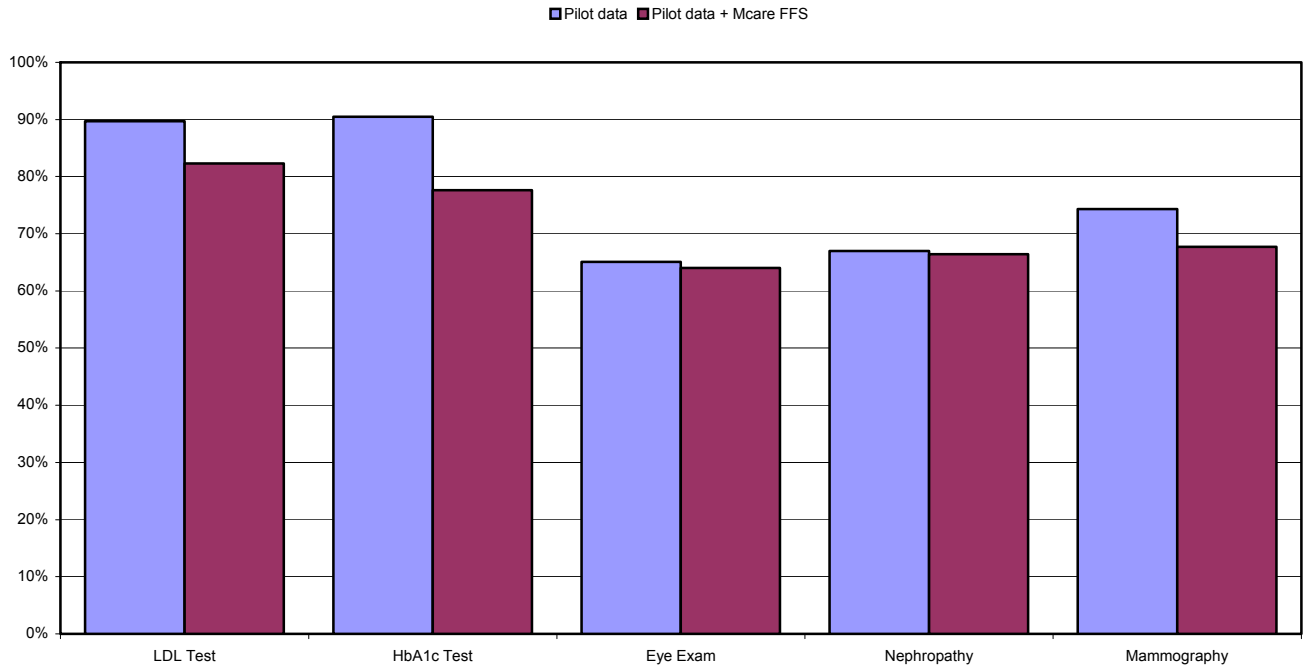
Measure	Average Rate	Standard Deviation	Rate Range	Denominator Range	Number of Physicians Reporting
Colorectal Cancer Screening	32.58%	23.85%	0.0% - 100.0%	1 - 865	48,068
HbA1c Testing for Diabetics	71.83%	30.63%	0.0% - 100.0%	1 - 361	18,230
LDL Testing for Diabetics	80.64%	27.83%	0.0% - 100.0%	1 - 360	18,230
Cardiovascular LDL Testing	75.25%	32.55%	0.0% - 100.0%	1 - 238	12,308
Mammography Screening	68.08%	28.77%	0.0% - 100.0%	1 - 293	22,554

Comparing Medicare Results to Other Payers

Preliminary results from the pilots after the first cycle of data aggregation indicate that including the Medicare FFS data is having an impact on measure results as well as attribution rates. These results warrant additional analysis and investigation, which is taking place now and will continue to take place throughout the rest of the project.

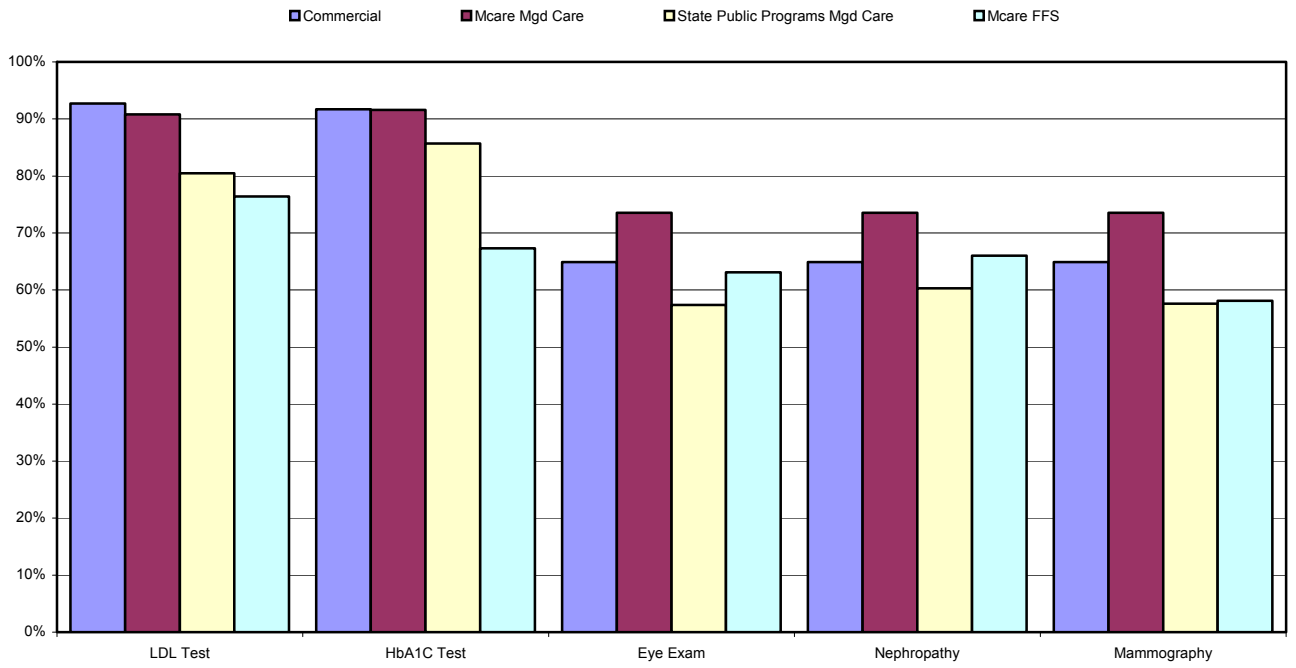
In one pilot, including the Medicare FFS results with the other data resulted in a decrease in all the measures, for the community-wide rates, as shown in Figure 1.

Figure 1: One Pilot's Experience: Impact of Medicare FFS on Community Rates



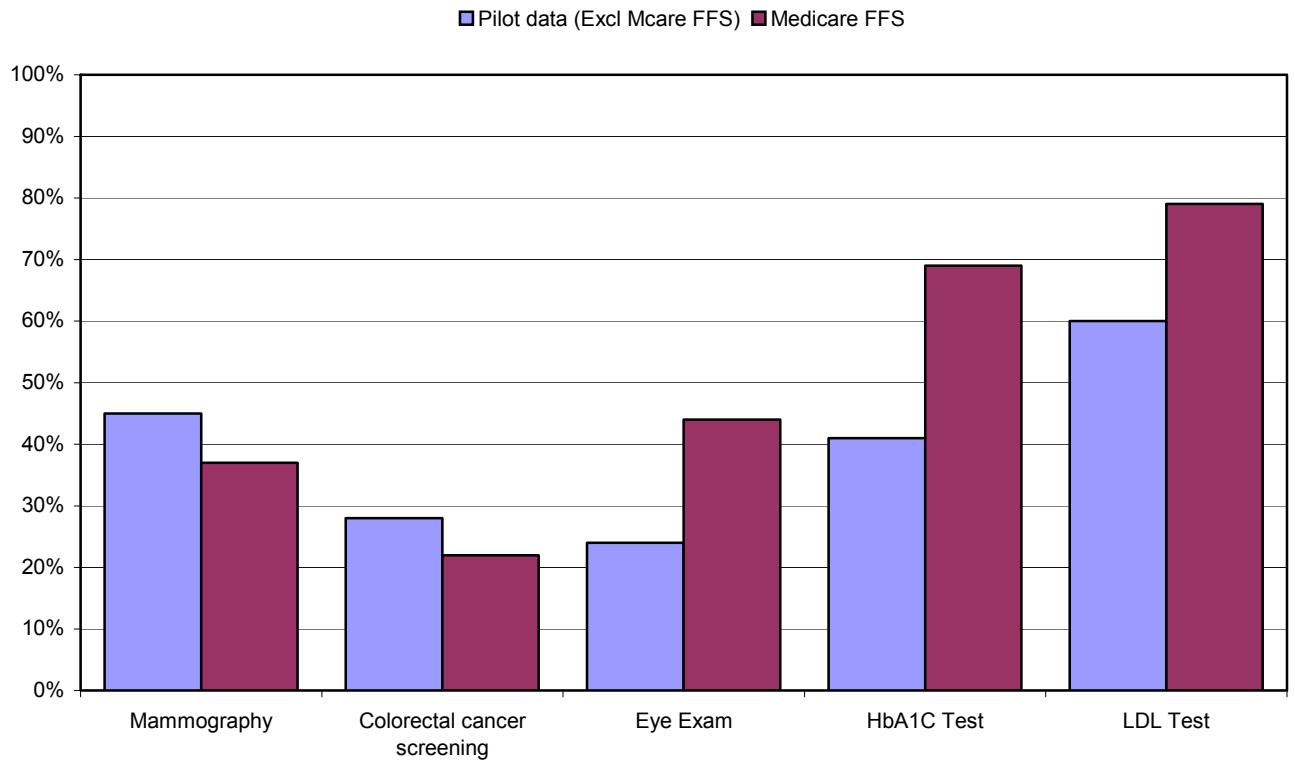
This same community was able to stratify the measure results by payer. Medicare FFS was the lowest rate for two of the measures, and lagged behind the Medicare managed care rate for all the measures, as shown in Figure 2.

Figure 2: One Pilot's Experience: Rates, by Payer/Data Source



Another pilot's Medicare FFS results were actually higher than their other data, for the diabetes measures. Both prevention measures, however, were lower, as shown in Figure 3.

Figure 3: Another Pilot's Experience: Rates, by Payer/Data Source



Analyses to Address Methodological Questions

Delmarva Foundation and the pilot sites have identified several choices regarding methodology that must be made in the process of calculating measures and the physician group and individual physician levels. Through discussion with CMS and the pilot sites we have focused these choices into seven research questions that we believe we can address within the BQI contract. The seven questions are:

1. What are the considerations for selecting a physician to physician group attribution method?
2. What denominator sizes are required for sufficient reliability of measures?
3. What are the considerations for selecting a patient to physician attribution method?
4. What are limitations of methods available for making data requests, i.e. dealing with snowbirds/border physicians, SNF data, lag time of data, age of beneficiaries?
5. Should physicians believe that claims based measures accurately reflect their performance level?
6. What correlations are there between physician group measures for Medicare and physician group measures based on commercial or Medicaid?
7. What are the benefits and limitations of reporting data in a composite measure?

We have some preliminary results of Questions 2 (reliability) and 3 (attribution method) to include in this report.

Attribution Method

There are differences across the pilots in testing and use of various attribution methods (see Appendix B). In addition to analyses that the pilot sites are conducting, Delmarva is analyzing various methods of attribution, using the Medicare FFS data. We have identified four attribution rules that warrant investigation. They are:

1. One touch rule – Beneficiary is attributed to any provider who has had any Evaluation & Maintenance (E&M) visits with that beneficiary.
2. 30% rule - Beneficiary is attributed to provider who has 30% of a given beneficiary's E&M visits.
3. 50% rule - Beneficiary is attributed to provider who has 50% of a given beneficiary's E&M visits.
4. Maximum frequency (or plurality) rule – Beneficiary is attributed to the physician who has the largest number of that beneficiary's E&M visits.

CCHRI is testing three attribution rules; these rules are listed at the end of this section.

Within the selection of each of these rules there are several other methodological questions. These questions include:

- Should eligible providers be limited according to their specialties? If so, which specialties should be used for which measures? For all of the results reported here, Delmarva limited the list of eligible providers to those with specialties applicable to the measures, according to

the NCQA HEDIS and Scholle et al., 2006 definitions. Table 4 shows the provider specialties that we used.

- Should there be a minimum number of E&M claims for a beneficiary to be attributed to a physician? For all of the results reported here, Delmarva did not use a minimum number of claims, however it has been suggested that we require a physician to have at least 2 claims per beneficiary before making the attribution.
- Should the cost of claims be used instead of the number of claims in making the attribution decision? For all of the results reported here, Delmarva used the number of claims. Using the cost of claims is another option and makes the most sense when looking at measures of clinical efficiency.
- Should the attribution rule be limited to only E&M claims? If so, how should E&M claims be operationally defined. For all of the results reported here, Delmarva used only E&M claims to make the attribution and defined E&M claims as only those made in an office setting with one of the codes listed in Appendix C. There are alternative methods of defining E&M visits.

CCHRI Attribution Rules

- All physicians of any specialty
- Any physician of a relevant specialty (See Table 5)
- Single physician of a relevant specialty (See Table 5)

Table 4. Specialties Included for Cycle 1 Measures

Measure – Cycle 1	Primary Care Provider Specialties
Breast Cancer Screening	*Primary Care Physicians including:* *General Practice, Family Practice, Internal Medicine, Geriatrician, Obstetrics and Gynecology, Oncologist*
LDL testing for diabetics *HbA1c testing for diabetics* *Nephropathy testing for diabetics* *Retinal eye exam for diabetics*	*Primary Care Physicians including:* *General Practice, Family Practice, Internal Medicine, Geriatrician, Endocrinologist*
Cardiovascular – LDL testing	*Primary Care Physicians including:* *General Practice, Family Practice, Internal Medicine, Geriatrician, Cardiologist*
Colorectal cancer screening	*Primary Care Physicians including:* *General Practice, Family Practice, Internal Medicine, Geriatrician, Gastroenterologist*

Note: Primary Care Providers defined using both the Treatment and Referral UPIN for a given Beneficiary's Claim. Primary Care Physician as defined by NCQA, HEDIS 2006. Type Provider or Provider Specialty determined from Physician Primary Specialty Code in the Medicare MPIER file.

Table 5. Relevant Specialties included for Attribution Rules, CCHRI

Measure	Provider Specialties
Breast Cancer Screening	*Primary Care (Family Practice, Internal Medicine), Obstetrics and Gynecology*
HbA1c testing for diabetics	*Primary Care (Family Practice, Internal Medicine), Endocrinology*
LDL Screening for Diabetics	*Primary Care (Family Practice, Internal Medicine), Cardiology*
Cardiovascular – LDL testing	*Primary Care (Family Practice, Internal Medicine, Geriatrician, Cardiology*
Colorectal cancer screening	*Primary Care (Family Practice, Internal Medicine), Gastroenterology*

Attribution Analyses Preliminary Results

Appendix D allows us to compare results of Cycle 1 measures using the 1-touch rule, the 30% rule, and the 50% rule. Table D1 presents preliminary analysis results, including the overall statewide rate for each of the six measures and for five of the six sites (California data have not yet been analyzed). The table, when fully populated will also include the mean individual physician rate, the standard deviation of the physician rate, the percent of eligible physicians with at least one beneficiary attributed to them, and the percent of eligible beneficiaries attributed to at least one physician for each of the three attribution methods used. Measures were calculated by Delmarva using only Medicare FFS data according to the specifications defined by the pilots for their regions.

The mean physician rates for all three attribution methods are higher than the overall rate for all measures in all states reported. This indicates that there is some systematic bias in the selection of beneficiaries who can be attributed to physicians. It makes sense that beneficiaries who have had an E&M visit with a physician from a relevant specialty (and are therefore attributed to at least one physician) are more likely to have the screenings than beneficiaries who have not had an E&M visit with a physician from a relevant specialty (but who are included in the overall rate).

We can also use the mean physician rates to compare the three attribution methods. However the direction of the results varies by measure. For breast cancer, colorectal, and eye exam screening rates, the 1-touch rule produces higher rates than the 30% and 50% rules, which produce results that are very similar. This may be because the 1-touch rule provides the greatest opportunity to count participants multiple times and those beneficiaries who have seen multiple providers are more likely to have received these screenings. For HbA1c, LDL, and nephropathy screening rates however, the 50% rule tends to produce higher rates than the 1-touch and 30% rules. This may be because beneficiaries who see one physician regularly tend to have these screenings more often. Further investigation to explain these observations is warranted.

The percentage of eligible physicians with at least one beneficiary also varied by measure. Eligible physicians include any physician with at least one E&M claim in an office setting during the time period and who has a specialty relevant to the measure. Looking at the 1-touch rule, for breast cancer screening the proportion varied between 39 and 54%, for colorectal screening it varied between 62 and 73%, and for the diabetic measures it was between 35 and 41%. Minnesota consistently had the lowest percent of physicians with at least one beneficiary. It will be interesting to note whether this trend continues as more data are added. It is also clear that the 1-touch rule allows the most physicians to have at least one beneficiary (ranging between 27 and 61% more). Similarly, the 30% rule allowed more physicians to have at least one beneficiary than the 50% rule (ranging between 21 and 31 %). The data on the percent of beneficiaries attributed to a physician is similar in that the 1-touch rule has the highest percentage; however there is not much difference between the 30% rule and the 1-touch rule, while there is a large decrease for the 50% rule.

Physicians with a denominator less than ten (10) tend to have rates that are erratic and unreliable. For this reason, we duplicated table D1 in table D2 after excluding physicians with a denominator less than 10. Ten (10) was selected as an arbitrary minimum denominator size. Twenty (20) and thirty (30) are often used by statisticians as the minimum denominator (N) required to be able to assume a normal distribution. However, the purpose of this table is to show the effect of applying a crude and liberal minimum denominator size.

Only results for Minnesota are available for Table D2 at this time. From the limited data available, we see that many of the trends noted for table D1 remain true. However the mean physician rates tended to decrease after excluding physicians with less than ten beneficiaries. The nephrology rates decreased the most, while the eye exams did not change. The standard deviations decreased consistently after excluding physicians with less than ten beneficiaries. As would be expected the percent of physicians with at least ten beneficiaries was much smaller than the percent of physicians with at least one beneficiary (ranging between 11 and 55% fewer physicians). This indicates that even with a very crude and liberal cutoff for a minimum denominator size, many physicians will not have a reportable rate. Likewise, fewer beneficiaries are represented in the rates from physicians when the less than 10 cutoff criterion is added.

Reliability and Minimum Denominator Size

Tables D1 and D2 showed that the attribution methods produce differing results. We added a minimum denominator size of ten for table D2, because we know that rates with a small denominator are unstable. However using an arbitrary cutoff point such as ten is crude and is probably overly liberal. Using the between physician variation to calculate reliability scores is an evidence-based approach to identifying appropriate minimum denominator sizes.

Cronbach's alpha correlation is a standard method for calculating reliability. However, this method requires that for each calculation, we limit the data for each physician to a specified number of participants. Even with a minimum denominator size, we see in the real world that there is much variation between physicians denominator size. Therefore using the Cronbach's alpha requires that we exclude some of our data for the calculation and this does not mimic the real world. The NCQA introduced an innovative approach using a beta binomial model methodology to calculate reliability. Reliability estimates require an estimate of physician-to-physician variation. To estimate the physician-to-physician variance, one would fit the beta-binomial models to each measure using a SAS macro developed for this purpose. Delmarva is currently working with NCQA to assess whether NCQA's beta binomial model approach to reliability would be a more statistically robust and valid approach to assessing reliability of attributed measures for providers, as opposed to using the Cronbach alpha correlation.

Table D3 shows reliability results for selected measures and states using the 30% rule. This is an attempt to understand the differences between the methods for calculating reliability and to develop a process for

identifying a minimum denominator size. This will then be added to our statistics used to compare the many attribution methods. We started the analysis by identifying the deciles of denominator size for each measure and state. We then calculated the mean physician rate, standard deviation, percent of physicians meeting that minimum denominator size, the Cronbach's alpha reliability, the beta binomial reliability, and the beta binomial reliability at the minimum sample size. For each decile with a minimum denominator size greater than ten, we listed out these descriptive statistics. Most of the table is not completed because we have just recently started these analyses.

A reliability score of 0.70 or higher is generally recognized as "good". It appears that the beta binomial reliability scores are higher than both the Cronbach's alpha and beta binomial at the minimum denominator, except for eye exam rates where the beta binomial scores are very low. This is probably due to the fact that eye exam rates are very high and the distribution of scores is very compact. For LDL and nephrology scores from Minnesota, the beta binomial model failed to converge and we have been unable to calculate a score. We anticipate using the beta binomial approach to calculate reliability, but further investigation is warranted to understand its limitations.

It is too early to identify minimum denominator sizes, even using the reliability scores for the 30% rule in Table D3. Using 0.70 as the reliability score cutoff point, the lowest denominator size revealed was 12, the highest was 24. However, for some of the diabetic measures even the 90th percentile of denominator sizes was too small to produce reliable scores. This indicates that fewer than 10% of physicians with at least one beneficiary will have a reportable score for some measures.

Conclusion

Cycle 1 data aggregation was a learning experience for all six BQI pilots and for Delmarva. As discussed above, numerous challenges were faced and, for the most part, addressed or overcome. At the time of this report, the BQI pilot sites were well into Cycle 2 data aggregation, and were applying the lessons learned during Cycle 1 to their subsequent data aggregation and analysis processes.

Appendix A – BQI Project Measures

1 = Measure calculated during Cycle 1 aggregation
2 = Measure calculated during Cycle 2 aggregation

Measure	Specifications	NQF endorsed	AQA Starter Set	AZ	CA	IN	MA	MN	WI[3]
PREVENTION									
Breast Cancer Screening	HEDIS – BCS	Y	Y	1	1	1	1	1	1
Colorectal cancer screening	HEDIS – COL	Y	Y	1	1	1	1	2[4]	1
Pneumonia vaccination in 65+ yrs pop	CMS / NCQA	Y	N						2[5]
DIABETES									
LDL testing	HEDIS – CDC[6]	Y	Y	1	1	1	1	1	1
Diabetes LDL cholesterol control	HEDIS– CDC	Y	Y			2		2	2
HbA1c testing	HEDIS – CDC	Y	Y	1	1	1	1	1	1
HbA1c control	HEDIS - CDC	Y	Y			2		2	1
Nephropathy testing	HEDIS – CDC	Y	N	2	2	1	2	1	2
Retinal eye exam	HEDIS– CDC	Y	Y	1		2	1	1	
Blood pressure control	HEDIS– CDC	Y	N					2	2
Optimal diabetes composite	MNCM	N	N					2[7]	
Diabetes – smoking status	MNCM	N	N					2	
Diabetes – daily aspirin use	MNCM	N	N					2	
CARDIOVASCULAR CONDITIONS									
Cardiovascular – LDL testing	HEDIS–CMC[8]	Y	N	2	1	2	2		2
Cardiovascular – LDL control	HEDIS–CMC	Y	N			2			2
CAD patients receiving lipid-lowering therapy	PCPI	Y	Y	2	2		2		
HEART ATTACK / HEART FAILURE									
Beta blocker treatment after heart attack[9]	HEDIS–BBH	Y	Y			2	2		
Persistence of beta blocker therapy – Post MI	HEDIS–PBH	Y	Y	2	2	2	2		
HF: LVF testing	PCPI	Y	Y		2				
Heart failure: ACE inhibitor/ARB therapy	PCPI	Y	Y	2					
Warfarin for patients w/ CHF & atrial fibrillation	PCPI	Y	N	2	2				
MEDICATION MANAGEMENT									
Antidepression Medication Management – Acute Phase	HEDIS-AMM	Y	Y	2					

[3] WCHQ will use HEDIS specifications for Medicare data and WCHQ specifications for its own data.
[4] MN will collect this measure via Hybrid data collection method.
[5] Administrative and EMR data only, no survey.
[6] CDC = Comprehensive Diabetes Care. This measure is one of those included in this category.
[7] MNCM Optimal Diabetes Composite: HbA1C control <7.0, LDL – control <100, BP control (collect <130/80 and <140/90), Daily Aspirin Use, Documented Tobacco free
[8] CMC = Cholesterol Management for Patients with Cardiovascular Conditions. This measure is one of those included in this category.
[9] 2008 HEDIS specifications for public comment indicate potential retirement of this measure (http://www.ncqa.org/Programs/PublicComment/HEDIS2008/index.htm)

Measure	Specifications	NQF endorsed	AQA Starter Set	AZ	CA	IN	MA	MN	WI[3]
Annual monitoring for patients on persistent medications: Ace inhibitors, Digoxin, Diuretics, Anti-convulsants, Statins	HEDIS–MPM	Y	N		2[10]		2[11]		
OTHER									
Hypertension – blood pressure control	HEDIS-CBP	Y	N						2
Use of Spirometry Testing in Assessment/Diagnosis of COPD	HEDIS–SPR	N	N				2		
Osteoporosis Management for Women w/ Fracture	PCPI	N	N		2				

[10] Annual monitoring for patients on persistent medications: CCHRI:5- Components: Diuretics, Ace inhibitors, Digoxin, Anti-Convulsants, Statins Cycle 1 based on 2005 data, HEDIS 2006; Cycle 2 based on 2006 data, HEDIS 2007

[11] Annual monitoring for patients on persistent medications: MHQP: ARBs & Ace inhibitors

Appendix B – Cycle 1 Measures, by Pilot

	Breast Cancer Screening	HbA1c Screening Diabetics	LDL Screening Diabetics	Nephropathy Testing Diabetics	Retinal Eye Exam Diabetics	Colorectal Cancer Screening	LDL Testing CVD	HbA1c Control Diabetics
Center for Health Information & Research (CHIR)								
Age Range	52-69	>18	18-75	X	>18	51-80	X	X
Measurement Period	2 years	1 year (2 tests)	1 year (1 test)	X	1 year (1 test)	10 yrs if poss	X	X
Denominator Period	2 years	27 months	27 months	X	27 months	10 yrs if poss	X	X
Specifications	Symmetry EBM Connect (http://www.symmetry-health.com/products/product_SEBM.php)							
Population Covered[12]	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care, Medicaid FFS; Medicaid Managed Care							
Data Sources	Inpatient, Outpatient, Physician Office claims, some lab data, some point of care data							
Specialties / Medical Professionals	Attribution to Primary Care Physicians (Doctors of Medicine, Doctors of Osteopathy) in Internal Medicine, Family Practice, Pediatrics, Geriatrics, and OB/GYN							
Patient to Physician/Group Attribution	EBM Connect – PCP Imputation Process 1. Phys exam or assessment performed by physician w/ allowed specialty (limited to selected specialties) and who is the PCP assigned via enrollment process 2. Most recent phys exam or assessment performed by physician other than assigned PCP, limited to allowed specialities 3. Eval and mgmt type visit performed by phys w/ allowed specialty and who is assigned PCP via enrollment 4. Phys who is allowed specialty who is other than the assigned PCP and who performed largest number of evaluation and management type visits 5. Prenatal, postpartum, or antepartum visit, or routine obstetrical service performed by physician w/ allowed specialty and who is assigned PCP via enrollment 6. Physican w/ allowed specialty who is other than the assigned PCP and who performed the largest number of prenatal, postpartum, or routine obstetrical care services 7. PCP assigned via enrollment. For gyn PCP imputation, this PCP must have Ob/Gyn specialty. For medical primary care giver imputation, the PCP at this point can be any specialty. 8. No PCP can be assigned.							

[12] Populations Covered include Product Line, Product and/or Payment Arrangement

	Breast Cancer Screening	HbA1c Screening Diabetics	LDL Screening Diabetics	Nephropathy Testing Diabetics	Retinal Eye Exam Diabetics	Colorectal Cancer Screening	LDL Testing CVD	HbA1c Control Diabetics
California Cooperative Healthcare Reporting Initiative (CCHRI)								
Age Range	52-69	18-75	18-75	X	X	51-80	18-75	X
Measurement Period	2 years	1 year	1 year	X	X	2 yrs	1 year	X
Denominator Period	2 years	2 years	2 years	X	X	2 yrs	2 years	X
Specifications	Administrative. HEDIS 2006, HEDIS 2007 (LDL testing for CVD only). For Breast Cancer and Colorectal Cancer Screening, HCPCS codes from HEDIS 2007 used for numerator determination.							
Population Covered	Commercial PPO (3 plans); Medicare FFS							
Data Sources	Administrative only: Commercial and Medicare professional service, institutional, and pharmacy claims and encounters.							
Specialties / Medical Professionals	Attribution to MDs (PCPs, and excluding such MDs such as radiology, etc), DOs, Nurse Practitioners and Physician Assistants. Specialties as per measure specifications, including primary care (family practice, internal medicine), OB/GYN, endocrinology, cardiology, gastroenterology.							
Patient to Physician/Group Attribution	Method #1: Single physician of relevant specialty for a given measure with the majority of E&M visits during the measurement period.(See Table 1 for relevant specialties) Method #2: Any physician of relevant specialty for a given measure with at least one E&M visit during the measurement period. Method #3: All providers of any specialty (specialties not constrained to Table 1). For each patient identified in the denominator, identify any physician who had a minimum of one E&M visit with the patient during the measurement period. In contrast to Methods 1 and 2, the physician does not have to be of relevant specialty for the measure.							
Indiana Health Information Exchange (IHIE)								
Age Range	52-69	18-75	18-75	18-75	X	50-80	X	X
Measurement Period	2 years	1 year	1 year	1 year	X	120 months	X	X
Denominator Period	2 years	2 years	2 years	2 years	X	As far back as possible	X	X
Specifications	HEDIS 2006, modified for LDL Screening for Diabetics.							
Population Covered	Commercial Managed Care and FFS; Medicare FFS; Medicare Managed Care; some Medicaid Managed Care and FFS.							
Data Sources	Administrative including claims, some point-of-care data							

	Breast Cancer Screening	HbA1c Screening Diabetics	LDL Screening Diabetics	Nephropathy Testing Diabetics	Retinal Eye Exam Diabetics	Colorectal Cancer Screening	LDL Testing CVD	HbA1c Control Diabetics
Specialties / Medical Professionals	Attribution to MDs, DOs, PAs, and NPs.							
Patient to Physician/Group Attribution	Three categories of patients: 1. Pat's who have not had any interactions with providers 2. Pt's who have had interactions with only one provider that meets criteria for PCP designation – (1 & 2 identifies 80% Patients) 3. Pt's who have had interactions with multiple providers that meet criteria to be PCP – Identifies PCP 75% of patients Also identified "providers" that are groups, practices, or not single physician who can act as PCP so that data attributed to those providers is ignored by the PCP attribution algorithm.							
Massachusetts Health Quality Partners (MHQP)								
Age Range	52-69	18-65	18-75	X	18-75	50-80	X	X
Measurement Period	2 years	1 year	1 year	X	1 year	1 year	X	X
Denominator Period	2 years	2 years	2 years	X	2 years	As far back as possible	X	X
Specifications	HEDIS 2006, with claims adjustment factor if necessary.							
Population Covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care							
Data Sources	Commercial administrative claims, Medicare managed care measure results at numerator/denominator level							
Specialties / Medical Professionals	Attribution to MDs and DOs (PCPs), except where the patient did not see any primary care physician for a preventive or routine visit in the measurement period. Endocrinologists are included as PCPs where they are the only physician that has been seen for routine care. Other specialists to be included as PCPs if a plan had so designated them.							
Patient to Physician/Group Attribution	No attribution method for private data – Will use assigned PCP according to enrollment data							
Minnesota Community Measurement (MNCM)								
Age Range	52-69	18-75	18-75	18-75	18-75	X	X	X
Measurement Period	2 years	1 year	1 year	1 year	1 year	X	X	X
Denominator Period	2 years	2 years	2 years	2 years	2 years	X	X	X

	Breast Cancer Screening	HbA1c Screening Diabetics	LDL Screening Diabetics	Nephropathy Testing Diabetics	Retinal Eye Exam Diabetics	Colorectal Cancer Screening	LDL Testing CVD	HbA1c Control Diabetics
Specifications	HEDIS 2006. Hybrid data collection for diabetes measures.							
Population Covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care; Medicaid FFS							
Data Sources	Aggregated claims data by medical group for administrative measures, patient level medical record data attributed to each medical group for hybrid measures, FFS claims.							
Specialties / Medical Professionals	All as appropriate.							
Patient to Physician/Group Attribution	Maximum claims to Tax ID Numbers (TIN) w/ highest number of claims during measurement year – selected specialties							
Wisconsin Collaborative for Healthcare Quality (WCHQ)								
Age Range	40-68	18-85	18-85	X	X	50-79	X	18-85
Measurement Period	2 years	1 year	1 year	X	X	1 year	X	1 year
Denominator Period	3 years	2 years	2 years	X	X	As far as possible. 1 visit w/in 2 years or 2 visits w/in 3 years	X	2 years
Specifications	Medical groups/providers determine denominators as per WCHQ methodology, and then provide numerator information.							
Population Covered	All payers; Medicare FFS.							
Data Sources	Medicare FFS claims; membership organizations measure results (numerators and denominators)							
Specialties / Medical Professionals	PAs, NPs, DOs, MDs.							
Patient to Physician/Group Attribution	Three questions are used to determine the denominator a. Is this a patient with the diagnosis? b. Is this a patient our physician group manages? c. Is this a patient that is current in our system?							

Appendix C – List of Evaluation and Management codes used to identify physicians eligible to be attributed to beneficiaries

Category / Subcategory	Codes	Category / Subcategory	Codes
Office or Other Outpatient Services	99201-99205	Subsequent Nursing Facility Care	99311-99313
New Patient	99211-99215	Nursing Facility Discharge Services	99315-99316
Established Patient	99217	Domiciliary, Rest Home or	
Hospital Observation Discharge Services	99218-99220	Custodial Care Services	
Hospital Observation Services		New Patient	99321-99323
Hospital Observation or Inpatient Care		Established Patient	99331-99333
Services (Including Admission and Discharge	99234-99236	Home Services	
Services)		New Patient	99341-99345
Hospital Inpatient Services	99221-99223	Established Patient	99347-99350
Initial Hospital Care	99231-99233	Prolonged Services	
Subsequent Hospital Care	99238-99239	With Direct Patient Contact	99354-99357
Hospital Discharge Services		Without Direct Patient Contact	99358-99359
Consultations	99241-99245	Standby Services	99360
Office Consultations	99251-99255	Case Management Services	
Initial Inpatient Consultations	99261-99263	Team Conferences	99361-99362
Follow-up Inpatient Consultations	99271-99275--	Telephone Calls	99371-99373
Confirmatory Consultations	99281-99288	Care Plan Oversight Services	99374-99380 —
Emergency Department Services	99289-99290	Preventive Medicine Services	
Pediatric Patient Transport		New Patient	99381 -99387 --
Critical Care Services	99291-99292	Established Patient	99391-99397-
Adult (over 24 months of age)	99293-99294	Individual Counseling	99401-99404
Pediatric	99295-99296	Group Counseling	99411-99412
Neonatal	99298-99299	Other	99420-99429
Intensive Care (Low Birth Weight)		Newborn Care	99431-99440
Nursing Facility Services		Special E/M Services	99450-99456
Comprehensive Nursing Facility Assessments	99301-99303	Other E/M Services	99499 —

Appendix D – Preliminary Analysis Results, Cycle 1

Table D1: Comparison of 3 attribution methods, cycle 1 data CY2005 measure rates by state

Measure	State	Overall Rate*	1-touch Rule				30% Rule				50% Rule			
			Mean physician rate[13]	SD	% physicians w/ at least 1 bene[14]	% benes w/ at least 1 physician[15]	Mean physician rate[1]	SD	% physicians w/ at least 1 bene[2]	% benes w/ at least 1 physician[3]	Mean physician rate[1]	SD	% physicians w/ at least 1 bene[2]	% benes w/ at least 1 physician[3]
BCS	AZ	0.576	0.726		46.7	85.2	0.68		33.4	83.6	0.678		26.1	67.1
	CA													
	IN	0.564	0.693		54	89.7	0.643		39.2	88.3	0.641		31.5	73.2
	MA	0.649	0.796		50.6	79.7	0.758		38.6	78.9	0.758		31.8	67
	MN	0.649	0.765	0.180	38.6	80.5	0.757	0.208	30.3	78.7	0.767	0.215	24.1	62.1
	WI	0.593												
COL	AZ	0.311	0.51	0.245	62.3	86.3	0.392	0.234	40.3	84.4	0.367	0.224	32	67.8
	CA													
	IN	0.247	0.427		73.4	90.8	0.308		45.7	89.4	0.293		36.8	74
	MA	0.371												
	MN													
	WI	0.297												
HbA1c	AZ	0.658	0.813		40.5	85.9	0.799		26.1	83.6	0.824		19.9	65.2
	CA													
	IN	0.716												

[13] Proportions calculated at the individual physician level, then average of the proportions reported

[14] Eligible physicians include any one with at least 1 Evaluation and Management (E&M) claim in an office setting during the time period of the rate and have speciality relevant to the rate. E&M claims used are identified in Appendix C. Relevant specialities vary according to measure: Breast cancer screening: General Practice, Family Practice, Internal Medicine, Geriatrician, Obstetrics and Gynecology, Oncologist; Colorectal screening: General Practice, Family Practice, Internal Medicine, Geriatrician, Gastroenterologist; Diabetes measures: General Practice, Family Practice, Internal Medicine, Geriatrician, Endocrinologist.

[15] Eligible beneficiaries includes all beneficiaries included in the denominator for the measure.

Measure	State	Overall Rate*	1-touch Rule				30% Rule				50% Rule			
			Mean physician rate[13]	SD	% physicians w/ at least 1 bene[14]	% benes w/ at least 1 physician[15]	Mean physician rate[1]	SD	% physicians w/ at least 1 bene[2]	% benes w/ at least 1 physician[3]	Mean physician rate[1]	SD	% physicians w/ at least 1 bene[2]	% benes w/ at least 1 physician[3]
	MA	0.788												
	MN	0.793	0.837	0.193	35	86.2	0.855	0.206	24.1	83.8	0.877	0.195	18.6	66.6
	WI	0.835												
Eye Exam	AZ	0.959	0.993		40.5	85.9	0.987		26.1	83.6	0.987		19.9	65.2
	CA													
	IN													
	MA	0.972												
	MN	0.962	0.992	0.040	35	86.2	0.985	0.066	24.1	83.8	0.986	0.065	18.6	66.6
	WI													
LDL Testing	AZ	0.665	0.8		40.5	85.9	0.794		26.1	83.6	0.81		19.9	65.2
	CA													
	IN	0.788												
	MA	0.774												
	MN	0.721	0.773	0.211	35	86.2	0.8	0.226	24.1	83.8	0.825	0.219	18.6	66.6
	WI	0.734												
Nephropathy Testing	AZ													
	CA													
	IN	0.539												
	MA													
	MN	0.617	0.716	0.216	35	86.2	0.722	0.252	24.1	83.8	0.729	0.257	18.6	66.6
	WI													

Table D2: Comparison of 3 attribution methods, cycle 1 data CY2005 measure rates by state, excluding physicians with denom < 10

Measure	State	Overall Rate*	1-touch Rule				30% Rule				50% Rule			
			Mean physician rate[16]	SD	% physicians w/ at least 1 bene[17]	% benes w/ at least 1 physician[18]	Mean physician rate[1]	SD	% physicians w/ at least 1 bene[2]	% benes w/ at least 1 physician[3]	Mean physician rate[1]	SD	% physicians w/ at least 1 bene[2]	% benes w/ at least 1 physician[3]
BCS	AZ	0.576												
	CA													
	IN	0.564												
	MA	0.649												
	MN	0.649	0.728	0.139	21.3	78.1	0.713	0.161	11.6	66.1	0.72	0.163	7.8	44.5
	WI	0.593												
COL	AZ	0.311												
	CA													
	IN	0.247												
	MA	0.371												
	MN	NA												
	WI	0.297												
HbA1c	AZ	0.658												
	CA													
	IN	0.716												
	MA	0.788												

[16] Proportions calculated at the individual physician level, then average of the proportions reported

[17] Eligible physicians include any one with at least 1 Evaluation and Management (E&M) claim in an office setting during the time period of the rate and have speciality relevant to the rate. E&M claims used are identified in Appendix C. Relevant specialities vary according to measure: Breast cancer screening: General Practice, Family Practice, Internal Medicine, Geriatrician, Obstetrics and Gynecology, Oncologist; Colorectal screening: General Practice, Family Practice, Internal Medicine, Geriatrician, Gastroenterologist; Diabetes measures: General Practice, Family Practice, Internal Medicine, Geriatrician, Endocrinologist.

[18] Eligible beneficiaries includes all beneficiaries included in the denominator for the measure.

Measure	State	Overall Rate*	1-touch Rule				30% Rule				50% Rule			
			Mean physician rate[16]	SD	% physicians w/ at least 1 bene[17]	% benes w/ at least 1 physician[18]	Mean physician rate[1]	SD	% physicians w/ at least 1 bene[2]	% benes w/ at least 1 physician[3]	Mean physician rate[1]	SD	% physicians w/ at least 1 bene[2]	% benes w/ at least 1 physician[3]
	MN	0.793	0.819	0.164	13	76.9	0.833	0.183	3.7	40.9	0.85	0.168	2.1	23.1
	WI	0.835												
Eye Exam	AZ	0.959												
	CA													
	IN	NA												
	MA	0.972												
	MN	0.962	0.992	0.024	13	76.9	0.986	0.034	3.7	40.9	0.984	0.038	2.1	23.1
	WI													
LDL Testing	AZ	0.665												
	CA													
	IN	0.788												
	MA	0.774												
	MN	0.721	0.746	0.166	13	76.9	0.769	0.185	3.7	40.9	0.796	0.181	2.1	23.1
	WI	0.734												
Nephropathy Testing	AZ	NA												
	CA	NA												
	IN	0.539												
	MA	NA												
	MN	0.617	0.662	0.166	13	76.9	0.635	0.198	3.7	40.9	0.624	0.214	2.1	23.1
	WI	NA												

Table D3: Reliability scores of physician specific rates calculated using 30% rule at differing minimum denominator sizes, by measure and state.

Measure	State	Percentile[19]	Minimum denom	Mean physician rate[20]	SD	% physicians w/ at least 1 bene[21]	% benes w/ at least 1 physician[22]	Cronbach's Alpha Reliability	Beta-Binomial Reliability (@ min denom)
BCS	AZ	90th	36					0.73	0.79(0.72)
		80th	25					0.67	0.75(0.64)
		70th	18					0.63	0.74(0.59)
		60th	12					0.55	0.70(0.49)
	CA	90th							
		80th							
		70th							
		60th							
	IN	90th	42					0.76	0.79(0.74)
		80th	32					0.73	0.77(0.69)
		70th	24					0.68	0.74(0.63)
		60th	18					0.61	0.72(0.56)
		50th	12					0.52	0.69(0.47)
	MA	90th							
		80th							
		70th							
		60th							
	MN	90th	24	0.720	0.151	3.2	31.7	0.68	0.72(0.66)
		80th	17	0.723	0.149	6.3	48.9	0.62	0.66(0.55)

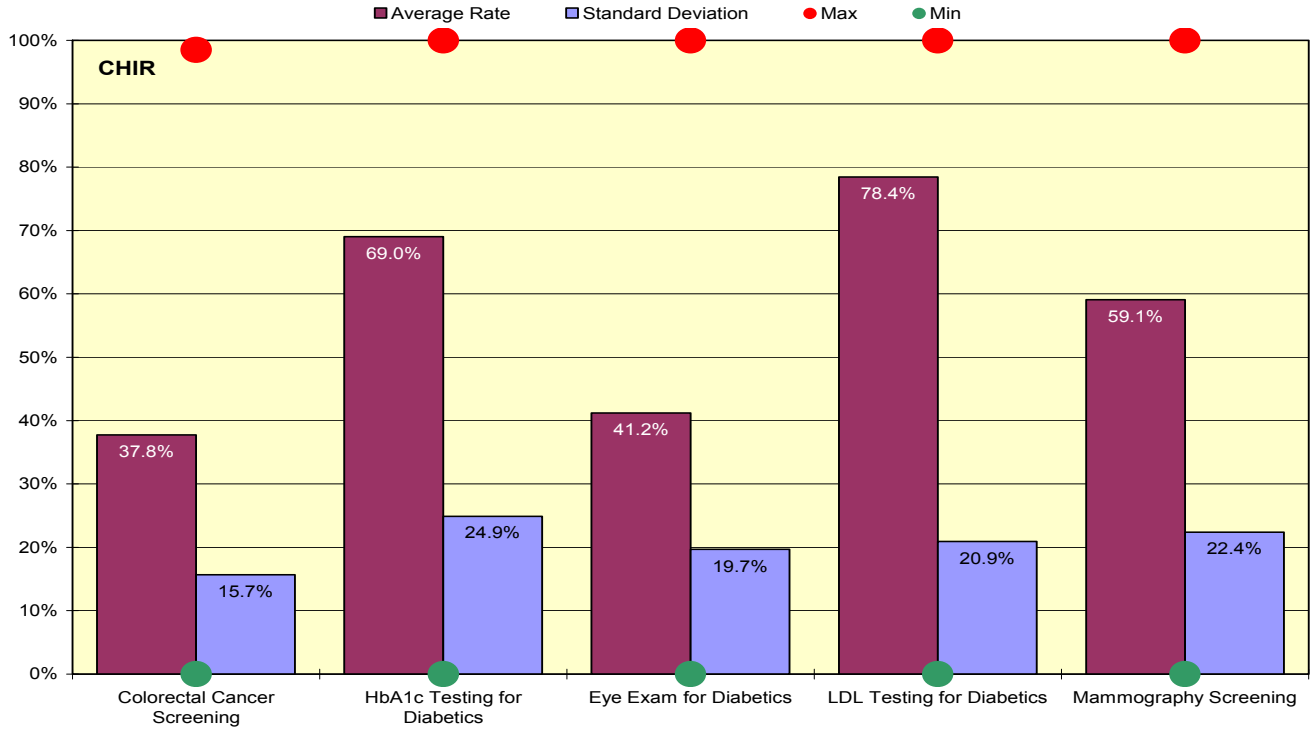
[19] Deciles are reported for 90% and decreasing until the minimum denominator size is < 10.
[20] Proportions calculated at the individual physician level, then average of the proportions reported.
[21] Eligible physicians include any one with at least 1 Evaluation and Management (E&M) claim in an office setting during the time period of the rate and have specialty relevant to the rate. E&M claims used are identified in Appendix C. Relevant specialties vary according to measure: Breast cancer screening: General Practice, Family Practice, Internal Medicine, Geriatrician, Obstetrics and Gynecology, Oncologist; Colorectal screening: General Practice, Family Practice, Internal Medicine, Geriatrician, Gastroenterologist; Diabetes measures: General Practice, Family Practice, Internal Medicine, Geriatrician, Endocrinologist.
[22] Eligible beneficiaries includes all beneficiaries included in the denominator for the measure.

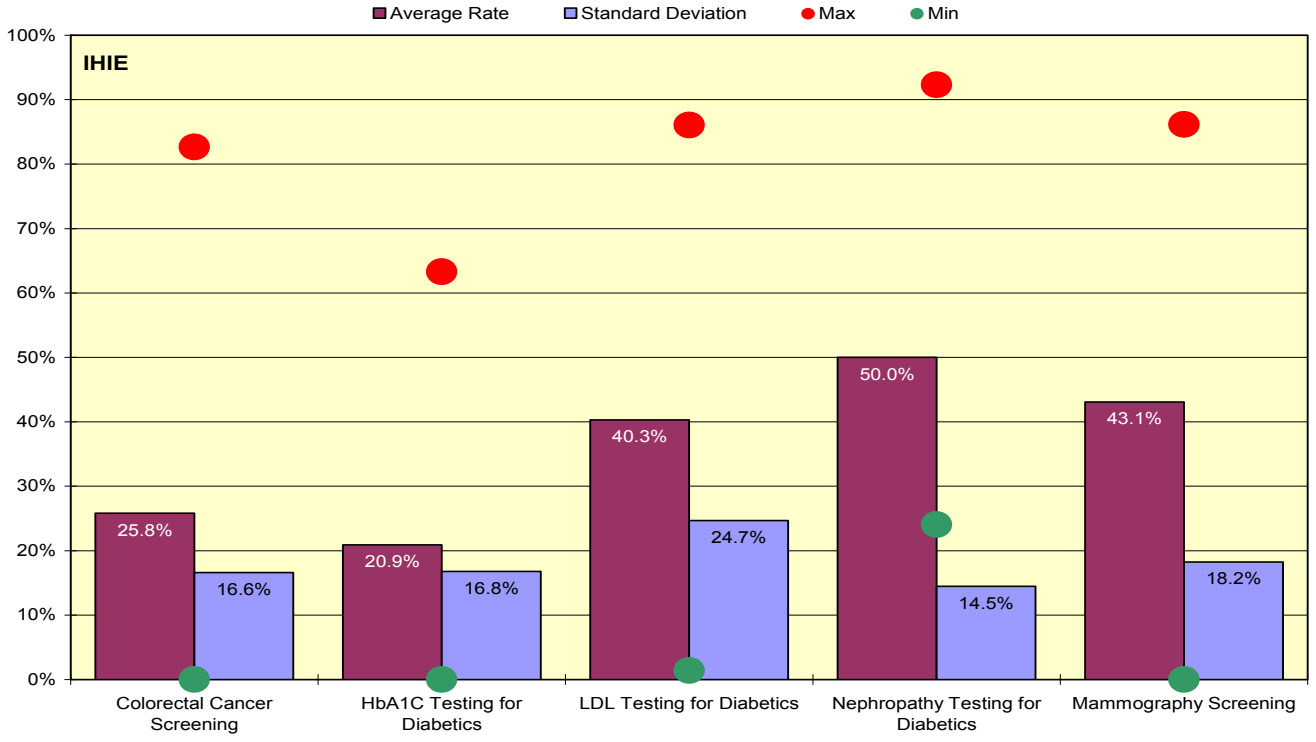
Measure	State	Percentile[19]	Minimum denom	Mean physician rate[20]	SD	% physicians w/ at least 1 bene[21]	% benes w/ at least 1 physician[22]	Cronbach's Alpha Reliability	Beta-Binomial Reliability (@ min denom)
		70th	13	0.729	0.150	8.9	58.9	0.53	0.62(0.49)
	WI	90th							
		80th							
		70th							
		60th							
COL	AZ	90th	173	0.369	0.141			0.94	0.96(0.93)
		80th	103	0.358	0.139			0.9	0.94(0.89)
		70th	60					0.83	0.93(0.82)
		60th	35					0.74	0.91(0.73)
		50th	19					0.64	0.90(0.60)
	CA	90th							
		80th							
		70th							
		60th							
	IN	90th	224					0.88	0.90(0.87)
		80th	32					0.73	0.77(0.69)
		70th	24					0.68	0.74(0.63)
		60th	18					0.61	0.72(0.56)
	MA	90th							
		80th							
		70th							
		60th							
	MN	90th							
		80th							
		70th							
		60th							

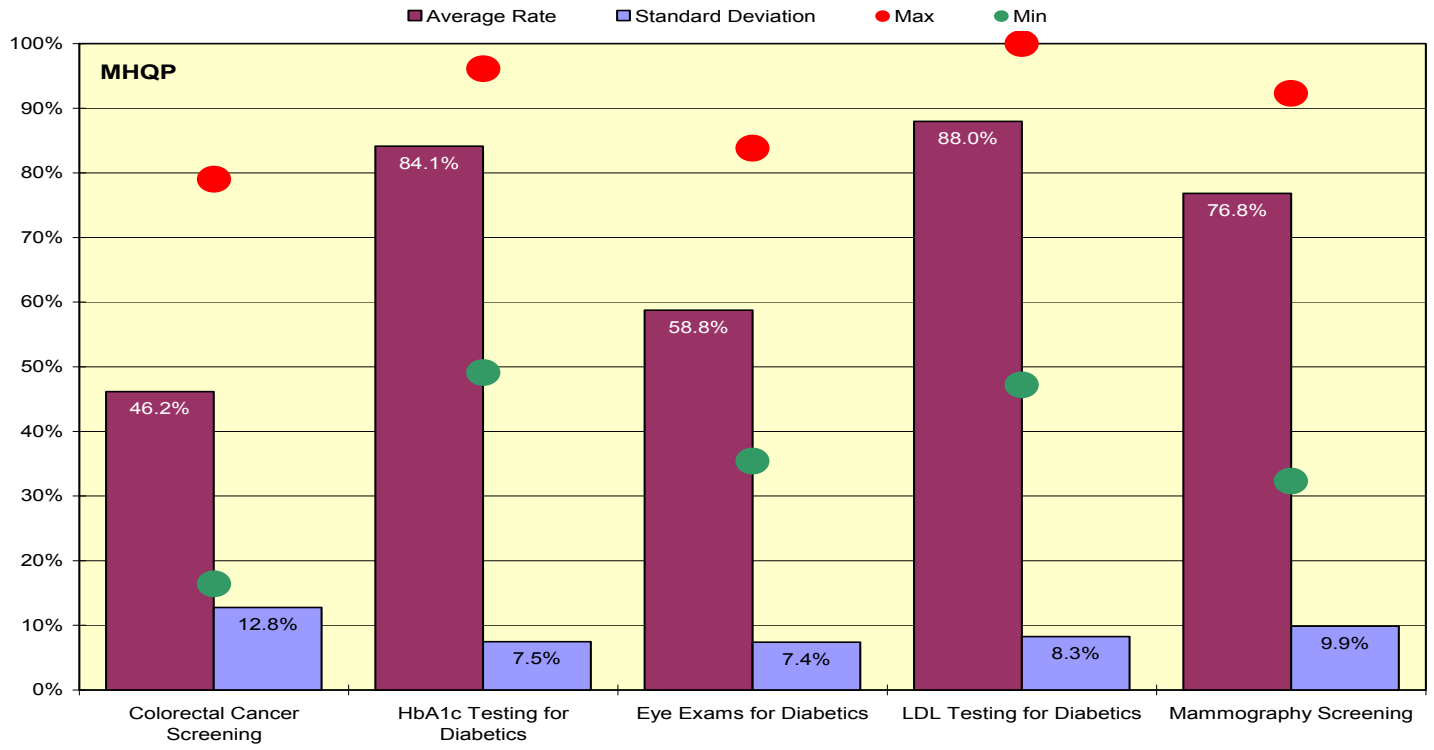
Measure	State	Percentile[19]	Minimum denom	Mean physician rate[20]	SD	% physicians w/ at least 1 bene[21]	% benes w/ at least 1 physician[22]	Cronbach's Alpha Reliability	Beta-Binomial Reliability (@ min denom)
	WI	90th							
		80th							
		70th							
		60th							
HbA1c	AZ	90th	15					0.65	0.74(0.65)
	CA								
	IN								
	MA								
	MN	90th	12	0.842	0.165	2.4	30.9	0.64	0.65(0.58)
	WI								
Eye Exam	AZ	90th	15					0.6	0.42(0.32)
	CA								
	IN								
	MA								
	MN	90th	12	0.986	0.031	2.4	30.9	0.53	0.13(0.10)
	WI								
LDL Test	AZ	90th	15					0.63	0.72(0.63)
	CA								
	IN								
	MA								
	MN	90th	12	0.774	0.172	2.4	30.9	0.6	.
	WI								
Nephropathy	AZ								
	CA								
	IN								
	MA								

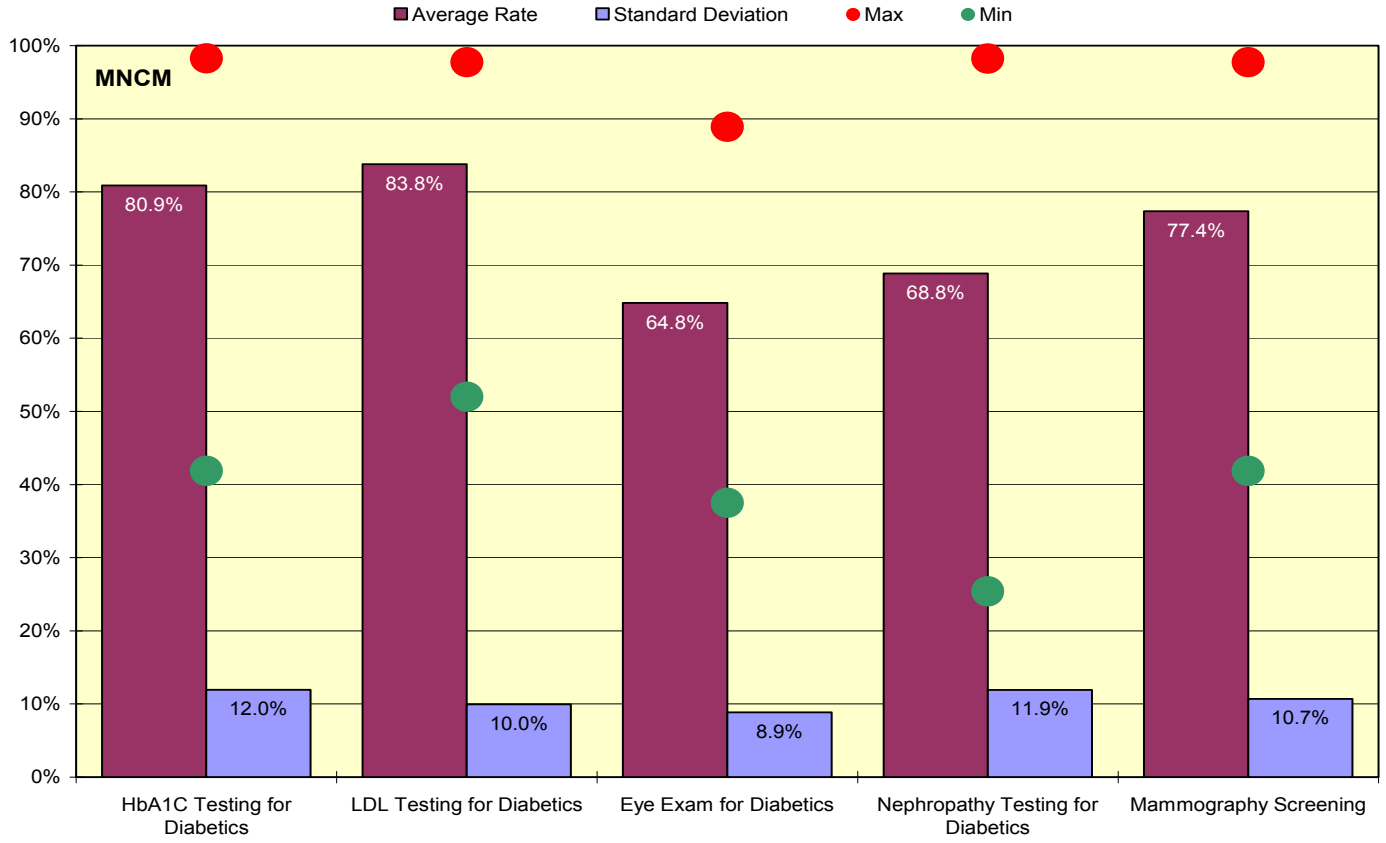
Measure	State	Percentile[19]	Minimum denom	Mean physician rate[20]	SD	% physicians w/ at least 1 bene[21]	% benes w/ at least 1 physician[22]	Cronbach's Alpha Reliability	Beta-Binomial Reliability (@ min denom)
	MN	90th	12	0.640	0.197	2.4	30.9	0.56	.
	WI								

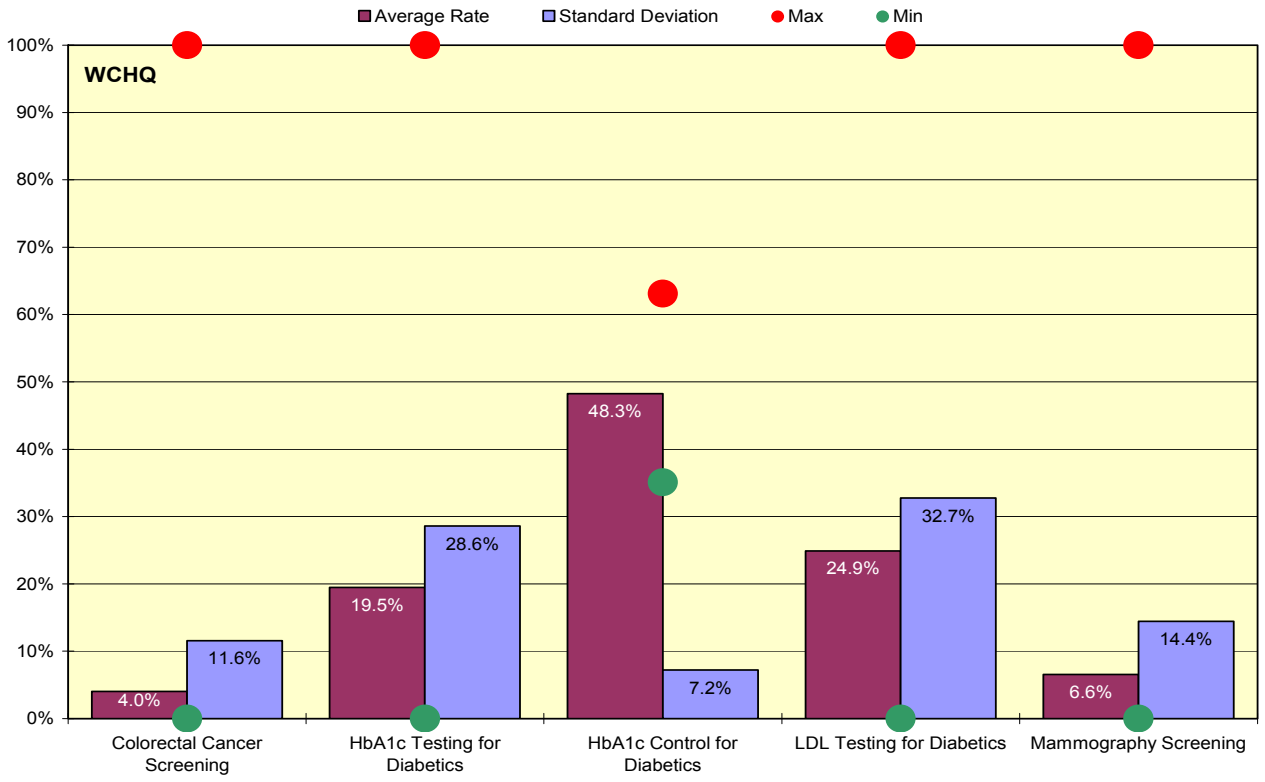
Appendix E – Summary Statistics Figures

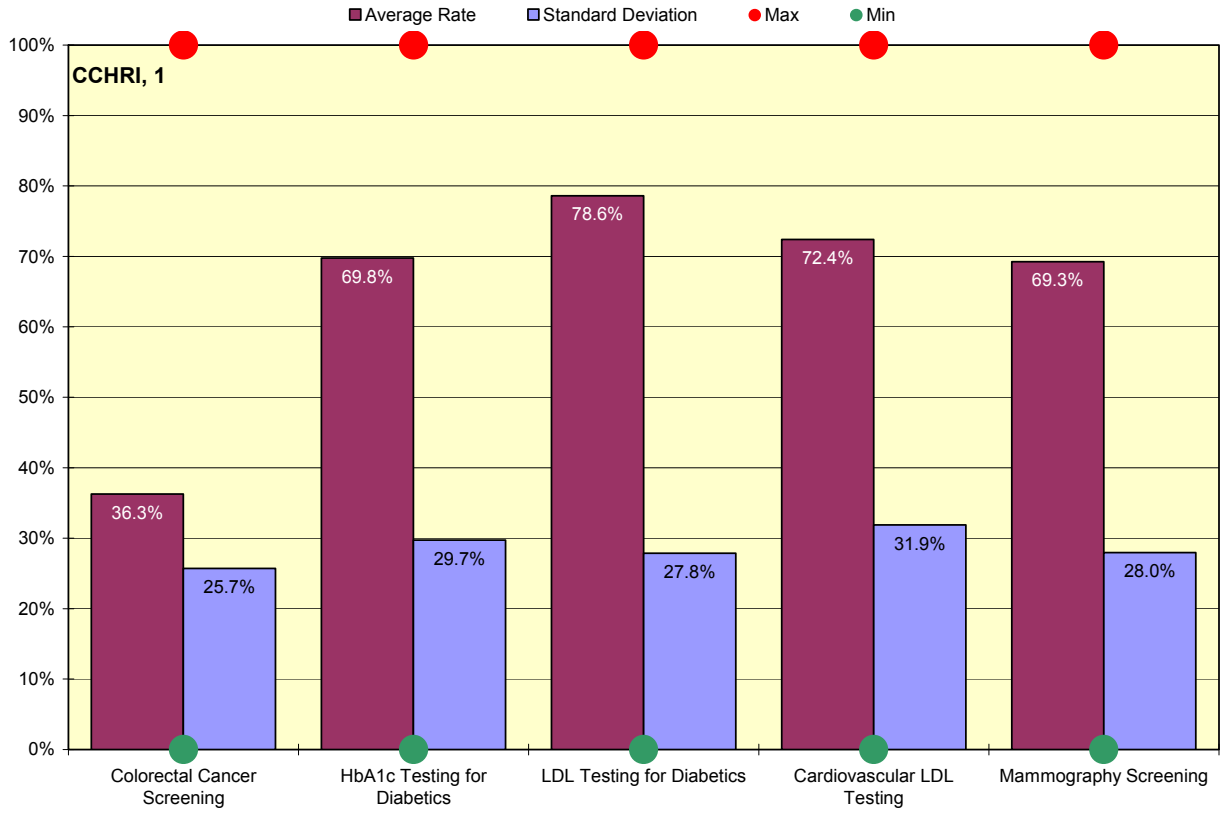


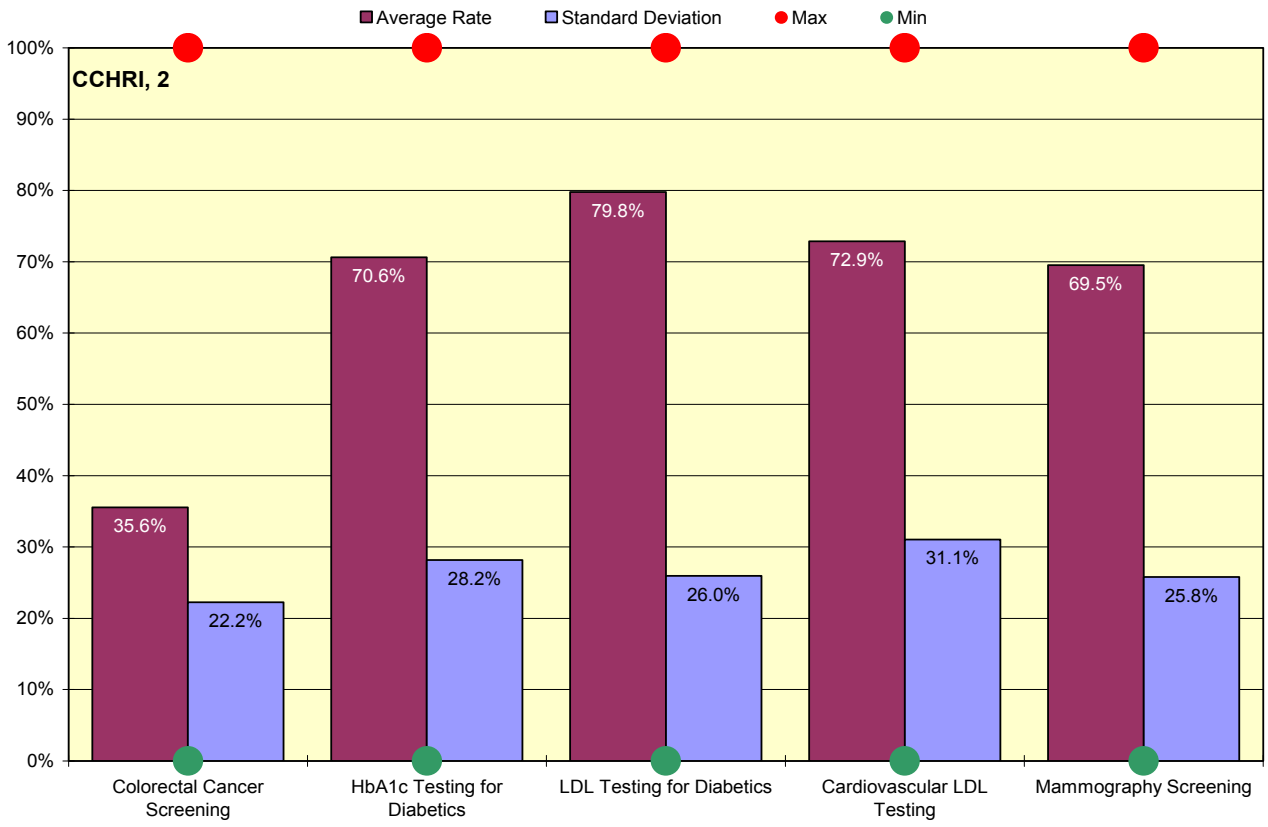


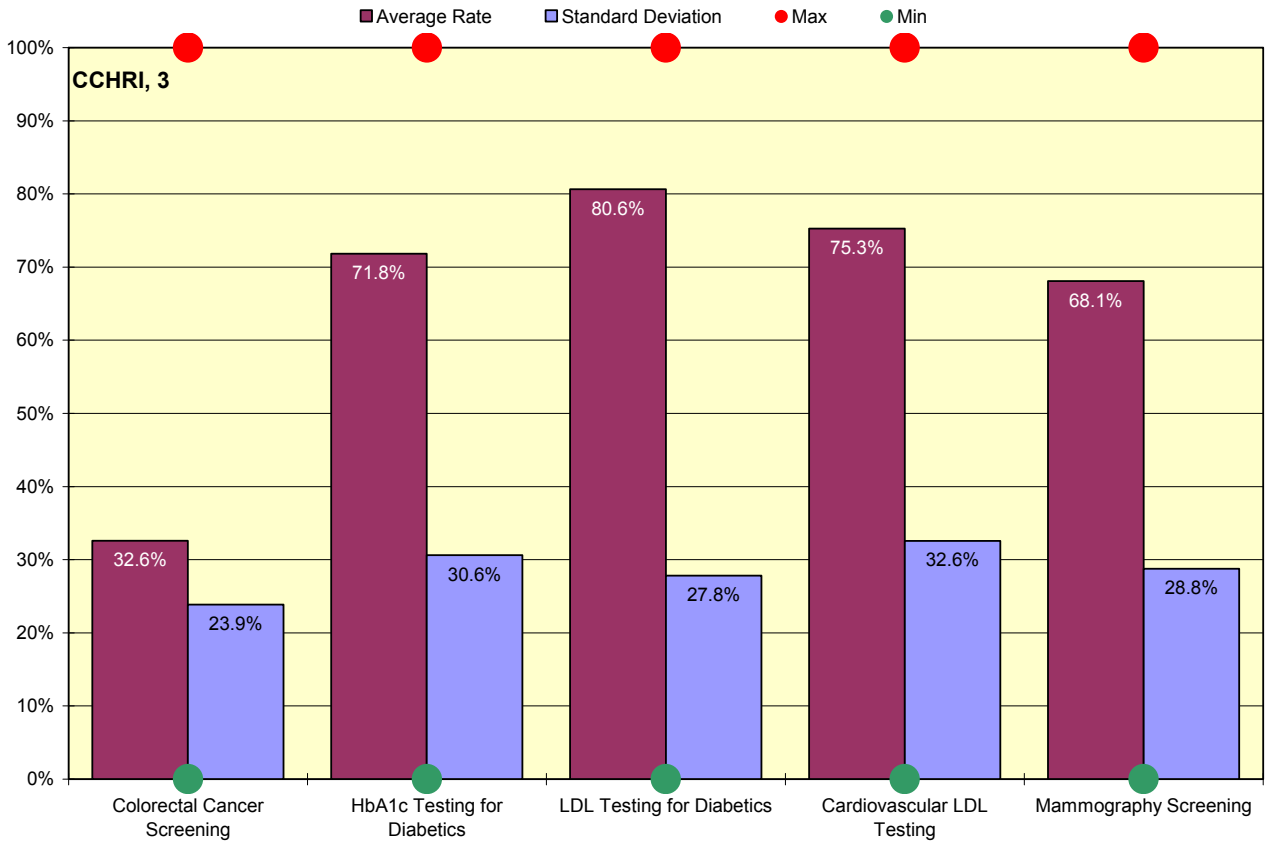

Average Rate
Standard Deviation
Max
Min
MHQP
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
46.2%
12.8%
84.1%
7.5%
58.8%
7.4%
88.0%
8.3%
76.8%
9.9%
Colorectal Cancer Screening
HbA1c Testing for Diabetics
Eye Exams for Diabetics
LDL Testing for Diabetics
Mammography Screening


Average Rate
Standard Deviation
Max
Min
MNCM
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
80.9%
12.0%
83.8%
10.0%
64.8%
8.9%
68.8%
11.9%
77.4%
10.7%
HbA1C Testing for Diabetics
LDL Testing for Diabetics
Eye Exam for Diabetics
Nephropathy Testing for Diabetics
Mammography Screening


Average Rate
Standard Deviation
Max
Min
WCHQ
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
4.0%
11.6%
19.5%
28.6%
48.3%
7.2%
24.9%
32.7%
6.6%
14.4%
Colorectal Cancer Screening
HbA1c Testing for Diabetics
HbA1c Control for Diabetics
LDL Testing for Diabetics
Mammography Screening


Average Rate
Standard Deviation
Max
Min
CCHRI, 1
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
36.3%
25.7%
69.8%
29.7%
78.6%
27.8%
72.4%
31.9%
69.3%
28.0%
Colorectal Cancer Screening
HbA1c Testing for Diabetics
LDL Testing for Diabetics
Cardiovascular LDL Testing
Mammography Screening


Average Rate
Standard Deviation
Max
Min
CCHRI, 2
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
35.6%
22.2%
70.6%
28.2%
79.8%
26.0%
72.9%
31.1%
69.5%
25.8%
Colorectal Cancer Screening
HbA1c Testing for Diabetics
LDL Testing for Diabetics
Cardiovascular LDL Testing
Mammography Screening


Average Rate
Standard Deviation
Max
Min
CCHRI, 3
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
32.6%
23.9%
71.8%
30.6%
80.6%
27.8%
75.3%
32.6%
68.1%
28.8%
Colorectal Cancer Screening
HbA1c Testing for Diabetics
LDL Testing for Diabetics
Cardiovascular LDL Testing
Mammography Screening

Appendix 9

Final Cycle 2 Data Aggregation Report
Report Appendix

Appendix A: BQI Project Measures

Appendix B: BQI Cycle 2 Measures, by Pilot

Appendix C: Pilot Physician-Physician Group Attribution Methods

Appendix D: Physician Group Attribution Results

Appendix E: List of Evaluation and Management Codes used to identify physicians eligible to be attributed to beneficiaries

Appendix F: Definition of PCP by Provider Specialty and Measure

Appendix G: Attribution Analysis Results

Appendix H: Attribution & Reliability Analysis Results

Appendix I: Reliability Scores for Selected Denominators

Centers for Medicare &
Medicaid Services

Better Quality Information to Improve Care for Medicare Beneficiaries

HHSM-500-2006-MD002C
Modification No. MD0008

Delmarva Foundation
9240 Centreville Road
Easton, MD 21601

July 25, 2008

Better Quality Information to Improve Care for Medicare Beneficiaries

(BQI) Project

Final Cycle 2 Data Aggregation Report

This Cycle 2 Data Aggregation report has been prepared by Delmarva for the purpose of providing a comprehensive picture of the activities, accomplishments, and challenges associated with the second round of data aggregation. This report describes the experiences of all six pilots, and contains the following sections:

Executive Summary 3
Summary and Common Themes 5
Working with the Medicare Data Challenges and Lessons Learned 13
Pharmacy Data: Part D Claims and Enrollment Files Experience 16
Data Aggregation and Analysis Challenges and Lessons Learned 17
Comparison of Cycle 1 Experience to Cycle 2 Experience 21
Experiences with Producing Individual Physician Level Results 22
Preliminary, Unvalidated Cycle 2 Measure Results 27
Comparing Medicare Results to Other Payers 33
Analyses to Address Methodological Questions 35
Defining Physician Groups to Report Quality Measures 35
Attribution of Beneficiaries to Physicians Preliminary Analysis Results 43
Reliability and Minimum Denominator Size Preliminary Results 50
Conclusion 55

Appendices

Appendix A: BQI Pilot Measures ..... A 1-2
Appendix B: BQI Cycle 2 Measure Specifications ..... B 1-6
Appendix C: Pilot Physician-Physician Group Attribution Methods ..... C 1
Appendix D: Physician Group Attribution Results ..... D 1-8
Appendix E: List of Evaluation & Management Codes ..... E 1
Appendix F: Definition of Primary Care Providers, by Provider Specialty and Measure ..... F 1
Appendix G: Attribution Analysis Results ..... G 1-5
Appendix H: Attribution & Reliability Analysis Results ..... H 1-56
Appendix I: Reliability Scores for Selected Denominators ..... I 1-30

Executive Summary

The Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project is a special study funded by the Centers for Medicare & Medicaid Services (CMS) under the Quality Improvement Organization (QIO) 8th Statement of Work (SOW). The BQI project will analyze, evaluate and report the advantages and challenges of six separate regional pilots ranking the quality of outpatient ambulatory office medical care provided in their geographic regions.

A national system to measure the quality and value of health care would give patients and consumers the information they need to make informed decisions, as well as give care providers the measurements they need to manage and improve their performance. In addition to nationally recognized standard measures of quality, such a reporting system would require accurate and reliable collection of useful information. Experience in collecting and using medical quality information in the outpatient office setting is scarce.

Leaders from the AQA (formerly, the Ambulatory Quality Alliance), the CMS and the Agency for Healthcare Research and Quality (AHRQ) noted by the spring of 2006 that leading organizations in several regions around the United States were already independently developing quality reporting schemes for their markets. The concept of the BQI project grew out of the proposal that the existing experience of these pilot sites could be leveraged into a valuable, transferable knowledge base by:

1) Providing each regional site access to Medicare Fee-for-Service (FFS) data, allowing aggregation of multiple public and private sources of data regarding the care being delivered by physicians to Medicare beneficiaries in those communities, and
2) Evaluating and comparing the results, reliability, advantages and challenges of each existing method of quality measurement and reporting.

CMS assumed the funding and oversight responsibility, and the BQI project began in October 2006 through CMS' QIO 8th Statement of Work. The BQI project will be completed by October 31, 2008. The Delmarva Foundation for Medical Care (Delmarva), the established QIO for Maryland, was awarded the CMS contract for overall BQI project administration and management and the evaluation and reporting of results. Subcontracts were awarded to six regional sites to receive Medicare FFS data, aggregate that data with their existing databases, and report and evaluate the results. Those six sites are:

- Center for Health Information and Research, Arizona State University (CHIR)
- California Cooperative for Healthcare Reporting Initiative (CCHRI)
- Indiana Health Information Exchange (IHIE)
- Massachusetts Health Quality Partners (MHQP)
- Minnesota Community Measurement (MNCM)
- Wisconsin Collaborative for Healthcare Quality (WCHQ)

During the BQI project, the six pilot sites will each report on twelve quality measures, five of which are consistent across all six sites. The sites use differing techniques for collecting data, various regional data bases, and

independent self-designed ranking methodologies which pre-existed the BQI project. In spite of these differences, it is the goal of the BQI project to distill valuable lessons learned from the experiences of these sites, to leverage this experience by combining information from multiple public and private sources, and to transfer this knowledge to the next generation of medical quality measurement and reporting.

By completion of the BQI project in October 2008, each pilot site will have aggregated Medicare FFS and other data through at least three cycles. The first cycle reported five measures from each site, based on calendar year 2005 data. Cycle 1 was completed in November 2007. Cycle 2 of the BQI project was completed in Spring 2008 and each site reported results for twelve measures based on calendar year 2006 data. Four of the pilot sites (CHIR, IHIE, MNCM, and WCHQ) performed an additional data aggregation cycle and reported results in June-July 2008. The last cycle will be completed in the late summer of 2008, and will again report twelve measures from each site, this time based on calendar year 2007 data.

This Cycle 2 Data Aggregation Report has been prepared by Delmarva to provide a comprehensive picture of the activities, accomplishments, challenges and lessons learned from the second of at least three rounds of data aggregation. This report describes the experiences of all six pilots, and contains the following sections:

- Summary & Common Themes
- Measure Selection Challenges & Lessons Learned
- Working with the Medicare Data Challenges & Lessons Learned
- Data Aggregation & Analysis Challenges & Lessons Learned
- Comparison of Cycle 1 to Cycle 2 Experience
- Experiences with Producing Individual Physician Level Results
- Preliminary, Un-validated Cycle 2 Measure Results
- Comparing Medicare Results to Other Payers
- Analyses to Address Methodological Questions

Summary and Common Themes

Measure Selection Challenges and Lessons Learned

At the beginning of the BQI project, it was anticipated that the measures to be calculated would be taken form the list of AQA Starter Set of 26 Measures[1]. As the project evolved, and inclusion of Medicare FFS data became the focus, the number of appropriate measures in the AQA set was significantly reduced. Two principle reasons for this were the availability of necessary measure data in Medicare administrative databases, and a measure's applicability to the Medicare population. Of the original 26 measures, only 5 were selected for Cycle 1 (CY 2005 data.) In Cycle 1, there were three measures calculated across all six pilot sites, and in addition each site selected two additional measures that varied across the pilots, for a total of seven different Cycle 1 measures.

For Cycle 2, CY 2006 data was used. The six sites refreshed the measures they had used in Cycle 1 for the Cycle 2 data. In addition, for Cycle 2 there were two more measures that were calculated for all six sites, for a total of five measures used across all six sites. These five measures were Breast Cancer Screening, Colorectal Cancer Screening, LDL Testing for Diabetics, HbA1c Testing for Diabetics, and Nephropathy Testing for Diabetics. In addition, each pilot site selected and calculated seven additional measures for Cycle 2 that varied from site to site, for a total of 12 measures for each site in Cycle 2, out of an overall population of twenty-five measures.

In addition to the lessons learned in Cycle 1, Cycle 2 further illuminated the trade-offs between desire for consistent measures across all pilot sites, and the benefits of local relevance and acceptance of the selected measures. For example, sites that had access to outcome measures preferred these measures as more clinically relevant, while sites using administrative data only could not calculate measures that included outcomes. There were also variations in the definitions of some of the rates, based on the site-specific methods of collecting the data. This Cycle 2 report contains specific information about each pilot site's experience with issues of measure selection.

Working with the Medicare Data Challenges and Lessons Learned

The experience gained in Cycle 1 informed the data handling in Cycle 2, yet there remained many problematic data issues in Cycle 2, including several new ones.

Based on the experience of Cycle 1, the pilot sites and Delmarva investigated requesting data from CMS in a smaller, more efficient format. In the end five of the six pilots remained with the longer format because of concerns surrounding the required change process. Also, a conscious decision was made to change the method of pulling the Medicare FFS data for Cycle 2. In Cycle 1 data was pulled for all care provided within the pilot site's state, regardless of where the beneficiary lived. After prolonged analysis the consensus of the pilots and Delmarva was that a "patient-centric" approach was more relevant to the Medicare beneficiaries, and therefore for Cycle 2 and future cycles the data is requested based on the residence of the patient, even if the care is provided in other states.

[1] http://www.aqaalliance.org/files/RevisedStarterSetApril2006.doc

Two data aggregation challenges arose in Cycle 2 that significantly impacted the timeline of the project. One of these was related to a new encryption program implemented by CMS between the Cycle 1 and Cycle 2 data release. This program caused unanticipated errors in some of the statistical packages. The strategies to overcome this problem are described. The second challenge was a single line of code found to be missing from supplied statistical programs. This report describes the detection and needed corrective actions for this problem as well.

The most important lesson learned about data acquisition and processing continues to be how complex and subtle the technical issues are, and how easily unintended errors may be introduced. Meticulous attention to detail and double-checking processes are essential.

Cycle 2 included importing and processing Pharmacy Data from Part D claims for the first time. There was a significant learning curve for handling this data as well. There was no documentation of how each field in the data set was defined. Physicians and other providers were not identified by the name codes as other data sets, but rather largely by their Drug Enforcement Agency (DEA) numbers. Strategies to address these issues are reported.

Data Aggregation and Analysis Challenges and Lessons Learned

In addition to the issues discussed above, there was extensive experience and learning with the first use of pharmacy data from the initial Part D data sets. For Cycle 2, Part D data was to be used for some measures in the denominator, to define the patient with a certain diagnosis based on their medications. Part D data was also used in the numerator of other measures, such as measures requiring a prescription for a positive event.

This report provides detailed experience of each of the pilot sites with Part D data from the 2006 Calendar Year. The major lesson learned was that this year's data had very significant completeness issues. There appear to be a substantial number of Medicare beneficiaries that have Part D coverage for whom there is no claims data in the 2006 database. Since 2006 is the first year of Part D, this issue may be due to enrollment and/or a Part D claims data problem. Because of these data issues, Delmarva believes measures reported in Cycle 2 based on Part D data should be considered exploratory and not indicative of provider performance. CCHRI in California provided an extensive analysis of the issues surrounding Part D data from 2006, and this is included in the Cycle 2 report.

Comparison of Cycle 1 to Cycle 2 Experience

All six pilot sites had lessons learned from the additional measures used in Cycle 2, from data aggregation issues, and from Part D data as described above. In addition, several of the sites have other lessons learned by comparing Cycle 2 to their Cycle 1 experience.

In Massachusetts, MHQP revised its attribution rules for Cycle 2 to include additional specialists for certain measures when there was no primary care provider. They believe this improved the accurate picture of care provided to Medicare beneficiaries.

In Indiana, IHIE included for Cycle 2 measures that required lab results and pharmacy data which they obtained from their community-wide electronic medical record. Their most important lessons learned involved the acquisition of this data. For both lab and pharmacy data, there were gaps in data resulting from patients using providers that were not included in the electronic database.

Wisconsin provided an analysis of the effect of using patient-based data extraction in Cycle 2 compared to physician-based data extraction in Cycle 1. In particular they noted an improvement in physician identification and CMS-group identification with the patient-centric data method.

Experiences with Producing Individual Physician Level Results

When the BQI project started, it was primarily focused on reporting performance measures at the individual level. Over the course of the project, interest in physician group level reporting increased. Most of the pilot sites have expressed a strong preference for group-level reporting at the present time. Only in California, which has been reporting at the group level for several years, is there some interest in moving to individual physician-level reporting.

At the completion of Cycle 2, all six of the pilot sites have fulfilled their commitments to aggregate and report data at either the individual or group provider level as documented in the Technical Proposal of February 3, 2008 (Table 3-6, p.3-18). In this section of the Cycle 2 report we review the lessons, issues and concerns of each pilot site regarding the level of reporting, both for provider feedback and for public and stakeholder evaluation.

There are several reasons for the preference of most pilot sites for group-level measures. First, two of the pilot sites, Wisconsin and Minnesota, collect measures at the group level and therefore cannot drill down the measures to individuals. Second, physician office care is often done in a group environment with individual physicians intentionally handing off specific types of care to other members of the group. This indicates that the group level is a more natural and appropriate level to report on than the individual physician level. Third, there are concerns that small denominator sizes at the individual physician level produce performance measures that are not reliable. Last, physicians who are beginning measure reporting are more comfortable with the group level and the pilot sites value and need the active involvement of the physicians in their efforts, in order to be successful.

Preliminary, Un-validated Cycle 2 Measure Results

The preliminary and not fully validated, aggregated measure results from the six BQI pilot communities for CY 2006 (Cycle 2) data are presented in site-specific tables in this section. In general in Cycle 2 inclusion of the Medicare FFS data had varying impact on measures, with strong effect on some, both positive and negative, and only weak effect on others. All six sites have provided to Delmarva results stratified by payer when possible for Cycle 2, and analysis of this data will be included in the final report on the BQI project. In this Cycle 2 report, we include the results of an extensive analysis performed by MNCM of the impact of Medicare FFS on measures in their region, compared to four other insurance products. In addition, MNCM included further analysis of the five

measures from Cycle 1, using the Wilcoxon Sign-Rank test, and found that most clinics for most measures maintained their relative rank ordering when Medicare FFS data was added to their community data. They concluded that, in general, Medicare FFS beneficiaries received the same quality of care for these measures as other patient subgroups.

Analyses to Address Methodological Questions

Of the seven methodological questions identified for further study in the Cycle 1 report, this Cycle 2 report has presented in the previous sections further insights into two: the limitations of methods for making data requests and the comparison of measures for Medicare FFS and Commercial or Medicaid beneficiaries. The benefits and limitations of composite measures and the validation by doctors of claims-based measures will also be discussed in the BQI final report. In this section and the related appendices, Delmarva presents extensive further work to date on the final three research issues, attribution of physicians to groups, attribution of patients to physicians, and reliability of measures.

Measure Selection Challenges and Lessons Learned

At the beginning of the BQI project, it was anticipated that the measures to be calculated would be taken from the list of AQA Starter Set of 26 Measures. This list represents measures that aligned with agreed-upon parameters and addressed agreed-upon specific conditions and areas[2]. However, while these measures can be used to measure ambulatory care performance in general, the BQI project discovered that not all of them were applicable to the project. The Data Conquerors (a workgroup comprising the technical and analytical representatives from the pilot sites and Delmarva) reviewed each measure in detail for applicability to the project. In evaluating the measures, the BQI project had to consider the availability of data as well as the applicability of the measure to the Medicare population.

After it became clear that the AQA starter set of 26 measures would not provide sufficient measures for the project (each pilot site was required to select 12 for the entire project), the pilot sites and Delmarva worked together to identify additional measure sets that could be reviewed for applicability to the BQI project. The measure sets reviewed include the National Committee on Quality Assurance (NCQA) Health Plan Employer Data and Information Set (HEDIS®) measurement set (for which physician-level specifications were in draft) as well as other National Quality Forum (NQF) endorsed measures, such as those from the Physician Consortium for Performance Improvement (PCPI). In addition, several pilot sites (MNCM, WCHQ) opted to include their measure specifications, which were largely based on existing measure sets such as NCQA HEDIS®.

The BQI pilot sites were active participants throughout the measure selection process. Cycle 2 includes a total of twenty-five measures, across all six pilots. Each pilot continued and refreshed the five measures they reported in Cycle 1, and each site also calculated seven additional measures for Cycle 2. The selection of measures varies by pilot; however, , as of Cycle 2, five measures are being calculated across all six pilot sites: Breast Cancer Screening, Colorectal Cancer Screening, Low-Density Lipoprotein (LDL) Testing for Diabetics, Hemoglobin A1c (HbA1c) Testing for Diabetics, and Nephropathy Testing for Diabetics. The pilot's selection of their additional measures was based not only on data availability, but also relevance to and acceptance by the local community. For example, some pilots have access to laboratory results data either from the labs themselves (IHIE) or through their data collection methodology (WCHQ and MNCM), and are therefore able to include disease control measures in their list. The full list of measures that are being calculated can be found in Appendix A.

While the measures across the pilots appear to be the same, there are differences across pilots in the detailed measure specifications. Since the pilots are aggregating the Medicare FFS data along with other payer data to which they have access, they must apply the same measure specifications that are used for the other payer data to the Medicare data. It was understood and accepted by all participants that there would be methodological differences across the pilots. Each pilot's selection of measures and specifications was discussed on the Data Conqueror's calls, and the group accepted these differences, acknowledging that it was important to address local community issues throughout the project. These variations must be taken into consideration when conducting any analyses on all the

[2] http://www.aqaalliance.org/files/AQAParametersforSelectingAmbulatoryCare.doc

data, as the variations severely limit, if not preclude, any measure rate comparisons across the pilot sites. Appendix B presents the measure specification details.

Each BQI pilot site expressed specific issues regarding the measures they selected for the project.

In Arizona, the most challenging aspect of measure selection was "identifying measures that reflect national guidelines, were applicable to the population of interest, and could be measured with available data." For example, CHIR chose not to select Pneumococcal Vaccination as a measure. The administration of this vaccination does not often get billed, thus evidence of its administration does not appear in administrative claims data that serves as the principal data source. In addition, while CHIR has continued its effort to work with a local laboratory, it has not yet been able to obtain any laboratory data to include in its aggregation activities.

At CCHRI, the first challenge with Cycle 2 measure selection was finalizing its 12th measure. Upon review of the measure specifications for its measures, CCHRI discovered an issue with the originally selected 12th measure (Osteoporosis Management for Women with a Fracture). The measure specifications dictate a look-back period that requires 2005 pharmacy data to ensure that a patient with a fracture that qualifies for the measure was not on osteoporosis drugs during the year prior to the fracture. Since the Medicare pharmacy benefit did not begin until 2006, the data available were insufficient to adequately calculate the measure. In addition, intuitively, not being able to look far enough back in the claims could have a material impact on rates since it may be very likely that, if a woman had a fracture, she was already on an osteoporosis drug (and more likely to be positive for the numerator, increasing the rate). Not being able to exclude these patients could result in higher rates, potentially changing the meaning of the measure.

Initially, CCHRI's Physician Advisory Group (PAG) recommended replacing the osteoporosis measure with the "Pharmacotherapy Management of Chronic Obstructive Pulmonary Disease (COPD) Exacerbation (PCE)" measure, after following these parameters when reviewing measures for inclusion:

1. AQA approved, or NQF endorsed, or part of the Physician Quality Reporting Initiative (PQRI)[3], or approved by Delmarva,
2. Relevant to Medicare population, and
3. Obtained by administrative claims data.

The PAG preferred to move forward with Pharmacotherapy Management of COPD Exacerbation (PCE) instead of other measures (see below). The PAG felt the COPD population is high risk and measuring this could provide very useful information for patients with this diagnosis. While this measure meets many of the AQA parameters for selecting physician performance measures[4], it is neither AQA nor NQF endorsed. Its selection for the BQI project was not supported by CMS since other NQF endorsed and/or AQA approved measures were available for selection.

[3] http://www.cms.hhs.gov/pqri
[4] http://www.aqaalliance.org/files/AQAParametersforSelectingAmbulatoryCare.doc

CCHRI reviewed other AQA approved/NQF-endorsed measures, as well as the list of measures being calculated by other pilots, and eliminated the following from consideration:

- *Glaucoma screening in older adults* - This measure depends on vision data which may be missing. In addition, CCHRI does not report this measure on the commercial side because the clinical evidence base is relatively weak.
- *Disease Modifying Anti-Rheumatic Drug (DMARD) Therapy in Rheumatoid Arthritis (ART)*- in this case there is some controversy about the clinical evidence base. In addition, the eligible population rate is low and will result in low denominators at the physician level.
- *Drugs to be avoided in the elderly* - the PAG felt that this measure was rather controversial, and relatively new for inclusion.
- *Beta Blocker treatment after Heart Attack* – CCHRI chose not to move forward with this measure because the denominator at the physician level of reporting will be too low to make this a useful measure. In addition, nationally rates are already 97%, and the measure is set to be retired by NCQA.
- *Eye Exam for Diabetics* - CCHRI opted not to consider retinal eye exam for diabetics due to the fact that there is limited vision data and the measure ideally requires chart review or having lab values for determining low risk status.

CCHRI felt that measures related to COPD were of most relevance to its community, and suggested selecting the "Use of Spirometry Testing in the Assessment/Diagnosis of COPD" measure, which was also being calculated by MHQP. CMS approved the selection of this measure, but in the timeframe between its submission to CMS and the CMS response, CCHRI confirmed that the data to support the measure were not collected on its commercial Preferred Provider Organization (PPO) side (their other data source), and that the measure requires up to 3.5 years of data history. In addition, the CCHRI Health Maintenance Organization (HMO) reporting committee has determined that the measure is not suitable for reporting, has a weak evidence base, and is not very relevant to the community.

After further discussion, it was agreed that CCHRI would include the "Beta Blocker Treatment after Heart Attack" in Cycle 2 data aggregation. Two other pilots (IHIE and MHQP) had also selected this measure for Cycle 2 data aggregation. After finalizing the 12th measure, CCHRI moved forward with its Cycle 2 aggregation and analysis activities. CCHRI has discovered potential issues with several of measures, the details of which are provided in the "Data Aggregation and Analysis Challenges and Lessons Learned" section.

At IHIE, the largest challenge it faced when selecting measures was data robustness, especially with regard to the "Beta Blocker Treatment after Heart Attack" and the "Persistence of Beta Blocker Treatment after Heart Attack" measures. Pharmacy data is and has been a challenge, since "consumers have various ways of getting prescriptions that are not always [captured] in the [administrative] claims systems." In addition, "Retinal Eye Exams are also a challenge to capture due to the fragmented health care system that exists today," and "measures that rely on laboratory results… are also a challenge due to the various data collection and processing as well as billing practices." With access to the Indiana Patient Care database (INPC), IHIE "felt confident that their lab results would be more robust with the connections to various facilities around central Indiana that participate in the INPC.

IHIE still faces challenges in getting several Laboratories to participate and to get outpatient lab results so those rates continue to be lower unless providers use the Point of Care data capture system or the Reconciliation Process that IHIE has in place to capture those areas of fallout."

At MHQP, measure selection was driven not only by the parameters already discussed (relevancy to the Medicare population and the community, projected to have a large enough denominator in commercial and Medicare populations to be reportable), but also by the lack of available lab data in the community. MHQP asked its plan representatives about the data, and discovered that the information was not consistently provided for all patient populations. MHQP has begun to test the feasibility of obtaining lab results for several of its large groups; this effort is underway and results will be presented in the final BQI project report.

In addition to examining data availability, MHQP vets the measure choices with its Physician Council. "One member was particularly concerned about including patients that were largely nursing home bound who might receive needed care in the Nursing Home or a Hospital and therefore might not show up as having had a test that is listed. This [Physician] Council member also strenuously argued that a terminally ill patient in a nursing home might be better served by NOT getting their mammogram." To mitigate this potential issue, Medicare Skilled Nursing Facility (SNF) claims were included in Cycle 2 in an attempt to help identify and eliminate patients who were in a nursing home for 45 days or more during the measurement period.

MNCM has not experienced much difficulty in measure selection following the parameters above; however, the pilot reports that the administrative-only measures have not been met with much enthusiasm. MNCM has moved towards increased reporting of outcome measures since it began reporting on Diabetes outcomes in 2002. For Cycle 2, MNCM was encouraged to select outcome measures if at all possible. The MNCM "Optimal Diabetes Care" measure, which includes three "outcome" components (HbA1c control, LDL control, and blood pressure control), met the pilot's desire to report outcome measures while still including measures relevant to, and calculable for, the Medicare population.

WCHQ did not experience any significant challenges selecting measures.

Working with the Medicare Data Challenges and Lessons Learned

Based on lessons learned during Cycle 1 aggregation and analysis, Delmarva submitted a modified data request for Cycle 2 data (calendar year 2006), on behalf of the pilots. There were two significant changes made to the Cycle 2 data request: file formats, and method of pulling the data.

The Cycle 1 Standard Analytic Files (SAFs) contained more data than was required to calculate measures, and were therefore very large and cumbersome. The pilots effectively determined which fields were required during Cycle 1, and were given the option of receiving the data in the same format as Cycle 1 (SAFs), or receiving the data files via the QIO Standard Data Processing System (SDPS) ad hoc request format, which is an abbreviated, fixed-width ASCII text format. After much discussion of the file formats on the Data Conquerors calls, only one pilot (MNCM) opted for the QIO ad hoc request format for subsequent aggregation cycles. The remaining five pilots determined that it would be less efficient and cost-effective to change data formats in the middle of the project, and requested that they receive the data in the same format as Cycle 1.

For Cycle 1, the claims files were pulled by the provider state of residence. Pulling the data in this fashion introduced what has been termed the "snowbird/robin/border doctor" issue, as the claims files may be missing any or all of the following claims:

- For any beneficiaries serviced by in-state providers, claims for services they received from an out-of-state provider. ("snowbird" issue)
- For any in-state resident beneficiaries, claims for services they received from an out-of-state provider. ("snowbird" issue)
- For any providers who are part of the pilot communities whose state of residence is across the border from the pilot state, claims for services billed by providers in the border state.

The BQI Data Conquerors group discussed the relative merits of the method of selecting the data, and all pilots agreed that for Cycle 2 (and subsequent) rounds of data aggregation, the data should be requested based on the beneficiary's state of residence (i.e., for IHIE, claims for all Indiana beneficiaries were requested, regardless of where the beneficiary received the care).

To accommodate the change to the "beneficiary centric approach," Delmarva added a step to its data request process: creating "finder files" of Medicare beneficiaries for each pilot state. This effort involved accessing the Medicare Beneficiary Database, through the QIO network, and pulling the Medicare numbers and sex of all beneficiaries who were recorded as having been eligible for Medicare in the reference year, 2006. One "finder file" was created for each of the six pilot states. These finder files were used to identify the related claims with service dates in 2006.

In addition to making the changes to the data request described above, two new claims types were introduced in Cycle 2: Skilled Nursing Facility (SNF) claims and Pharmacy (Part D) claims.

During Cycle 1 aggregation, there was considerable discussion regarding the inclusion or exclusion of institutionalized beneficiaries in the measure denominators. There was concern that a primary care provider for a beneficiary in an institutionalized setting may no longer be able to deliver the care once in the setting, but would still be held accountable. The pilots and Delmarva decided that they would investigate ways that institutionalized beneficiaries might be excluded from the Medicare FFS data; however, the group felt it might be challenging. As such, Delmarva requested the SNF claims for Cycle 2, to aid in the investigation of this issue.

Pharmacy claims were available to the BQI project through the QIO ad hoc data request process described above, therefore Delmarva submitted requests for the Part D claims for all six pilots, based on the six finder files described above. In addition, for MNCM only, Delmarva submitted a separate ad hoc data request for Inpatient, Outpatient, Carrier, and SNF files. [Since the other five pilots had opted to continue receiving the data in the Cycle 1 formats, Delmarva submitted a formal data request to the CMS Office of Information Services, through the Government Task Leader, for these files.]

Table 1 summarizes the data requests submitted for Cycle 2 claims data files.

Table 1: Claims Files Received in Cycle 2

Claim type	Dates	Format	Pilots	Delivery method
Carrier	2006	EBCDIC	AZ, CA, IN, MA, WI	Encrypted tape cartridge
Carrier	2006	ASCII	MN	Delivery to QIO database server
Inpatient	2006	EBCDIC	AZ, CA, IN, MA, WI	Encrypted tape cartridge
Inpatient	2006	ASCII	MN	Delivery to QIO database server
Outpatient	2006	EBCDIC	AZ, CA, IN, MA, WI	Encrypted tape cartridge
Outpatient	2006	ASCII	MN	Delivery to QIO database server
SNF	2006	EBCDIC	AZ, CA, IN, MA, WI	Encrypted tape cartridge
SNF	2006	ASCII	MN	Delivery to QIO database server
Pharmacy (Part D)	2006	ASCII	AZ, CA, IN, MA, MN, WI	Delivery to QIO database server

In addition to the claims files, Delmarva requested and provided the following reference files:

- Master Physician file, all physicians (requested from CMS)
- Beneficiary denominator file (Delmarva pulled data from QIO databases)
- Beneficiary supplemental information (Delmarva pulled data from QIO databases)
- Part D enrollment file (Delmarva pulled data from QIO databases)

Since CMS had discontinued making updates to the physician group membership file and the group UPIN file, these files were not provided in Cycle 2. As with Cycle 1, Delmarva converted the EBCDIC claims files for IHIE and MHQP. No conversion was necessary for the Part D claims files.

There were two significant issues with Cycle 2 claims files that impacted the original project timelines. These issues affected all the pilots.

The first issue involved a change to the "header" information in the EBCDIC claims files. In Cycle 1, Delmarva received unencrypted tape cartridges. However, effective January 1, 2007, CMS required that all portable media be

encrypted. However, CMS did not itself have the technology to perform the encryption, therefore, that process was contracted out to RDDC/Acumen. This change in process resulted in a change in the header structure of the EBCDIC files. As such, the SAS programs that Delmarva used to convert the IHIE and MHQP files, which were also used by CHIR and WCHQ, failed. After an email exchange regarding the issue, RDDC provided other SAS programs to Delmarva, which appeared to convert the files correctly. Delmarva used these programs to convert the IHIE and MHQP files, and passed the programs along to CHIR and WCHQ for their use.

MNCM was not affected by this change, as this pilot had already chosen to request the files in the "ad hoc" format. CCHRI, on the other hand, who has subcontracted its data aggregation work to Thomson Reuters, experienced significant delays in processing the data. Thomson uses its own data import software, which they were unable to modify to accept the new "header" information. As such, Thomson had to switch over to using a SAS process to import the files into its analytic software. This change resulted in a one-month delay in the timeline.

Unfortunately, subsequent to using these SAS programs to convert the IHIE and MHQP data, and distributing the programs to the other pilots (except MNCM), Delmarva discovered a problem with the Carrier program. Delmarva's research into the issue revealed that the SAS programs provided for carrier claims for Cycle 2 and forward were missing a critical "output" statement at the end of the program. The omission of this "output" statement meant that for claims with more than one HCPCS (procedure code) detail, only the **last** HCPCS code was being written to the output file, resulting in total record counts far lower than expected due to the missing data.

This issue was extremely hard to spot since the SAS programs ran without errors and created output files that had the correct variables. MHQP's aggregator (ViPS) had commented on the drop in the number of claims, however, since we had changed the method of requesting the data (from provider-centric to beneficiary-centric), the SAS programs ran without errors, and the records read by the programs matched the documentation provided by CMS, we did not suspect that the output files were missing claim line items.

The effect of this issue varied across pilot sites. The issue did not affect MNCM at all since this pilot had chosen to receive the "ad hoc" format files. For IHIE and MHQP, for whom Delmarva converted the claims, Delmarva created replacement files as per the pilot's preference: revised files containing all the claims and associated line items, or new files containing only the missing claim line items. CHIR and WCHQ used the corrected SAS programs, posted for all pilots' reference, to re-import their carrier files directly. CCHRI (Thomson Reuters) also used the corrected SAS programs to modify its import algorithms.

It is important to note that at the time of this report, Cycle 3 and Cycle 4 data (first round) had been received in the same formats as Cycle 2; therefore, these issues have been resolved. The CMS requirement for encryption occurred between 2 data request cycles; therefore, while it is frustrating, it is not surprising that the process changes that took place adversely affected the project.

A continuing issue is the timeliness and the currency of the Medicare FFS data, as compared to that of other payers. For example, at IHIE, its other payer data system is able to create measure reports that are based on 3-month old data. However, when adding the Medicare FFS data, the IHIE reports become at least 1 year old. In that community, the providers feel that reporting of dated information is not fair to the provider or to the patients. In addition, physicians are reticent to spend valuable time reconciling or justifying old results, and would rather focus on what they need to be doing to improve care today.

Pharmacy Data: Part D Claims and Enrollment Files Experience

Importing and processing the Part D claims and Part D enrollment files was simple (as compared to the other claims files). However, both of these datasets were new to all parties, and it has taken much of Cycle 2 to understand the contents of the files and assess their utility. For example, there was no data dictionary for the specific claims or enrollment files that Delmarva and the pilots received. Instead, Delmarva had to piece together documentation from various other sources. Delmarva sent this documentation to the Iowa Foundation (IFMC) for review. IFMC reviewed the document, added some clarifications, and thanked Delmarva for its efforts.

Delmarva and the BQI pilots also discovered that the Part D files provided through the QIO ad hoc process were not "final action paid claims," as was the case for the other claims files provided through OIS. Two pilots (IHIE and CHIR) first noticed duplicate values in the supposed key field for the dataset (RX_CLAIM_NUMBER). Delmarva asked IFMC to investigate the issue, and it was revealed that the files included adjustments and deletions (thus explaining the duplicate claim number). The claims files contain a final action flag, and using that flag and additional information received from IFMC, Delmarva was able to provide guidance to the pilots on how to remove adjustments and deletions as necessary.

Interpretation of the contents of the Part D dataset required additional effort. The Part D claims data uses an 11-digit National Drug Code (NDC) to identify the prescribed drug. Some pilots were not familiar with this code; therefore Delmarva provided a link to the NDC directory for reference[5].

CHIR and IHIE identified, and Delmarva confirmed, that the prescribing provider identifiers on the Part D claims were not UPINs, TaxIDs, or Group UPINs. Most of the time, the prescribing provider identifier on the Part D claims was the provider's Drug Enforcement Agency (DEA) number. For example, IHIE found that in 99.6% of their Part D claims, the prescribing provider ID type was the DEA number. MHQP confirmed the same issue with the Medicare Part D data but also sees the issue with its other payer data. Since there is no master DEA directory that will ultimately link back to the MPIER file, there is no way to link the claims back to a specific physician or physician group. This inability to link the prescribing provider on the Part D claims file precludes the use of the Part D claims data for patient to physician attribution (for example, using the most recently billed claim as a tie-breaker). However, since the Part D claims data are used to identify beneficiaries with specific prescriptions, this issue with provider identifiers does not impact the ability to calculate measures.

[5] http://www.fda.gov/cder/ndc/

CCHRI and MHQP did not note any additional challenges with the Part D data, other than what has been listed above. MNCM used the Part D data in the identification of the diabetes denominator but did not report any challenges with using the data. WCHQ did not use the Part D data as described further in the next section.

More discussion on the analytic impacts of using the Part D data can be found in the next section.

Data Aggregation and Analysis Challenges and Lessons Learned

As mentioned above, there were significant challenges using the Cycle 2 data as provided. The BQI pilots and Delmarva worked through these issues and eventually were able to use the data in analysis. This analysis revealed additional challenges and considerations when aggregating Medicare FFS data with other payer data to calculate quality measurement results. Most importantly, Cycle 2 data aggregation and analysis revealed substantial issues with using the Part D claims and enrollment files for quality measurement.

In Cycle 2, the Part D data has been used for two specific reasons:

1. Definition of the denominator: For all the diabetes measures, the measure specifications include the presence of a prescription as a qualifying denominator event. Several other measures use the existence of a prescription in the denominator definition criteria.
2. Inclusion in the numerator: Several of the Cycle 2 measures require a prescription for a positive numerator event.

At MNCM, the Part D data were only used to enhance the definition of the diabetes measure denominators, and this pilot experienced no issues with using the data for this purpose. None of MNCM's selected measures required pharmacy claims to determine whether a patient had met the numerator criteria.

WCHQ did not use the Part D data in its analyses. There are no criteria in the WCHQ measurement specifications that require pharmacy data for either patient to physician group attribution or for numerator inclusion.

Those pilots that are relying on pharmacy claims for their measure results have expressed grave concerns with the results. CCHRI's extensive Cycle 2 analysis has revealed data completeness issues, both in the Part D enrollment and the Part D claims files. These issues severely limit the ability to use Part D dependent measures for physician and/or physician group performance.

When CCHRI reviewed their preliminary, draft, Cycle 2 measure results, they noted that several of the measure results, for the Medicare FFS population, were surprisingly low.

For example, CCHRI looked closely at the patients who qualified for the denominator for the "Coronary Artery Disease (CAD) patients receiving lipid-lowering therapy" measure. For inclusion in the denominator, members must have 11 months of Part D enrollment. For inclusion in the numerator, members must have a prescription for

a statin. CCHRI found that 55% of the denominator members had no prescriptions at all, for any drug. Given that Medicare PPO members have an average of about 30 prescriptions a year, they probed further. They first examined the numerator definition: beneficiaries with at least one prescription in 2006; then examined the denominator definition: beneficiaries continuously enrolled for 12 months with Medicare Part A, B and D coverage.
The results showed that only 39% of beneficiaries with Part A, B and D coverage had at least one prescription in the data (defined as both covered by Part D and available in the data). Then, CCHRI analyzed the average number of prescriptions for beneficiaries who had at least one prescription during the year. The average was 52.6 prescriptions per beneficiary.

This analysis implies that there are a substantial number of beneficiaries that have Part D coverage but for whom there is no claims data. This issue may be a Part D enrollment and/or Part D claims data problem. Some Part D plans may have submitted incomplete/erroneous data to CMS. Alternatively, beneficiaries may have signed up for Part D coverage but continued to receive and pay for their drugs through alternate plans or independently.

However, for those beneficiaries for whom any claims data exists, the data may be relatively complete; given the average number of prescriptions per beneficiary, per year is 52.6 prescriptions per beneficiary.

The analysis implications are that the Medicare rates for measures that are pharmacy dependent are substantially understated due to apparently missing pharmacy claims data. The level of understatement probably varies from measure to measure, depending on the specifications.

Table 2 presents the potentially affected measures (that is, those that use Part D data), by pilot.

Table 2: Part D Dependent Measures

Measure	Part D used for denominator	Part D needed for numerator	AZ	CA	IN	MA
Nephropathy testing	✓		✓	✓	✓	✓
CAD patients receiving lipid-lowering therapy		✓	✓	✓		✓
Beta blocker treatment after heart attack		✓		✓	✓	✓
Persistence of beta blocker therapy -Post MI	✓	✓	✓	✓	✓	✓
Heart failure: ACE inhibitor/ARB therapy		✓	✓			
Warfarin for patients w/ CHF & atrial fibrillation		✓	✓	✓		
Antidepression Medication Management – Acute Phase		✓	✓			
Annual monitoring for patients on persistent medications: Ace inhibitors, Digoxin, Diuretics, Anti-convulsants, Statins	✓	✓		✓		✓

CCHRI's extensive analysis of several Cycle 2 measures reveals the following issues with certain measures:

Measure: Colorectal Cancer Screening
Medicare Cycle2 result: 35.04%
Medicare Cycle2 benchmark[6]: 53.20%
Difference: -18.16%

[6] Medicare HMO national rates

Discussion:
CCHRI has expressed its concern that the colorectal cancer screening rate based on claims data alone understates the true rate by a significant margin and may lack face validity with physicians. CCHRI is only able to examine a 2-3 year look back period when relying on claims data to construct this measure, whereas the specification allows for screening within the past 10 years. When comparing the rate calculated based on Medicare FFS claims to an external benchmark (i.e., the Medicare HMO national rates, which include chart review) the rate is 18 percentage points lower.

Measure:	Nephropathy Testing for Diabetics
Medicare Cycle2 result:	54.63%
Medicare Cycle2 benchmark[7]:	85.30%
Difference:	-30.67%

Discussion:
The measure is run using the NCQA HEDIS® specifications, which do not require a pharmacy benefit; however, one of the ways to satisfy the numerator is an ACE/ARB prescription. So some Medicare beneficiaries in the denominator might have received an ACE/ARB but don't have prescription coverage. This will result in the rate being understated. Based on CCHRI's analysis of the Part D enrollment file, only 32% of beneficiaries had a prescription benefit in 2006. This suggests that the 2006 Part D data may not be correct and/or complete. Without being sure that the data are correct and/or complete, the currently calculated testing rate for the Medicare population is likely to be substantially understated than the actual rate. This problem is not observed with the CCHRI commercial data, where there are more complete pharmacy data. When compared to the Medicare HMO rate (which includes chart review), the Medicare FFS rate is 31 percentage points lower.

Measure:	Beta Blocker Treatment after Heart Attack
Medicare Cycle2 result:	19.6%
Medicare Cycle2 benchmark[8]:	93.7%
Difference:	-74.1%

Discussion:
The claims-based version of this measure results in a lower rate compared to the benchmark rate. The Beta Blocker Treatment after a Heart Attack (BBH) measure doesn't require prescription benefits coverage for the member to get into the denominator. This is because the NCQA HEDIS® measure specification can be done through chart review; therefore, even if the member doesn't have prescription coverage, the prescription would be found during the chart review. In California, the HMO rate including chart review is about 18 percentage points higher than the administrative-only rate. The really low rate on this measure for the Medicare population is most likely a function of the fact that most Medicare beneficiaries in the 2006 files did not have Part D prescription coverage, and thus the

[7] Medicare HMO national rates
[8] Medicare HMO national rates

pharmacy claims for many beneficiaries included in the denominator do not exist in the Part D claims files that were provided [the prescriptions, however, could very well have been filled through other means not captured in the Part D claims data]. Additionally, NCQA has retired this measure given that rates based on chart review are at 98%. This measure is also problematic because the number of beneficiaries who qualify for the measure is quite low (n=11,390 in California for measurement year 2006, with 10,494 being attributed to physicians of a relevant specialty for the measure). When examined in terms of the average number of events assigned per doctor for this measure, the minimum was n=1, the maximum was n=11, and the average was n=4.4 patient events based on aggregated commercial and Medicare data.

Measure:	Persistence of Beta Blocker Therapy Post MI
Medicare Cycle2 result:	48.4%
Medicare Cycle2 benchmark[9]:	69.6%
Difference:	-21.2%

Discussion:
Though CCHRI applied the specification that limited the denominator to only those with a prescription benefit, they still observed a substantial discrepancy against the external benchmark comparison (Medicare HMO because no Medicare PPO benchmark is available). The difference could reflect true differences between Medicare FFS and Medicare HMO. This measure is problematic because of the small number of beneficiaries who qualify for the measure (n=1,952 in California for measurement year 2006). When examined in terms of the average number of events assigned per doctor for this measure, the minimum was n=1, the maximum was n=3, and the average was n=1.4 patient events based on aggregated commercial and Medicare data.

MHQP reported a specific data issue with the Retinal Eye Exams for Diabetics measure, not related to the Medicare FFS data, but affecting the aggregated results. For one plan, MHQP observed that many of the claims for the eye exam measure were not rendered by eye doctors, but rather other types of doctors, including some gastroenterologists and radiologists. That plan confirmed the correctness of those rendering doctor types. This confirmation meant those claims could not be included in the measure calculation, since the measure specification requires that the exam be given by an ophthalmologist or an optometrist. Of the eye exam claims, only 8.9% were given by a provider with a specialty of ophthalmology. Since this number seems low, MHQP has asked the plan to further investigate the issue. At the writing of this report, MHQP continues to investigate the issue with the plan.

As mentioned above, for Cycle 2 aggregation, the SNF claims were requested. These were requested to address the concern about including institutionalized patients in the measures. We await MHQP's findings relating to the inclusion of these data and anticipate their inclusion in the final BQI report.

IHIE's challenges with using the Part D can be found in the next section. CHIR did not note any additional challenges with Cycle 2 aggregation other than those listed in other sections of this report.

[9] Medicare HMO national rates

Comparison of Cycle 1 Experience to Cycle 2 Experience

This report has already discussed the differences in the way the data were requested between Cycles 1 and 2, as well as the issues surrounding the introduction of new datasets. In this section we present some observations from the pilots regarding their Cycle 1 and Cycle 2 experiences.

In Arizona, CHIR's experience was similar with the exception of the introduction of the new datasets. The challenges CCHRI experienced in Cycle 2 as compared to Cycle 1 have been addressed in the previous section. MHQP's Cycle 2 experience was similar to Cycle 1, with the exception of the introduction of the new datasets and the observations discussed in the previous section. In addition, for the new Cycle 2 measures, MHQP revised its attribution rules to include additional specialists where there were no primary care providers for the cardiac measures (cardiologists), and the Spirometry measure (pulmonologists). They added Obstetricians and Gynecologists (Ob-Gyns) for the breast cancer measure, and continued the attribution to endocrinologists for the diabetes measures, as in Cycle 1. These new rules were processed with their "rapid response team," representing their Physician Council.

In Indiana, Cycles 1 and 2 were very similar in that they continued to use existing measures and conduct aggregation in the same manner, but very different in what experiences and lessons they learned from conducting the Cycle 1 data aggregation. Cycle 2 included measures that relied on lab results and pharmacy data which included additional testing and programming efforts. There were many lessons learned, but the most important ones revolved around laboratory and pharmacy data acquisition. IHIE discovered that laboratory data is not always easily obtained as there are various sources and ways to collect lab data. Pharmacy data is similar in that there are various ways for patients to obtain prescription medications including pharmaceutical company samples and local deals offered at grocery stores and department stores that are less expensive to purchase but these are very difficult to track in a system dependent on electronic data capture.

At MNCM, Cycle 2 data aggregation went relatively smoothly. While programming algorithms had to be modified to accommodate the "ad hoc" data files, these changes were relatively simple.

In WCHQ, the data aggregator (CHSRA) provided observations about the Cycle 2 experience as compared to Cycle 2. In Cycle 1 claims data was selected using Date of Service (DOS) and Wisconsin physicians as the selection criteria while in Cycle 2 the claims were selected based on DOS and Wisconsin patients. The preliminary observation is that, as expected, the difference in selection criteria was reflected not so much in performance, although that improved slightly year to year. Rather, using the beneficiary-centric approach resulted in:

1. Fewer total claim lines, but much better identification of patients and obviously a huge improvement in the number of unidentified patients from out of state,
2. An improvement in physician identification and CMS-group identification,
3. Cycle 2 data revealed a significant jump in the average number of physicians treating each patient and revealed a bump, although also part of an overall trend, in the average claims lines per patient. These

increased claims lines may also simply reflect that Wisconsin citizens tend to stay home and thus extend the period of time they might require or request care.

4. Although more physicians seem to be seeing each patient and creating more claim lines, the actual percentage of the Medicare population with any claim has dropped steadily from 2003-2006.

Experiences with Producing Individual Physician Level Results

When the BQI project started, it was primarily focused on reporting performance measures at the individual level. Over the course of the project, interest in physician group level reporting increased. Most of the pilot sites have expressed a strong preference for group-level reporting at the present time. Only in California, which has been reporting at the group level for several years, is there some interest in moving to individual physician-level reporting.

There are several reasons for this preference for group-level measures. First, two of the pilot sites, Wisconsin and Minnesota, collect measures at the group level and therefore cannot drill down the measures to individuals. Second, physician office care is often done in a group environment with individual physicians intentionally handing off specific types of care to other members of the group. This indicates that the group level is a more natural and appropriate level to report on than the individual physician level. Third, there are concerns that small denominator sizes at the individual physician level produce performance measures that are not reliable. Last, physicians who are beginning measure reporting are more comfortable with the group level and the pilot sites value and need the active involvement of the physicians in their efforts, in order to be successful.

Pilot Sites that committed to a non-patient identifiable provider reporting file at the individual physician level

As agreed in the Technical Proposal dated February 13, 2008, (Table 3-6, p.3-18) two of the pilot sites committed to creating a reporting file of non-patient identifiable provider measures at the level of the individual physician. With Cycle 2, both of these two pilot sites, CCHRI and CHIR, have calculated their measures at the physician level and have created reporting files at the individual physician level which have been submitted with this Cycle 2 report.

CCHRI has been calculating and giving feedback to providers at the individual physician level for several years. To date, their public reporting has been at the group level only. CCHRI was the only pilot site to express that the providers in their market were requesting a move to public reporting of measures at the individual level. The reason stated was a feeling that progress based on professionalism and group self-improvement had reached a place where more public scrutiny of individual would be a useful motivator. The BQI project has afforded CCHRI the ability to test individual physician-level quality measure reporting. CCHRI is in the initial stages of this measure reporting effort. They have begun sharing individual-physician quality measure reports to the Medical Directors of a sample of physicians in California, as a way to pilot test the report process. The second phase of the feedback / reporting process will be a staggered mailing of reports to approximately 15,000 physicians. In both instances, responses and inquiries from the physicians are welcomed. The results of this pilot effort will be included in the final BQI report.

CHIR has completed measure calculation at the individual physician level, but has not yet begun validation or public reporting of the information. Public reporting in general has been an issue since the project's inception. CHIR has never engaged in public reporting before, and the climate in Arizona has not historically been conducive to public reporting. There has not been substantial change in this sentiment over the life of the BQI project. As a corrective plan to address the issue of public reporting, CHIR has employed multiple strategies to garner buy-in at the provider level and the institution level. CHIR has been working with the state medical associations for more than a year to establish buy-in and to help form their Expert Physician Panel. This engagement process took months, and continues to be impacted by work schedules as well as the overall perception of the project and its relationship to pay-for-performance.

Despite the organization's neutral stance, they continue to be met with what appears to be a philosophical opposition to public reporting; there has been voiced concern that the Federal government will use this mechanism to reduce reimbursement for services rendered. CHIR has worked with its Expert Physician Panel to devise a process by which physicians may feel confident in result reporting. CHIR felt the idea of physicians leading each other through the process with CHIR's help would improve the chance of buy-in, as the panel would be able to demystify the process for its peers and their medical associations would be supportive of the effort. When CHIR assembled the panel, it was clear that there was a wide range of understanding about all of the different initiatives they had heard of--there was also an evident range of interest. Therefore, CHIR selected one physician who met their criteria for ability to turn around the results as well as being "on board" with the idea. CHIR is engaging this physician as a "physician champion." Having worked through the process with him, CHIR had him present to the Center stakeholders as someone who had "lived" through the process. CHIR has secured two more physicians on the panel who are going to participate in the process in July. All three will attend a meeting in July to discuss their experiences with other interested parties.

The process developed by CHIR and its physician panel includes five steps:

1. Participating physicians select one of the twelve project measures for review,
2. Physicians who have been identified as a primary care physician, who have electronic health records (EHRs) or a similar mechanism to validate results shall be provided with lists of individual physician results,
3. Physicians will first validate the patient attribution methodology,
4. Physicians will verify the measurement methodology, and
5. CHIR will review the physician's process and findings.

This process seems to, for those who understand the overall purpose of the project, be effective when working with a physician or physician group one at a time. Unfortunately, the scope of BQI (both time required to secure participation and conduct validation) will not allow for results to be reviewed with each physician or even each physician group in order for all participants to feel that their results have been adequately validated.

Pilot sites that committed to aggregating and reporting patient-identifiable feedback to providers at the individual physician level

In the Technical Proposal of February 13, 2008, (Table 3-6, p.3-18) two of the four remaining pilot sites that planned to create reporting files only at the physician group level still committed to aggregating and giving feedback to the providers at the individual physician level. Of these two sites, IHIE has successfully begun the process as of this report, and has reported on its preliminary experience. WCHQ has identified logistical limitations to individual-level calculations, has analyzed the issues, and has planned and executed an alternative to individual level reporting in their community.

IHIE has been giving feedback data to physicians through its Quality Health First SM Program (QHF) for approximately one year. IHIE has about 900 physicians mostly from large area group contracts. IHIE has developed and provided education to designated quality managers in each physician group. The measures reported through the QHF Program are similar in many cases to those in the HEDIS measure specifications with which the physician groups have several years of experience. IHIE is now working with the quality managers of each physician group to validate the individual and group results. Some of the issues raised by providers in this early review of a new process include:

- Coding issues- patients may be included in denominators if their diagnosis coding was not correct. Education is being provided regarding accurate coding.
- Noncompliant patients- physicians feel that they should be given credit for recommending the correct test or treatment, and not held accountable for the patient actually accomplishing recommendation. Physicians accept the responsibility for diagnosis and treatment recommendations more than for effecting the treatment or test.
- Risk adjustment- providers perceive that their outcomes are lower because of unadjusted disease severity, socioeconomic factors, or patient compliance.
- Benefit packages affect access- restrictions on what tests and treatments are covered, and what the patient's co-payment is, effect the achieving of the recommended treatments.
- Data issues- questions around whether coding and billing data capture all treatments, including those paid for in cash or charity clinics
- Administrative burden of validation- working through issues surrounding new process and reports have been a major concern of providers. Time for validation has been an imposition and a provider dissatisfier.

At IHIE, the community feels that reporting must occur at the physician group level for several reasons:

1. The providers practice as a team, team effort is encouraged; therefore ratings should be for the team.
2. Physician extenders such as interns, residents, Nurse Practitioners, Physician Assistants and Nurse Midwives are not counted by CMS as providers, but assignment of the care provided by these individuals to any one member of the supervising team would be arbitrary and possibly distort the measurement of quality.
3. One patient may be attributed to multiple providers within a group, who all managed the patient at various times, and that this would result in this one patient's outcome having greater weight in the measures than a patient who was counted only once for the group.

WCHQ has identified some logistical limitations to fulfilling the commitment to report feedback to physicians at the individual physician level. The current process WCHQ uses to calculate measures is to receive data directly from providers as a numerator and denominator for each measure at the self-defined provider group level. Rates are then calculated and posted for public review on their website at www.wchq.org. As a result, WCHQ cannot calculate measure results at the level of the individual physician.

A second logistical issue is that each of the provider organizations participating with WCHQ has developed their own internal methodologies for attribution of patients to physicians within their groups. These attribution methodologies have not been reviewed or standardized by WCHQ. Trying to obtain this data from the organizations for the BQI project would therefore result in inconsistent data that would not be comparable site to site.

Based on the information WCHQ obtained in a literature review, their leadership is not fully convinced that measurement and reporting at the individual physician level have sufficient scientific rigor at this time. The WCHQ Board of Directors (BOD) is particularly concerned about some of the methodological and technological issues that come about when discussing the power of measures to assess individual performance due to sample sizes, sensitivity of measures and specificity of measures in small patient populations.

In addition, the WCHQ BOD feels strongly that the work of their organization has consistently promoted the idea of a care team, and that assessment of individual performance within this care team undermines the system of accountability to the team.

Because of the above concerns WCHQ does not intend to calculate measures and give provider feedback at the individual physician level at this time. As an alternative plan to address their commitment, WCHQ planned and sponsored a conference with the AHRQ to address the issues at the core of this decision. WCHQ feels this work will move the agenda forward toward building consensus and establishing some guidelines for future individual physician measurement. At this conference, there was agreement that there was value in physician level *measurement*, however, there was still dissention about the public release of such information. In addition, the participants agreed that a team approach to care giving is important to understand. However, more areas of further research existed than areas of agreement. Some of these areas of further research identified include:

- Regarding attribution: Is individual physician an adequate model, or a concept with individuals in separate practices? How can this be addressed?
- Is there a way to simplify measurement and reporting?
- How do we create a bridge between claims and clinical data?
- How do we develop standards that can be applied to the diverse physician practices across the US?
- How do we address the regional variations in coding?
- Have we done adequate research on the size of a practice and its impact on measure results?

Pilot sites that committed to reporting and giving feedback at the physician group level only

The remaining two pilot sites, MHQP and MNCM, have fulfilled their commitment in the Technical Proposal of February 13 (Table 3-6, p.3-18) to aggregate, report feedback to providers, and create provider reporting files at the group level only.

MHQP in Massachusetts has not previously done physician-level reporting for any of its annual HEDIS® measure reports. Though the measures are calculated down to the individual physician level, they are rolled up to a provider-defined group level for analysis and reporting. The MHQP Physician Council has concerns about the validity and reliability of the data at the individual physician level. MHQP has considered conducting studies around the validity and reliability of physician-level data in the future.

Based on its experience, however, MHQP believes that improvements in quality are more related to the organizational structure within which a physician practices than to the individual physician. Their reasoning regarding the delivery of care by teams and groups is similar to that reported by other pilot sites.

In Minnesota, MNCM also receives its data directly from provider groups as a group-level numerator and denominator, similar to the system in Wisconsin. For MNCM, the only means for reporting at the individual physician level would be through the expansion of their direct data submission process, and it would also require significant resources at the medical group level. This type of expansion work is not currently funded.

Beyond that, for the measures MNCM currently reports, their philosophy is that the unit of analysis and quality improvement is the clinic-site level, and they have recently considered moving towards a "team level." MNCM's chronic care measures, especially, are measures of a care system at a given medical group.

MNCM has thought that if and when they consider reporting specialty measures - especially those that are procedure based – these measures might be better suited to individual physician level unit of analysis.

Status of identifiable provider results supplied to Delmarva as of Cycle 2 report

Though CCHRI and CHIR have provided reporting files to Delmarva calculated at the individual physician level, the identities of the physicians are masked and not identifiable. Similarly, IHIE has provided reporting at the group level with the actual group identities masked. IHIE is still involved with its providers in validating the data and gaining provider buy-in. The remaining sites, MHQP, MNCM, and WCHQ have supplied Cycle 2 reporting files aggregated at the group level, with the groups identified.

Preliminary, Unvalidated Cycle 2 Measure Results

Despite the limitations and challenges encountered, the BQI pilot sites have been able to generate measure results, at either the physician group or individual physician level, and have provided their Cycle 2 results to Delmarva.

Descriptive statistics for these results are presented in Tables 3a-3f. Unless otherwise noted, the results presented are aggregated, multi-payer results. All results are not fully validated nor have the pilots completed all statistical testing (for validity and reliability).

Table 3a: Center for Health Information & Research, Cycle 2 Descriptive Statistics

Center for Health Information & Research (Individual Physician Level) Timeframe: 2006							
Measure	Average Rate (1)	Standard Deviation	Rate Range	Denominator Range	Average Denominator	Number of Physicians Reporting (2)	Number of Physicians (3)
Retinal Eye Exam	31.7%	25.3%	0.0% - 100.0%	1 - 389	20	1,233	1,334
HbA1c Testing	66.6%	23.5%	0.0% - 100.0%	1 - 414	29	1,306	1,334
LDL Testing	78.9%	20.5%	0.0% - 100.0%	1 - 414	29	1,306	1,334
Breast Cancer Screening	56.3%	23.0%	0.0% - 100.0%	1 - 210	25	1,321	1,334
Colorectal Cancer Screening	27.9%	14.1%	0.0% - 95.6%	10 - 898	110	1,334	1,334
Nephropathy testing	33.4%	31.7%	0.0% - 100.0%	1 - 266	14	1,098	1,334
Depression Medication Managementt	16.7%	30.6%	0.0% - 100.0%	1 - 39	4	372	1,334
Cardiovascular –LDL testing	65.5%	47.6%	0.0% - 100.0%	1 - 103	14	1,296	1,334
Beta blocker treatment after heart attack	41.8%	46.4%	0.0% - 100.0%	1 - 5	1	225	1,334
Heart Failure: ACE inhibitor/ARB therapy	47.6%	35.9%	0.0% - 100.0%	1 - 42	4	849	1,334
Warfarin for patients w/CHF & atrial fibrillation	57.0%	42.6%	0.0% - 100.0%	1 - 9	2	522	1,334
CAD patients receiving lipid lowering therapy	46.2%	29.0%	0.0% - 100.0%	1 - 142	10	1,098	1,334

(1) Average of individual physician rates.
(2) Number of physicians with at least 1 patient in the denominator.
(3) Number of physicians included in Cycle 2 data file.

Table 3b: California Cooperative Healthcare Reporting Initiative, Cycle 2 Descriptive Statistics

California Cooperative Healthcare Reporting Initiative (Individual Physician Level) - Physicians Meeting Minimum Reliability Level (1) Timeframe: 2006							
Measure	**Average Rate (2)**	**Standard Deviation**	**Rate Range**	**Denominator Range**	**Average Denominator**	**Number of Physicians Meeting Min Reliability Criteria (1)**	**Number of Physicians (3)**
Breast Cancer Screening	68.5%	14.8%	6.8% - 100.0%	17 - 1,055	75	13,085	25,647
Colorectal Cancer Screening	47.0%	13.7%	0.0% - 100.0%	22 - 2,251	166	16,754	24,221
LDL Testing	71.9%	12.0%	5.0% - 100.0%	20 - 629	80	8,192	22,186
HbA1c Testing	74.0%	10.8%	9.5% - 100.0%	24 - 629	88	6,735	21,659
Nephropathy Testing	63.1%	15.2%	9.1% - 100.0%	15 - 629	70	9,920	21,773
Cardiovascular - LDL Testing	76.3%	10.1%	20.5% - 100.0%	24 - 562	79	3,306	19,144
CAD patients receiving lipid lowering therapy	35.0%	12.3%	0.0% - 88.2%	21 - 562	69	4,289	20,120
Persistence of beta blocker therapy- Post MI	60.3%	0.0%	0.0% - 0.0%	0 - 0	0	0	4,605
HF - LVF Testing	43.8%	14.4%	5.1% - 89.8%	39 - 108	64	135	13,731
Warfarin for Patients w/CHF & atrial fibrilation	49.8%	13.5%	29.0% - 77.6%	58 - 86	67	16	13,738
Annual Monitoring for Patients on Persistent Medications (4)	86.9%	11.1%	16.7% - 100.0%	10 - 270	38	9,150	21,560
Beta Blocker treatment after heart attack	23.9%	0.0%	0.0% - 0.0%	0 - 0	0	0	11,489

(1) A physician's rate is deemed reliable if his or her sample size was at least as large as the sample size for which 90% of the physicians had a least a 70% reliability.
(2) Average of individual physician rates.
(3) Number of physicians included in Cycle 2 data file.
(4) ACEI, Digoxin, Diuretics, Statins

Table 3c: Indiana Health Information Exchange, Cycle 2 Descriptive Statistics

Indiana Health Information Exchange (Physician Group Level) Timeframe: 2006							
Measure	**Average Rate (1)**	**Standard Deviation**	**Rate Range**	**Denominator Range**	**Average Denominator**	**Number of Groups Reporting (2)**	**Number of Groups (3)**
Breast Cancer Screening	65.3%	12.7%	50.0% - 100.0%	1 - 5,866	1,657	12	12
Colorectal Cancer Screening	41.1%	17.1%	0.0% - 68.0%	3 - 21,178	5,511	12	12
Beta-Blocker treatment after heart attack	13.0%	17.4%	0.0% - 50.0%	2 - 29	16	7	12
Persistence of beta-blocker therapy- Post MI	39.5%	24.8%	27.0% - 100.0%	1 - 71	29	8	12
Cardiovascular LDL-testing	68.1%	14.2%	33.0% - 78.0%	6 - 4,585	1,317	11	12
Cardiovascular LDL-control	21.3%	14.0%	0.0% - 43.0%	6 - 4,585	1,317	11	12
HbA1c testing	63.8%	26.0%	0.0% - 100.0%	1 - 5,419	1,426	12	12
LDL-testing	42.4%	27.5%	0.0% - 100.0%	1 - 5,419	1,426	12	12
Diabetes LDL cholesterol control	10.6%	8.2%	0.0% - 26.0%	1 - 5,419	1,426	12	12
Nephropathy testing	52.5%	28.1%	0.0% - 100.0%	1 - 5,419	1,426	12	12
Retinal Eye Exam	85.8%	15.0%	42.0% - 100.0%	1 - 5,490	1,448	12	12
HbA1c control	11.5%	10.2%	0.0% - 28.0%	1 - 5,419	1,426	12	12

(1) Average of physician group rates.
(2) Number of groups with at least 1 patient in the denominator.
(3) Number of participating IHIE groups.

Table 3d: Massachusetts Health Quality Partners, Cycle 2 Descriptive Statistics

Massachusetts Health Quality Partners (Physician Group Level) Timeframe: 2006							
Measure	**Average Rate (1)**	**Standard Deviation**	**Rate Range**	**Denominator Range**	**Average Denominator**	**Number of Groups Reporting (2)**	**Number of Groups (3)**
Beta blocker Treatment after heart attack	68.9%	10.3%	40.0% - 100.0%	10 - 133	37	88	133
Breast Cancer Screening	75.3%	7.6%	38.6% - 96.4%	10 - 8,121	1,378	156	158
LDL testing	79.8%	7.8%	43.8% - 92.3%	10 - 4,999	766	155	161
Retinal Eye Exam	55.7%	9.1%	26.1% - 87.1%	10 - 4,999	766	155	161
HbA1c testing	86.0%	5.2%	50.0% - 96.2%	10 - 4,999	766	155	161
Nephropathy testing	77.0%	6.8%	51.2% - 93.7%	10 - 4,999	766	155	161
Cardiovascular LDL-testing	84.5%	6.9%	54.5% - 95.8%	10 - 1,941	290	150	150
Colorectal cancer screening	51.6%	9.5%	15.2% - 76.9%	11 - 20,063	3,201	155	156
Medication Management - ACEI	79.5%	7.3%	55.0% - 94.0%	12 - 2,735	477	150	156
Medication Management - ARBs	51.9%	15.0%	25.5% - 93.3%	10 - 2,049	297	151	152
Persistence of beta blocker therapy- Post MI	81.7%	11.3%	62.5% - 100.0%	10 - 30	16	29	98
Use of Spirometry Testing in Assessment/Diagnosis of COPD	26.3%	9.2%	4.8% - 54.2%	11 - 1,345	214	140	150

(1) Average of physician group rates.
(2) Number of groups with at least 2 physicians and 10 patients.
(3) Number of participating MHQP groups.

Table 3e: Minnesota Community Measurement, Cycle 2 Descriptive Statistics

Minnesota Community Measurement (Physician Group Level) Timeframe: 2006							
Measure	**Average Rate (1)**	**Standard Deviation**	**Rate Range**	**Denominator Range**	**Average Denominator**	**Number of Groups Reporting (4)**	**Number of Groups (5)**
Breast Cancer Screening (2)	75.6%	8.2%	34.7% - 97.6%	21 - 25,269	1,510	104	129
Colorectal cancer screening (3)	61.1%	9.5%	39.5% - 84.5%	27 - 541	117	39	39
LDL testing (2)	80.8%	7.6%	57.4% - 93.5%	53 - 12,470	1,406	52	52
HbA1c testing (2)	89.4%	3.8%	81.1% - 98.2%	1 - 14,011	1,544	52	52
Nephropathy testing (2)	77.6%	7.9%	54.5% - 91.7%	1 - 12,076	1,370	52	52
Retinal Eye Exam (2)	65.8%	7.7%	50.5% - 89.1%	49 - 10,739	1,154	52	52
HbA1c control (3)	50.1%	8.4%	24.6% - 71.6%	60 - 1,448	217	52	52
Diabetes LDL cholesterol control	47.1%	9.8%	26.0% - 67.2%	60 - 1,448	217	52	52
Blood pressure control	43.3%	7.5%	23.9% - 58.5%	60 - 1,448	217	52	52
Diabetes- daily aspirin use	69.2%	11.3%	49.8% - 90.4%	47 - 1,297	195	52	52
Diabetes - smoking status	70.0%	10.9%	46.1% - 83.9%	28 - 975	151	52	52
Diabetes Composite-	8.4%	4.9%	0.0% - 19.6%	60 - 1,448	217	52	52

(1) Average of physician group rates.
(2) Administrative rate
(3) Hybrid rate
(4) Number of groups with at least 30 patients in the denominator.
(5) Number of participating MNCM groups.

Table 3f: Wisconsin Collaborative for Healthcare Quality, Cycle 2 Descriptive Statistics

Wisconsin Collaborative for Healthcare Quality (Physician Group Level) (1) Timeframe: Varies by measure							
Measure	**Average Rate (2)**	**Standard Deviation**	**Rate Range**	**Denominator Range**	**Average Denominator**	**Number of Groups Reporting (3)**	**Number of Groups (4)**
HbA1c testing (3)	68.8%	6.3%	59.5% - 86.0%	3,082 - 25,374	7,072	18	18
HbA1c control (3)	48.3%	7.2%	35.1% - 63.1%	3,082 - 25,374	7,072	18	18
LDL testing (3)	83.1%	5.8%	71.8% - 97.3%	3,082 - 25,374	7,072	18	18
Diabetes LDL cholesterol control (3)	51.3%	6.0%	42.0% - 65.4%	2,283 - 25,374	6,937	18	18
Blood pressure control (3, 5)	39.2%	3.6%	36.4% - 44.5%	3,082 - 8,378	6,538	4	18
Nephropathy Testing (3)	73.7%	11.4%	48.0% - 96.5%	3,082 - 25,374	7,072	18	18
Hypertension-Blood Pressure Control (2)	62.2%	3.4%	58.1% - 69.2%	1,991 - 38,406	11,769	18	18
Cardiolvascular - LDL testing (4)	83.5%	4.6%	78.0% - 92.0%	1,023 - 14,287	4,994	6	18
Cardiovascular - LDL control (4)	62.1%	4.0%	58.0% - 69.0%	1,023 - 14,287	4,994	6	18
Pneumococcal vaccination in 65+ yrs pop (4)	33.6%	9.9%	22.5% - 42.0%	7,631 - 21,971	17,347	4	18
Breast Cancer Screening (4)	61.8%	3.8%	56.9% - 69.4%	14,816 - 124,279	35,723	18	18
Colorectal cancer screening (4)	71.1%	4.3%	59.2% - 78.5%	10,844 - 93,356	25,396	18	18

(1) All-payer results using WCHQ measurement methodology.
(2) Average of physician group rates.
(3) Time period July 1, 2005 – June 30, 2006
(4) Time period calendar year 2006
(5) New measure with this reporting cycle

Comparing Medicare Results to Other Payers

The inclusion of the Medicare FFS data has had a varying impact on measure rates. The sites have provided measure results stratified by payer, where possible, to Delmarva. Analyses of these data are ongoing and are not yet complete. These analyses will be completed in time to include the results in the BQI final report.

One pilot, MNCM, has performed extensive analysis of the impact of the Medicare FFS data on its measure rates. The addition of the Medicare FFS data dramatically changed the dynamics of the population on which rates were calculated. Diabetes measure rates for the 2007 MNCM reporting year (2006 Managed Care Organization data only) were based on a population that was approximately 63% commercially insured managed care enrollees, 25% Medicare managed care enrollees, and 13% state public program managed care enrollees. Following the addition of the Medicare FFS data, the commercially insured managed care portion reduced to approximately 30% of the population on which rates were calculated, Medicare managed care made up approximately 12%, state public program enrollees approximately 6%, and Medicare FFS comprised approximately 53%.

Similarly, for the mammography rate for women ages 52-69, the inclusion of Medicare FFS dramatically altered the population mix. Mammography rates for the 2007 MNCM reporting year were based on a population of women that was approximately 82% commercially insured, 12% Medicare managed care insured and 6% state public program managed care insured. Following the addition of the Medicare FFS data, the commercially insured managed care portion reduced to 51% of the population on which rates were calculated, Medicare managed care enrollees made up 7%, state public program enrollees approximately 4%, and Medicare FFS enrollees 38%.

Finally, the addition of Medicare FFS dramatically altered the population mix for Colorectal Cancer screening rates. Colorectal cancer screening rates for the 2007 Minnesota Community Measurement reporting year were based on a population that was approximately 73% commercially insured and 27% Medicare managed care. Following the addition of the Medicare FFS data, the commercially insured managed care portion reduced to approximately 38% of the population on which rates were calculated, Medicare managed care enrollees made up 14%, and Medicare FFS enrollees made up 49% of the population.

Of the four insurance products measured (Commercial Managed Care, Medicare Managed Care, State Public Programs Managed Care and FFS), Medicare FFS patients have the lowest rates for the administrative Diabetes measures (LDL test, HbA1c test, Retinal Eye Exam, Nephropathy testing) and conversely, Medicare Managed Care had the highest rates.

However, for the Diabetes "control" measures and the MNCM Optimal Diabetes Care composite measure, the differences in the rates of the five components of care (LDL <100, HbA1c<7, BP <130/80, no tobacco use, aspirin use) were not as significant. State Public Programs Managed Care had the lowest rate of optimally managed diabetes care, and in four of the five components of this composite measure. The differences between the rates for Medicare FFS and commercially insured managed care were not significant. Of interest is that for this measure, Medicare

Managed Care is the top performing insurance product when calculating the overall composite score and for four of the five components.

For the Cancer Screening measures, MNCM found the same pattern as with the Diabetes Administrative (claims-based calculation as opposed to hybrid) measures. For Breast Cancer screening and Colorectal Cancer screening rates, Medicare FFS rates were the lowest rates and Medicare Managed Care had the highest rates of the four insurance products measured.

During Cycle 2 aggregation, in addition to the analysis of the differences in the population mix and the measure rates, MNCM also analyzed the impact on a clinic's ranking on the five Cycle 1 measures. Using Cycle 1 data, MNCM analyzed if the changes that resulted from the addition of Medicare FFS data had a statistically significant impact on the order of clinic performance on the 5 Cycle 1 measures, using the Wilcoxon Sign-Rank test.

The results from the Cycle One analysis suggest that, while there do exist some changes in the rank order for clinics, and for some clinics large changes in rank order, in general, the Medicare FFS population gets relatively the same care, as shown by most clinics maintaining their relative rank ordering for most measures. It can be suggested, therefore, that for clinics that have a larger proportion of Medicare FFS patients, coupled with ***high*** performance on the measures within that population, there might be an improvement in their overall rank. Conversely, clinics with larger numbers of Medicare FFS patients with ***low*** performance on the measures might experience a decline in rank. The latter was dramatically demonstrated in at least one measure, where the rank order of the highest performing clinic in MNCM reporting was the lowest performing when Medicare FFS data was added. Such a change in rank would carry a strong message to examine health system factors within the clinic that resulted in lower performance among their most elderly patients. MNCM is currently repeating this analysis based on Cycle 2 results. The results of the analysis will be included in the final BQI report.

Analyses to Address Methodological Questions

Defining Physician Groups to Report Quality Measures

Given Medicare FFS administrative data, physician groups are not obvious. We created a workgroup to review the methods of group attribution across the six pilots sites, and to provide insights into which positive characteristics each method provided. In addition, Delmarva performed an analysis of possible group attribution methods using Medicare FFS data only. There are three main goals to our analysis plan. First, identify the methods available for attributing physicians to physician groups using the Medicare FFS data and narrow down the options. Second, describe and compare the methods for attributing physicians to physician groups identified through the first goal. Third, compare the methods identified above against methods created by the pilot communities that rely on provider feedback and/or provider interaction. This report focuses primarily on the first two goals.

Physician Group Workgroup

We started a workgroup focused on learning what each of the six pilot sites are doing to define physician groups. We then wanted to identify the characteristics that make a method for defining physician groups "good" or "bad". We had periodic phone conferences with representatives from each of the six pilot sites. We had each site describe what they were doing to define physician groups and provide a list of characteristics that they look for in a method to define physician groups. We developed a comprehensive list of characteristics that any "good "methods for defining physician groups should possess, asked each pilot site to rate how well the pilots' methods exhibited these characteristics, then combined and summarized the individual pilot ratings to produce an overall rating on each characteristic, for each method. Table 4 shows the results of this ranking on a four point scale. Appendix C summarizes each pilot's method.

Table 4. Rating of each pilot's method of attributing physicians to groups
0=N/A, 1=somewhat applies, 2=Mostly applies, 3=Completely applies

	CHIR	CCHRI	IHIE	MHQP	MNCM	WCHQ – 1	WCHQ - 2	TIN only
Relevant to Patients	1	3	2	3	3	2	2	1
Relevant to Providers	1	3	2	3	3	3	2	2
Can be done with Medicare data	3	0	1	0	1	0	3	3
Can be done with Admin data	3	0	1	0	1	0	3	3
Allows input of Providers	0	3	3	3	3	3	0	0
Obtains buy-in of Providers	0	3	2	3	3	3	1	1
Results in groups small enough	1	3	2	3	3	2	2	1
Can be used for differing payers	1		3	3	3	0	3	1
Easy to separate payers	1		3	3	3	0	3	1

Out of this workgroup there was strong consensus that methods that allow providers to interactively define their own membership are most desirable. Several of the pilots (IHIE, MHQP, MNCM, WCHQ) are doing this. While this method takes extra effort by the provider, the involvement of the provider gains their trust, validates the groups, and makes the results more relevant to them.

The antithesis of working interactively with providers is using only administrative data. Without the relationships that the pilots have in their community it is a much less expensive and easier option. In some cases, using administrative data is the only option. The workgroup thought that the tax ID method available through the administrative data only is preferred over other non-interactive methods, because it does allow the physicians to somewhat self-define their group through billing. Delmarva has the Medicare FFS administrative data and very limited information from the pilots about how their physicians define a group. Given this, we focus this report on comparisons of methods that use administrative data.

Methods

We used the 2006 Medicare administrative data from six states, Arizona, California, Indiana, Massachusetts, Minnesota, and Wisconsin. We identified four methods that allow us to create lists of groups and individual physicians who belong to these groups. We used data from the carrier part B claims files and the physician reference files, which include:

- Master physician identification and eligibility record (MPIER) – This file has one record per physician and physician profile number. A physician will have a different profile number for each practice location.
- Member – This file has one record per physician and physician group pairing. A physician will have multiple records if they are in more than one group.
- Group – This file has one record per group and contains information about the group.

The four methods that we identified are defined in Table 5.

Table 5. Four methods identified by Delmarva to define physician groups using Medicare FFS data only

Name	Description
UPIN to group UPIN6	All individual performing or referring UPINS linked to group UPIN6.
UPIN to group UPIN6+4	All individual performing or referring UPINs linked to group UPIN6 + group UPIN4. Group UPIN4 contains information about the location of the group.
Tax ID member	All physicians in the member file linked by Tax ID to an individual UPIN.
Tax ID claims	All performing UPINs linked with the Tax ID on the same claim.

For all methods except the Tax ID member method, we started with the carrier claims file. Outlines of the data flow for each of the methods are in Figures 1 through 3. For the UPIN methods we started by creating a list of eligible physicians. This was defined as any physician with a valid UPIN (first character alpha with a value of A-M, P, R-Z followed by 5 numbers) who was listed as a performing or referring UPIN on a carrier claim in 2006. We then merged this list of eligible physicians to the MPIER file to limit the list to only those licensed to practice in the state of interest. We then merged the list to the Member file to pair the physicians with groups (both UPIN6 and UPIN6+UPIN4 separately). The Member file contains dates of group membership. We limited the pairs to only those that were documented any time during 2006. We were unsure about excluding groups that had a business location outside of the state of interest, so we created the group definition both ways to compare the results. We were also unsure about whether a group could be considered a group with only one physician in it. In order to make a determination, we created groups without requiring a minimum number and then created groups that excluded any group that had only one member for comparison.

Figure 1. UPIN methods data flow chart

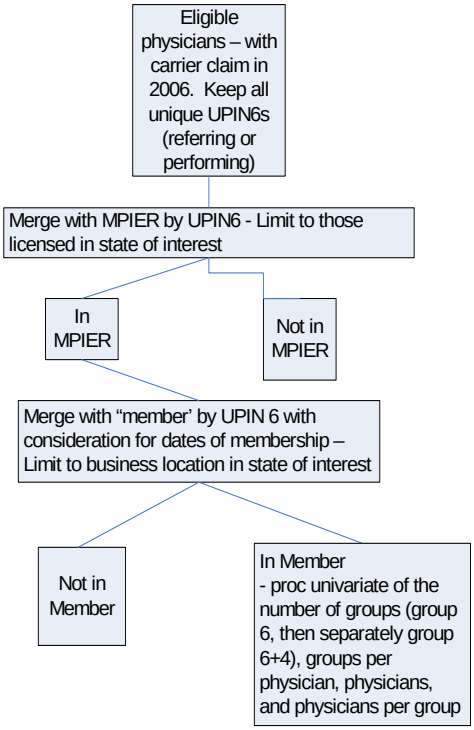
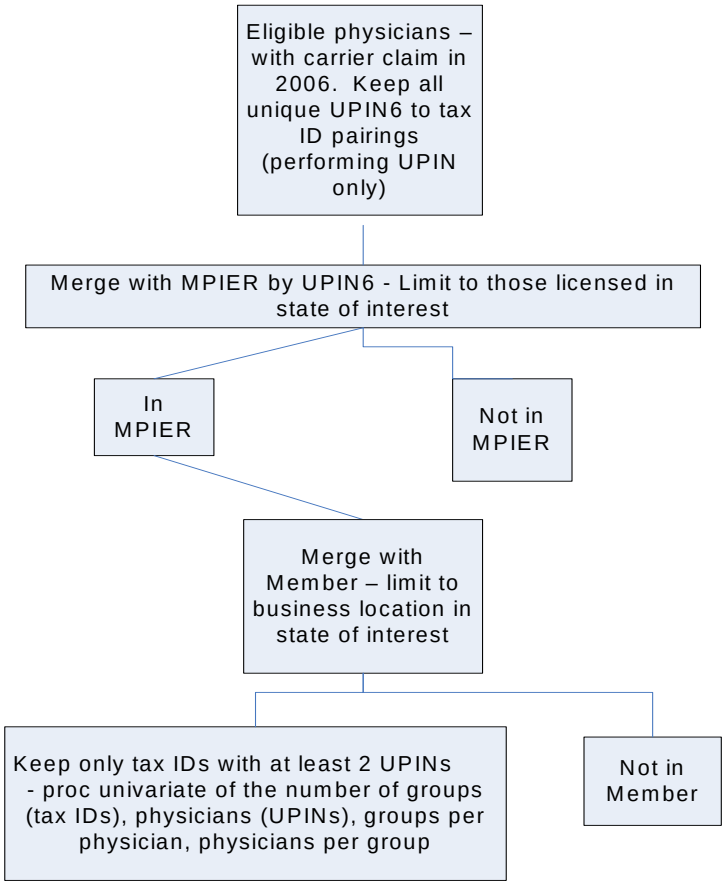
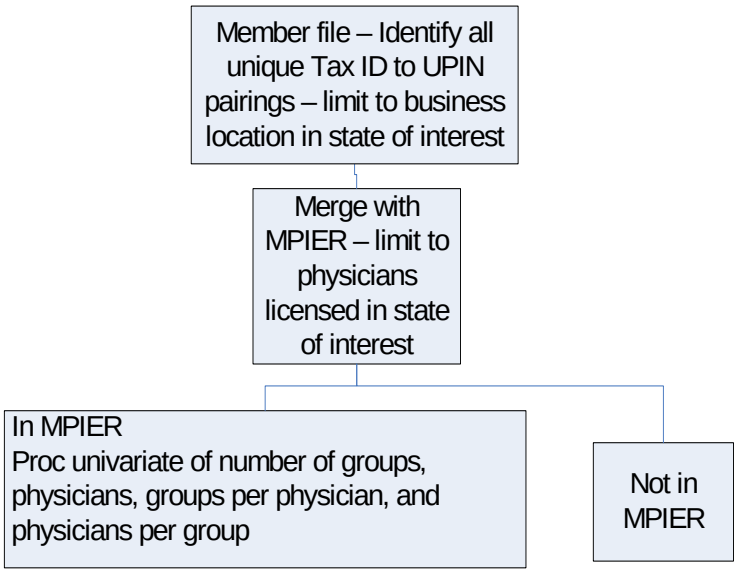


Figure 2. Tax ID method data flow chart



The tax ID member method was the simplest because it did not require the use of claims data. We started with the Member file and identified tax ID to UPINs pairings as physician to group pairings. Eligible physicians were defined as any physician with a valid UPIN (first character alpha with a value of A-M, P, R-Z followed by 5 numbers). We limited the pairs to only those that were documented any time during 2006. We then merged this list of pairs to the MPIER file to limit the list to only those licensed to practice in the state of interest. We were again unsure about excluding groups that had a business location outside of the state of interest, so we created the group definition both ways to compare the results. We were also unsure about whether a group could be considered a group with only one physician in it. In order to make a determination, we created groups without requiring a minimum number and then created groups that excluded any group that had only one member for comparison.

Figure 3. Tax ID member method



For each of our four methods, we calculated a set of descriptive statistics to compare the number of physicians required to define a group and the business location. This created four options for each of the four methods:

1. Any business location, no minimum number of physicians per group
2. Business location restricted to state of interest, no minimum number of physicians per group
3. Any business location, number of physicians per group greater than 1
4. Business location restricted to state of interest, number of physicians per group greater than 1

The descriptive statistics included the total number of groups, the number of physician to group pairings (the number of physicians allowing them to be double counted if they are in multiple groups), the number of unique physicians with at least one group, the number of beneficiaries attributed to at least one group (allowing for double counting of beneficiaries attributed to multiple groups) , the number of unique beneficiaries attributed to at least one group, the mean number of physicians per group, and the mean number of beneficiaries per group. We used the one-touch rule to attribute beneficiaries to groups for the purpose of these descriptive statistics. The descriptive statistics represent the size and utilization of the groups. For each of the four methods, we decided which option was most appropriate then compared the results of that option to the option selected for each of the other methods.

Results

We have generated results for Indiana and Arizona and only these states' data are included in this report. Results from the other four states will be included in the BQI final report. Appendix D shows the results of our analyses to determine whether to limit physician groups to only those with a business location in the state of interest and whether a group can be a group with only one member in it. The tables are arranged from the least restrictive definition of 'group' to the most restrictive definition, as follows: (1) one ore more physicians per group and any business location, (2) one or more physicians per group and business location limited to only the relevant State, (3)

two or more physicians per group and any business location, and (4) two or more physicians per group and business location limited to only the relevant State. The indecision about the business location was because we did not know if there were many businesses that had a corporate headquarters outside of the state where the physicians practiced.

The tables in Appendix D show the descriptive statistics for each of these four options, plus a trend arrow to indicate the direction the numbers go as the options get more restrictive. As the options get more restrictive, the number of groups, number of physician to group pairings, number of unique physicians, number of beneficiaries attributed to pairings, and number of unique beneficiaries all decrease while the mean number of physicians per group and mean number of beneficiaries per group increase. The comparison indicates that the number of groups decreases faster than the number of physicians and beneficiaries.

The small decrease in the number of beneficiaries when limiting to only the business locations in-state indicates that groups with business locations out of state did not account for many claims. There was a dramatic drop in the number of groups when limiting to in-state business locations. The data request was by the state of the beneficiary, which means that there would be claims for groups who happen to see a beneficiary when they were out of state. These groups would not account for many beneficiaries, so we decided to limit to only in-state business locations for subsequent analyses.

Similarly, we were uncertain about whether to include groups with only one physician because we did not know if this constituted a group. While there are many individual physicians practicing by themselves, we did not want to assume that a "solo" physician always constitutes a "group". Again we observed a dramatic decrease in the number of groups, but only a minor decrease in the number of beneficiaries. We heard anecdotal evidence that groups may systematically bill to only a single UPIN, thereby making it appear that there is only one physician in a group. We finally decided that a group need only have one physician in it to be considered a group for both UPIN to group UPIN methods because the group UPIN is inherently a group. For the tax ID methods, we decided to require at least two physicians to make a group because the tax ID is not designed to capture group identities. Subsequent analyses limit the methods to these selected options.

Tables 6a and 6b shows the final four methods' descriptive statistics, for Arizona and Indiana.

Table 6a. Comparison between four physician grouping methods, Arizona

Summary Statistics	Tax ID Method ≥2 phys./group		Group UPIN6 Methods ≥ 1 physician/group	
	UPIN6 By Tax ID Method (Claims)	UPIN6 By TAX ID method (Member file)	UPIN6 By Group UPIN6	UPIN6 By (Group UPIN6 +UPIN4)
# of physician groups	1,632	1,662	2,732	3,268
# of unique physicians in the group	10,458	10,349	9,537	9,537
# of physicians Attributed	12,967	16,193	14,494	16,349
Mean # physicians per group (SD)	7.95 (24.44)	9.74 (27.74)	5.30 (18.19)	5.00 (14.01)
Min #of physicians per group	2	2	1	1
Max #of physicians per group	604	494	380	380
Mode	2	2	1	1
Median	3	3	2	2
Distribution of beneficiaries				
# of beneficiaries attrib. to group	1,790,379	1,010,379	2,193,530	2,261,242
# of unique beneficiaries by group	469,654	396,903	487,413	487,413
Mean # beneficiaries per group (SD)	1,097.05 (2,588.17)	607.93 (2,058.93)	802.90 (3,025.42)	691.93 (2,623.32)
MIN # of beneficiaries per group	1	1	1	1
Max # of beneficiaries per group	46,059	44,135	99,335	99,335
Mode # of beneficiaries per group	2	1	1	1
Median # of beneficiaries per group	376	417	166	151

Table 6b. Comparison between four physician grouping methods, Indiana

Summary Statistics	Tax ID Method ≥2 phys./group		Group UPIN6 Methods ≥ 1 physician/group	
	UPIN6 By Tax ID Method (Claims)	UPIN6 By TAX ID method (Member file)	UPIN6 By Group UPIN6	UPIN6 By (Group UPIN6 +UPIN4)
# of physician groups	1,565	2,040	2,909	4,432
# of unique physician in the group	10,454	14,687	13,028	13,028
# of physician Attributed	12,837	22,381	19,704	23,958
Mean # physician per group (SD)	8.20 (14.25)	10.97 (21.03)	6.77 (15.62)	5.40 (9.85)
Min #of physician per group	2	2	1	1
Max #of physician per group	198	255	233	212
Mode	2	2	1	1
Median	4	4	2	2
Distribution of beneficiaries				
# of beneficiaries attrib. to group	2,708,855	2,323,403	3,559,114	3,804,221
# of unique beneficiaries by group	746,580	718,511	774,924	774,924
Mean # beneficiaries per group (SD)	1,730.90 (3,769.51)	1,138.92 (3,114.35)	1,223.48 (3,261.08)	858.35 (2,277.92)
MIN # of beneficiaries per group	1	1	1	1
Max # of beneficiaries per group	63,447	63,087	70,477	45,136
Mode # of beneficiaries per group	3	1	1	1
Median # of beneficiaries per group	612	193	329	185

Since these tables were generated we corrected a minor programming error that calculated the number of beneficiaries attributed to groups. We have rerun the program to calculate the numbers for the tables in Appendix D for the two states reported herein. We have also run the SAS program for the remaining four states. We are in

the process of populating these tables and generating the figures. The updated results will be included in the BQI final report.

We have received descriptive statistics for groups from the Wisconsin, Minnesota, and Massachusetts pilot sites. For each of these pilot sites we will add a column to the main results table that shows the pilots' descriptive statistics beside the data we generate using the Medicare FFS data. These results will be included in the BQI final report. For now the results are summarized in Table 7.

Table 7. Pilot Descriptive Statistics (Physician groups defined independently by each pilot)

Characteristic	WI 2006	WI PCPs only 2006	MN	MA
Number of physician groups	650 (1)		148	155
Number of physicians attributed	15,740 (2)	8,558 (2)	254	3802
Mean # physicians per group (SD)	26 (2)	15 (2)	1.72 (3.20)	23.9 (30.6)
Min # of physicians per group	1 (2)	1 (2)	1	2
Max # of physicians per group	1,360 (2)	758 (2)	23	223
	2003-2006			
Mean # of claims per group (SD)	3549 (78,714) (3)			
Min # of claims per group	1 (3)			
Max # of claims per group	7,596,160 (3)			

(1) Groups with Wisconsin billing zip codes and significant border groups only
(2) Only Wisconsin physicians and primary care providers (PCPs) with Wisconsin billing zip codes
(3) Statistics are for claim LINES and all groups across all years

Conclusions

Preliminary comparisons between our physician group definitions indicates that there are large differences between how we are defining groups and how the pilots are defining groups (see Table 7 above). There are many fewer groups identified by the pilot sites using their respective methods than the number identified by Delmarva using only Medicare FFS data. It is not clear if this indicates that the pilot only has data for the groups that work with them, while we have data for all groups who bill Medicare, or if the definition of a physician group is vastly different. This will need to be explored further.

Attribution of Beneficiaries to Physicians Preliminary Analysis Results

One of the first challenges faced when attempting to use administrative claims data from a fee for service environment to measure quality performance for physicians or physician groups is that there is no singular method for assigning the responsibility for a beneficiary to a physician or group of physicians. Given the Medicare data, there are many options for assigning responsibility for a given beneficiary to a physician and no clearly correct method. There are quantitative and qualitative differences between the choices. The goal of these analyses is to elucidate the quantitative differences between the attribution methods.

Methods

We conducted analyses using the Medicare FFS administrative data for six states (Arizona, California, Indiana, Massachusetts, Minnesota, Wisconsin) for the years 2003-2006. Five quality measures are calculated using Medicare FFS data for six states for 2006, using the prior years' data as look back periods. We used 2007 HEDIS® specifications to calculate the five measures – Breast cancer screening, Colorectal cancer screening, HbA1c testing, LDL testing, and nephropathy testing. These five measures were selected because they are consistent across all six of the pilot sites.

We identified five attribution algorithms – one-touch, two-touch, 30%, 50%, and maximum frequency. Table 8 lists these rules along with a brief description. There are many other rules. We chose these five because there are two fairly liberal methods (the one-touch and two-touch rules assign multiple physicians per beneficiary), two fairly conservative rules (the 50% rule and maximum frequency rules only attribute one physician per beneficiary), and one middle-ground approach (30% rule).

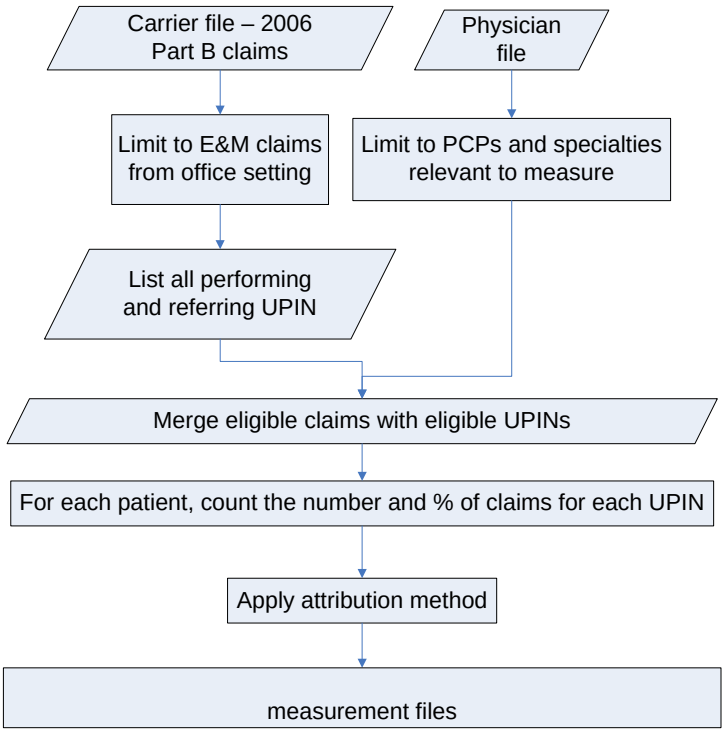
Table 8. Attribution rules tested

Rule	Description
One-touch rule	Every physician who touched the beneficiary is attributed to the beneficiary.
Two-touch rule	Every physician who touched the beneficiary at least 2 times is attributed to the beneficiary.
30% rule	Any physician who performed at least 30% of the claims for a beneficiary is attributed to them.
50% rule	Any physician who performed at least 50% of the claims for a beneficiary is attributed to them. If tied, only 1 physician is selected randomly.
Maximum frequency	The physician with the highest number of claims for a beneficiary is attributed to them. If tied, only 1 physician is selected randomly.

The data flow for calculating attribution methods is described in Figure 4. We started with the 2006 carrier file for each state and limited it to only E&M claims from an office setting (see Appendix E for codes to identify E&M claims and codes). We then create a list of eligible UPINs as physicians who were either performing or referring UPINs on an eligible claim and who had a specialty related to the measure (see Appendix F for list of specialties per measure). This step of limiting UPINs to those with relevant specialties required us to rerun the attribution algorithm 3 times – once each for the breast cancer screening, colorectal screening, and diabetes measures. We then

merged the list of eligible UPINs to the list of eligible claims and apply each of the five attribution algorithms in turn. At that point, we merged the resulting beneficiary – physician pairing list to a beneficiary level file that has the numerator and denominator status for the measure.

Figure 4. Data flow for attribution methods



For each state and measure, we calculated the numerator, denominator, overall statewide rate, and number of eligible physicians for the rate. To compare the five attribution methods, we calculated the mean of the physician rates, standard deviation, percent of eligible physicians with a rate, and the percent of eligible beneficiaries with a physician separately for each attribution rule. To evaluate the impact of a minimum denominator, we also calculated these measures for each attribution rule after excluding physicians with a denominator of 10, 20, 30 and 50.

Preliminary results

Appendix G allows us to compare results of 2006 measures using the 1-touch rule, 2-touch, the 30% rule, the 50% rule, and maximum frequency rule with no minimum denominator. The table in this appendix presents preliminary analysis results, including the overall statewide rate for each of the six measures and for five of the six sites. The table also includes the mean individual physician rate, the standard deviation of the physician rate, the percent of eligible physicians with at least one beneficiary attributed to them, and the percent of eligible beneficiaries attributed to at least one physician for each of the three attribution methods used, for the analyses that had been completed at the time of the submission of this report.

Updates not reflected in these results

Since the results in Appendix G were created, we made several corrections to the SAS programs that calculated the results. First, we updated the measure specifications to match HEDIS® 2007. For breast cancer screening, this involved lowering the minimum denominator to age 42 for denominator criteria, adding a HCPCS code and removing a revenue center code for identification of the numerator. We did not make any changes to the colorectal screening specifications. For the diabetes care measures, we added and deleted several diagnosis, procedure, and revenue codes used to identify diabetics and to identify the numerators of all three indicators. Additionally, we changed the algorithm for the 50% rule to add the tie breaker logic. For the results that appear in this report, the 50% rule allows multiple physicians to be attributed to the same beneficiary. We added a tie breaker to select only one physician randomly for attribution to the beneficiary when two physicians have 50% of their claims.

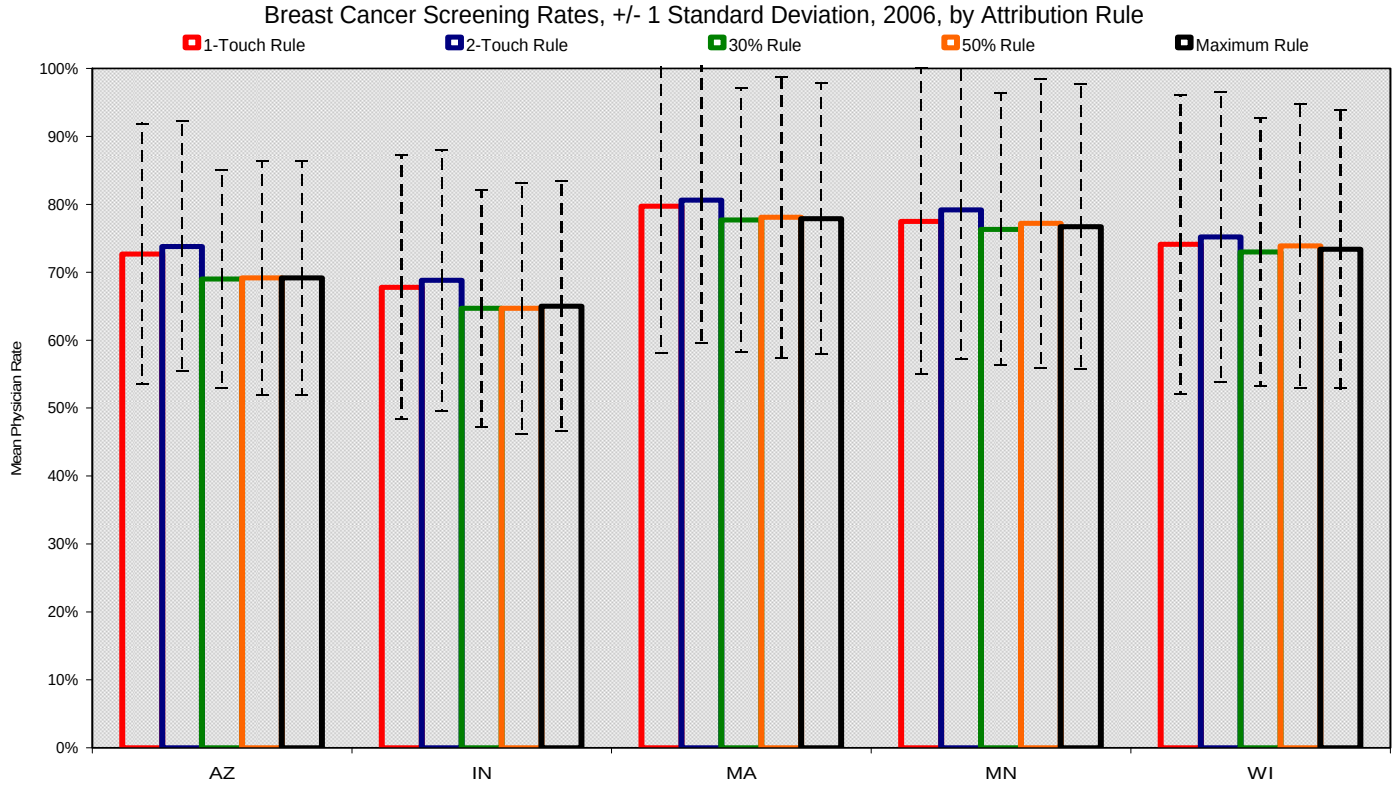
We have made all of these edits to the SAS code. Additionally, we have rerun the results for breast cancer screening and colorectal screening for all states except California. We will repopulate the tables in Appendix G when all of the updated results are available, for the final BQI report.

Discussion

The mean physician rates for all five attribution methods are higher than the overall rate for all measures in all states reported. This indicates that there is some systematic bias in the selection of beneficiaries who can be attributed to physicians. It makes sense that beneficiaries who have had an E&M visit with a physician from a relevant specialty (and are therefore attributed to at least one physician) are more likely to have the screenings than beneficiaries who have not had an E&M visit with a physician from a relevant specialty (but who are included in the overall rate).

We can also use the mean physician rates to compare the five attribution methods. However the direction of the results varies by measure. For breast cancer screening rates, the 2-touch rule produced the highest mean physician rate followed closely by the 1-touch rule. The 30%, 50%, and maximum frequency rule rates were grouped closely together and were lower than the more liberal attribution methods. This trend also occurred in the colorectal screening rate (see Appendix H, page H-1). The trend also appeared to occur for the three diabetes measures, however the differences between the scores was much smaller in these measures (See pages H-2 through H-5). This may be because, for diabetes care, beneficiaries who see one physician regularly tend to have these screenings more often than for other measures. Further investigation to explain these observations is warranted.

Figure 5. Breast cancer screening rates by attribution method for five states



The percentage of eligible physicians with a rate also varied by measure. Eligible physicians include any physician with at least one E&M claim in an office setting during the time period and who has a specialty relevant to the measure. The highest percentages of eligible physicians with a rate occurred for the 1-touch rule. Table 9a shows the percentages of physicians with a rate for each attribution rule, along with the relative change in the percent of eligible physicians, when each attribution rule is compared against the 1-touch rule. The relative change from 1-touch rule is calculated by dividing the percentage of eligible physicians with a rate for the attribution method in that row by the percentage for the 1-touch rule. For all measures in all states, the trend in the percentage of physicians with a rate was consistently from highest to lowest as – 1-touch rule, 2-touch rule, 30%, maximum frequency, and 50% rule. It will be interesting to see if this trend continues after our change in the algorithm for the 50% rule and after all measures and states are run. At the time of the submission of the initial Cycle 2 report, not all measures, states, and attribution analyses runs had been completed.

Table 9a. Percentage of eligible physicians with a rate by attribution method

Measure/State	Attribution Method	% of eligible physicians w/a rate	Relative change from 1-touch rule
Breast cancer screening - AZ	1-touch rule	70.4%	Reference
	2-touch rule	62.4%	88.6%
	30% rule	55.2%	78.5%
	50% rule	47.1%	67.0%
	Maximum frequency rule	51.8%	73.6%
Breast cancer screening - IN	1-touch rule	62.3%	Reference
	2-touch rule	55.8%	89.4%
	30% rule	50.8%	81.5%
	50% rule	43.4%	69.6%
	Maximum frequency rule	47.7%	76.5%
Breast cancer screening - MA	1-touch rule	77.5%	Reference
	2-touch rule	67.4%	87.0%
	30% rule	63.0%	81.4%
	50% rule	53.9%	69.5%
	Maximum frequency rule	58.7%	75.8%
Breast cancer screening - MN	1-touch rule	76.9%	Reference
	2-touch rule	65.9%	85.7%
	30% rule	62.4%	81.1%
	50% rule	51.7%	67.2%
	Maximum frequency rule	58.5%	76.0%
Breast cancer screening - WI	1-touch rule	77.2%	Reference
	2-touch rule	68.5%	88.8%
	30% rule	67.1%	86.9%
	50% rule	58.8%	76.3%
	Maximum frequency rule	63.3%	82.0%
Colorectal cancer screening - IN	1-touch rule	96.5%	Reference
	2-touch rule	86.3%	89.5%
	30% rule	80.0%	82.9%
	50% rule	69.6%	72.2%
	Maximum frequency rule	75.1%	77.8%
AZ – All 4 measures: LDL testing HbA1c testing Retinal eye exam Nephropathy testing	1-touch rule	73.3%	Reference
	2-touch rule	64.6%	88.1%
	30% rule	55.0%	75.1%
	50% rule	46.2%	63.1%
	Maximum frequency rule	50.9%	69.5%

The percentage of eligible beneficiaries with a rate also varied by measure and attribution method; however the differences between attribution methods were not as drastic for this measure. Eligible beneficiaries include any beneficiary included in the statewide denominator for the measure. Beneficiaries are considered to have a rate if they are attributed to at least one physician who has a rate. The highest percentages of eligible beneficiaries with a rate occurred for the 1-touch and maximum frequency rules. Percentages were identical for these two measures. Table 9b shows the percentages of beneficiaries with a rate along with the relative change from the observed percentages 1-touch percentage. The relative change from 1-touch rule is calculated by dividing the percentage of eligible beneficiaries with a rate for the attribution method in that row by the percentage for the 1-touch rule.

For all measures in all states, the trend in the percentage of beneficiaries with a rate was consistently from highest to lowest as – 1-touch rule and maximum frequency rule tied, 30%, 2-touch, then 50% rule. With the exception of the 50% rule, the change in percentage of eligible beneficiaries between the varying attribution methods is very small. It will be interesting to see if this trend continues after all measures and states are run.

Table 9b. Percentage of eligible beneficiaries with a rate by attribution method

Measure/State	Attribution Method	% of eligible beneficiaries with a rate	Relative change from 1-touch rule
Breast cancer screening - AZ	1-touch rule	84.3%	Reference
	2-touch rule	78.2%	92.8%
	30% rule	83.3%	98.8%
	50% rule	70.4%	83.6%
	Maximum frequency rule	84.3%	100%
Breast cancer screening - IN	1-touch rule	89.9%	Reference
	2-touch rule	84.2%	93.6%
	30% rule	89.0%	99.0%
	50% rule	76.6%	85.2%
	Maximum frequency rule	89.9%	100%
Breast cancer screening - MA	1-touch rule	82.4%	Reference
	2-touch rule	75.2%	91.3%
	30% rule	81.8%	99.3%
	50% rule	71.2%	86.5%
	Maximum frequency rule	82.4%	100%
Breast cancer screening - MN	1-touch rule	82.5%	Reference
	2-touch rule	72.4%	87.8%
	30% rule	80.9%	98.0%
	50% rule	66.1%	80.1%
	Maximum frequency rule	82.5%	100%
Breast cancer screening - WI	1-touch rule	84.2%	Reference
	2-touch rule	74.9%	89.0%
	30% rule	83.4%	99.0%
	50% rule	71.5%	84.9%
	Maximum frequency rule	84.2%	100%
Colorectal cancer screening - IN	1-touch rule	88.5%	Reference
	2-touch rule	83.0%	93.9%
	30% rule	87.8%	99.3%
	50% rule	77.2%	87.2%
	Maximum frequency rule	88.5%	100%
AZ – All 4 measures: LDL testing HbA1c testing Retinal eye exam Nephropathy testing	1-touch rule	85.3%	Reference
	2-touch rule	81.0%	94.9%
	30% rule	84.5%	99.0%
	50% rule	73.7%	86.3%
	Maximum frequency rule	85.3%	100%

Effect of minimum denominator

Physicians with a denominator less than ten tend to have rates that are erratic and unreliable. For this reason, we duplicated the statistics reported above after excluding physicians with a denominator less than 10. Ten was

selected as an arbitrary minimum denominator size. Twenty, thirty, and fifty are often used by statisticians as the minimum denominator (N) required to be able to assume a normal distribution. The measures are also calculated using these minimum denominator sizes (see Appendix I). However, the purpose of this table is to show the effect of applying a crude and liberal minimum denominator size of 10.

Many of the trends noted above continue. For the comparison between mean physician rates, the trend is the same as was noted for no minimum denominator size. The 1-touch and 2-touch rates had the highest rates, followed by a tight grouping of the other three attribution methods. This difference was again less pronounced in the diabetes care measure. The trend in percentage of eligible physicians with a rate was also identical to the one noted above for no minimum denominator size. The percentage of eligible physicians varied consistently between the methods from highest to lowest: 1-touch rule, 2-touch rule, 30%, maximum frequency, and 50% rule.

The trend in percentage of eligible beneficiaries with a rate changed when a minimum denominator of 10 was applied. The 1-touch rule continued to have the highest percentage, but the maximum frequency rule no longer had the same high percentage. For breast cancer screening, the attribution methods ranked in order of highest to lowest: 1-touch, 30%, followed very closely by 2-touch, maximum frequency, and finally the 50% rule. For colorectal screening, the maximum frequency rule scored slightly higher than the 2-touch rule. For the diabetes measures, the 2-touch rule scored higher than the 30% rule.

Despite these minor changes in the trends in the comparison of the attribution methods, the impact of the minimum denominator size was large. The mean physician rates tended to decrease after excluding physicians with less than ten beneficiaries. The nephropathy rates decreased the most. The standard deviations decreased consistently after excluding physicians with less than ten beneficiaries. As would be expected the percent of physicians with at least ten beneficiaries was much smaller than the percent of physicians with at least one beneficiary. This indicates that even with a very crude and liberal cutoff for a minimum denominator size, many physicians will not have a reportable rate. Likewise, fewer beneficiaries are represented in the rates from physicians when the less-than-10 cutoff criterion is added.

Next steps

First, we will fill out the remainder of the Appendix G. Next, we will re-evaluate all of the trends noted in this report. The goal of this next evaluation will be to identify two or three attribution methods that are representative of a "type" of attribution method. For example, the results for the 1-touch and 2-touch rules appear very similar. For subsequent data analyses, we can limit to only one of the algorithms of each type. Third, we will evaluate the impact of adding an additional 2-touch rule to any of the attribution methods where this is an option (Under the current version of these rules, a beneficiary may be attributed to a physician for only a single claim). To limit the quantity of analyses and results, we will again determine which two or three methods are best to continue using. Fourth, we will evaluate the impact of varying physician specialty definitions. We will alternatively use a more strict limitation than what we previously used to accept only primary care physicians and then use no limitation for

specialty of physicians. Finally, we will calculate the measures at the physician group level using the physician grouping method identified from that set of analyses. Calculating measures at the physician group level still requires the same attribution of beneficiaries to providers as we use here. We have the additional option of attributing beneficiaries to individual physicians, then rolling the individual physicians into groups, or we may attribute beneficiaries directly to groups. We will use two or three different attribution methods selected from the above analyses to get scores at the physician group level. We will then have a quantitative evaluation of most of the attribution method choices available in future BQI-like activities.

Reliability and Minimum Denominator Size Preliminary Results

Reliability

Small denominator sizes have always been a problem when reporting quality measurement results. This problem is intensified when reporting quality measurement results at the individual physician or even physician group level because only a few physicians have denominators large enough to meet crude minimum cut off points. There are four questions that we hope to address with our analyses on reliability:

1. Are physician (or physician group) level measures calculated based on a small denominator size predictive of a physicians true score?
2. Is the distribution of scores such that we can use scores to differentiate between high and low performers?
3. What happens if we impose a minimum denominator size?
4. How confident are we in a given practice's score?

Methods

We conducted analyses using the Medicare FFS administrative data for six states (Arizona, California, Indiana, Massachusetts, Minnesota, Wisconsin) for the years 2003-2006. Five quality measures are calculated using Medicare FFS administrative data in six states for the year 2006, using the prior years' data as look back periods. We used 2007 HEDIS® specifications to calculate the five measures – Breast cancer screening, Colorectal cancer screening, HbA1c testing, LDL testing, and nephropathy testing. These five measures were selected because they are consistent across all six of the pilot sites.

Reliability defined

Unfortunately, there are no well known statistics that quantify the answers to these questions given our circumstances. Question #1 cannot be answered directly without repeated measures of scores on the same physician. While we do have several years of data and have calculated some of the same measures for Cycle 1, the specifications of the measures themselves have changed. While we are unable to directly address Question #1, the reliability scores that we do obtain can be interpreted as estimates of test-retest reliability. Question #2 we can

address with our approaches to calculating reliability. We are calculating reliability as the variance between physicians divided by the sum of the variance between physicians and the variance within physicians:

Formula 1: Reliability

$$\text{Reliability} = \sigma^2_{\text{provider to provider}} / (\sigma^2_{\text{provider to provider}} + \sigma^2_{\text{error}}/n)$$

High reliability scores can suggest that it is easy to tell providers apart by their performance measures, while low scores can suggest that it is difficult to tell providers apart by their performance measures. Denominator size and the overall score on the performance measures have an impact on reliability scores. Reliability scores increase as the denominator size increases. Reliability scores decrease as the overall performance measure approaches 50%, with higher reliability scores as the performance measure approaches 0% or 100%. Because reliability scores are dependent on the tightness of the distribution of the scores, a measure that has little variation between physicians can have high reliability. In this case, the high reliability scores will make it appear as if it is easy to use that measure to differentiate between physicians, but in fact it will be very difficult to differentiate between the physicians. For this report, we calculate reliability scores to correspond with varying denominator sizes. While reliability scores always increase as a function of the denominator size, it is important to consider that there are other factors involved and that even measures with high reliabilities may not necessarily indicate that it is good measure to report. A reliability of 0.70 is generally considered "good" [10] and can be interpreted to mean that 70% of the variability between beneficiaries getting the screening is due to physician to physician variance. To directly address question #3, we calculated descriptive statistics and reliability scores at various cutoff points.

Question #4 is tricky. On one hand, it is quite simple to calculate a confidence interval around a proportion. On the other hand, it is probably not appropriate to use confidence intervals in this situation. Confidence intervals are used when point estimates are calculated from a sample of the data. In our Medicare FFS data, we may not always have all of the patients for a given physician, but we have close to all of the data for Medicare beneficiaries. If the performance measure is described as the "Medicare FFS performance measure", it can easily pass for the entire population and not a point estimate taken from a sample. Therefore there is no sampling error around our performance scores and it is not appropriate to calculate confidence intervals around the performance. For this reason, we have not attempted to address Question #4.

Analysis plan

For all five measures, all six pilots, and all five attribution methods we calculated reliability scores for different minimum denominator cutoff points (10, 20, 30, and 50). We chose these cutoff points based on previous work performed by NCQA. To estimate reliability, we used a powerful type of statistical model that can be used for analysis of correlated data - a mixed model. The model used a binomial (yes or no) variable that indicated the beneficiaries' numerator status as the outcome. The only predictor variable in the model is physician. The model is hierarchical because the variability must be calculated at the individual beneficiary level within physician. The

[10] Shrout, P.E. and Fleiss, J.L. 1979. Intraclass correlations: Uses in assessing rater reliability. Psychological bulletin, 2, 420-428.

proc NLMIXED procedure in SAS is used to accomplish this. Because the outcome measure is binomial and we used a hierarchical model, the parameters of the model need to be converted to approximate those of a normal distribution. We used the beta link function to convert the model parameters to approximate a normal distribution. We use the term "beta-binomial" throughout this report to refer to the results of this model. The beta-binomial model produced the estimates of between and within variability. We ran a separate model for each state, measure, attribution method, and minimum denominator. We plugged these estimates of variance into the Spearman-Brown prophesy formula to obtain reliability scores. The Spearman-Brown prophesy formula is identical to our definition of reliability in Formula 1. The "n" in the formula can easily be adjusted to produce differing reliability scores. For each set of model results, we plugged in the minimum denominator size used for that model run (beta binomial at the minimum) and the average denominator size of the physicians' included in that model run beta binomial at the average). To be conservative, we put the most stock in reliability scores calculated at the minimum.

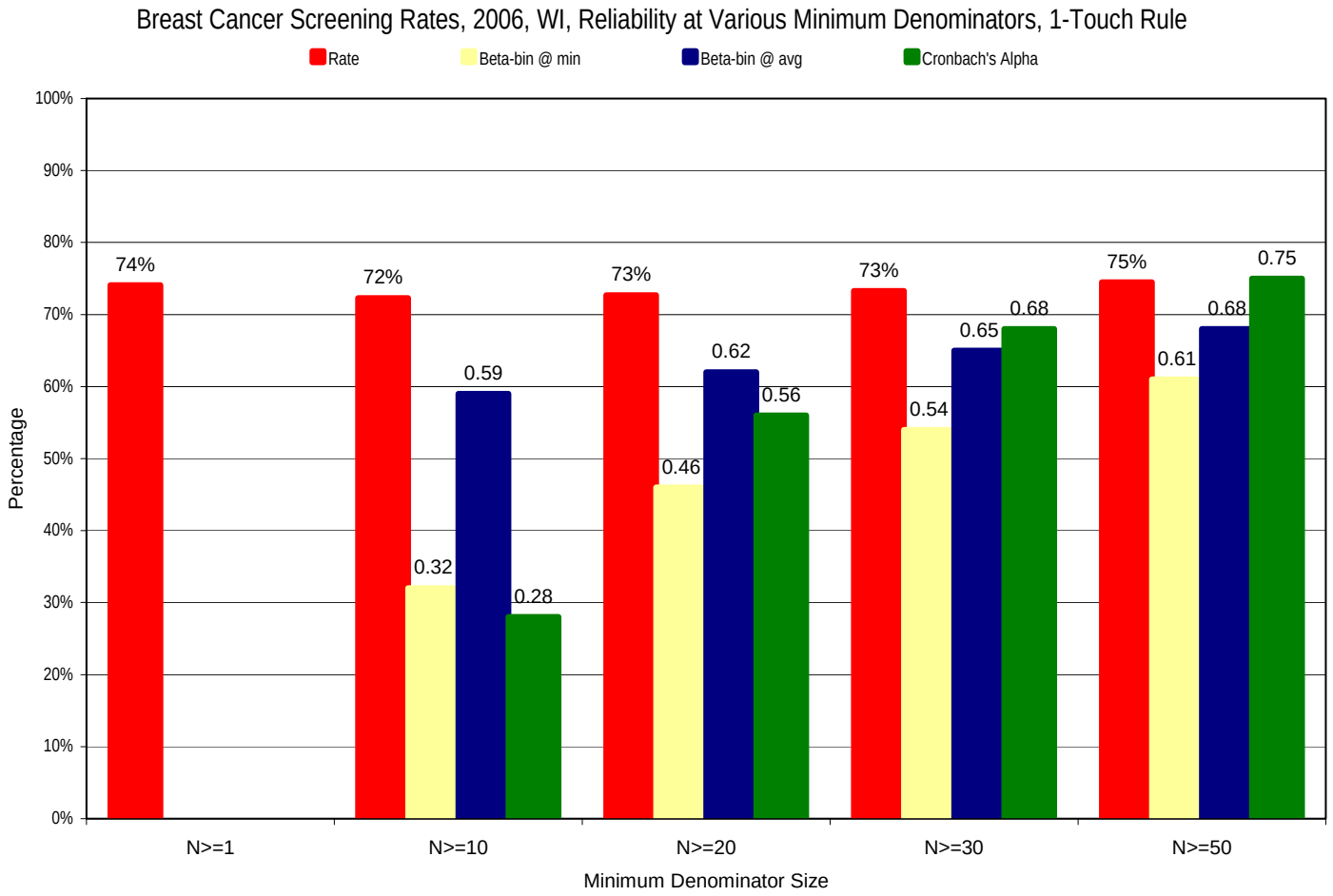
As an alternative to the beta-binomial approach to reliability, we also calculated Cronbach's alpha reliability. Cronbach's alpha measures internal consistency reliability. It is commonly used to evaluate whether items within a scale are correlated with one another. While the Cronbach's alpha is used out of its normal context in this instance, it provides a useful alternative to the beta-binomial method. The primary disadvantage to the Cronbach's alpha method is that it can only be run when all physicians have the same denominator size. To achieve this, we make a separate analysis for each cutoff point. We limit the data to only physicians who meet the minimum denominator criterion. For those who have more than the exact cutoff point, we randomly select patients to include only the exact number of patients specified by the cutoff point. This makes the Cronbach's alpha somewhat comparable to the beta-binomial method at the minimum denominator size. The advantage of the Cronbach's alpha is that it can be calculated even when the beta binomial model fails to converge.

Results

Results can be found in Appendices H and I. Figure 6 shows an example of the results from Appendix H (pages H-23 through H-47). The figure shows three different reliability scores for various minimum denominator sizes for breast cancer screening rates in WI calculated using the one-touch rule. We trust the beta binomial at the minimum method more than the other two. Even with a minimum denominator size of 50, the beta binomial at the minimum score is only 0.61 and it only exceeds the 0.70 cutoff point for the Cronbach's Alpha. Reliability scores rise for attribution methods that are more restrictive. For the 30%, 50%, and maximum frequency rules all three reliability methods achieve scores of at least 0.70 with a minimum denominator size of 50. There is some variation between the states in reliability scores for breast cancer screening. AZ had the highest reliability scores achieving beta binomial at the minimum scores of 0.70 for the one-touch, two-touch, and 30% rule at a denominator of 50 and for the 50% and maximum frequency rule at a denominator of 50 and 30. MA had the lowest reliability scores with no beta binomial at the minimum scores reaching 0.70. The results for MN breast cancer screening reveal a weakness in the beta binomial method. For the 30% rule at a minimum denominator of 50, no reliability scores were calculated because the beta binomial model would not converge. The biggest reason for this break down in the method is the small number physicians who had a score with a minimum denominator of 50. The results for IN

were very similar to those from WI – The beta binomial at the minimum score reached 0.70 for the two-touch, 30%, 50%, and maximum frequency rules with a denominator of 50.

Figure 6. Reliability scores for breast cancer screening rates in WI at various minimum denominator sizes, 1-touch rule



For colorectal screening in Indiana, the reliability results are similar to breast cancer screening reliability results. The beta binomial at the minimum reliability score reaches 0.70 at a minimum denominator of 50 for all attribution methods. However, there is a large difference between the beta binomial scores at the minimum and beta binomial scores at the average for this measure that was not seen for breast cancer screening. For the three diabetes measures, we only have reliability results for AZ. For HbA1c the beta binomial at the minimum reliability reached 0.70 at a denominator of 50 for the 30%, 50%, and maximum frequency rules. Reliability scores were slightly higher for the LDL measure. The beta binomial at the minimum reached 0.70 at a denominator of 50 for all five attribution methods and reached 0.70 at a denominator of 30 for the 30%, 50% rule, and maximum frequency rules. Reliability scores were highest for the nephrology measure. The beta binomial at the minimum reached 0.70 at a denominator of 50 and 30 for all five attribution methods and reached 0.70 at a denominator of 20 for the 30%, 50% rule, and maximum frequency rules.

Updates not reflected in the results presented and discussed above

Since the results in Appendices H and I were created, we made several corrections to the SAS programs that calculated the results. First, we updated the measure specifications to match HEDIS® 2007. For breast cancer screening, this involved lowering the minimum age to 42 for denominator inclusion, adding a HCPCS code and removing a revenue center code for identification of the numerator. We did not make any changes to the colorectal screening specifications. For the diabetes care measures, we added and deleted several diagnosis, procedure, and revenue codes used to identify diabetics and to identify the numerators of all three indicators. Additionally, we changed the algorithm for the 50% rule to add the tie breaker logic. For the results that appear in this report, the 50% rule allows multiple physicians to be attributed to the same beneficiary. We added a tie breaker to select only one physician randomly for attribution to the beneficiary when two physicians have 50% of their claims.

We have made all of these edits to the SAS code. Additionally, we have rerun the results for breast cancer screening and colorectal screening for all states except California. We will repopulate the tables in Appendices H and I when all of the updated results are available, and include these in the BQI final report.

Discussion

Under most situations, a minimum denominator of 30 or 50 is needed to achieve conventionally accepted levels of reliability. However reliability scores vary by measure, attribution method, and state.

Reliability and denominator size needs to be considered when reporting quality measures for individual physicians or physician groups or using physician quality measures for quality improvement initiatives (i.e. pay for performance).

Next steps

The first step is to add to our current algorithm the calculation of reliability with no minimum cutoff point. This will be calculated through a model run that includes all physicians regardless of their denominator size. Only a reliability at the average denominator size will be calculated. The second step will be to complete and update the results presented in this report for all measures all states and attribution methods. The third step in this analysis plan is to complete training on beta binomial modeling and reliability conducted by NCQA and John Adams from RAND. We have been mostly following NCQA's direction with the calculation of reliability scores, so we do not anticipate that the training will drastically change any of the numbers we are calculating. Instead, we hope it will help us to know which reliability scores are more appropriate for what interpretations. The fourth step will be to split the physicians into quartiles ranked based on their distribution of scores on the individual measures. We will then examine the means of the highest quartiles compared to the means of lowest quartiles in conjunction with the reliability scores to see if higher reliability scores are accompanied by clinically significant differences between the

highest and lowest quartiles. Finally, after we calculate scores at the physician group level, we will recalculate reliability scores for measures at the group level.

Conclusion

The aggregation of Cycle 2 data, representing Calendar Year 2006, raised new data handling issues and increased the number of measures for each pilot site from five to twelve. This report has presented in depth the challenges and issues as well as strategies used to address them by each pilot site. In addition, significant progress has been made on many of the research questions presented in the Cycle 1 report; and the results of that work have been presented.

Appendix A: BQI Project Measures

1 = Measure calculated during Cycle 1 aggregation
2 = Measure added during Cycle 2 aggregation
3 = Measure added during Cycle 3 aggregation
4 = Measure added during Cycle 4 aggregation

Measure	Specifications	NQF endorsed	AQA Starter Set	AZ	CA	IN	MA	MN[1]	WI[2]
PREVENTION									
Breast Cancer Screening	HEDIS®– BCS	Y	Y	1	1	1	1	1	1
Colorectal cancer screening	HEDIS®– COL	Y	Y	1	1	1	1	2[3]	1
Pneumonia vaccination in 65+ yrs pop	CMS / NCQA	Y	N						2[4]
DIABETES									
LDL testing	HEDIS®– CDC[5]	Y	Y	1	1	1	1	1	1
Diabetes LDL cholesterol control	HEDIS®– CDC[1]	Y	Y			2	4	2	2
HbA1c testing	HEDIS®– CDC[1]	Y	Y	1	1	1	1	1	1
HbA1c control	HEDIS®– CDC[1]	Y	Y			2	4	2	1
Nephropathy testing	HEDIS®– CDC[1]	Y	N	2	2	1	2	1	2
Retinal eye exam	HEDIS®– CDC[1]	Y	Y	1		2	1	1	
Blood pressure control	HEDIS®– CDC[1]	Y	N					2	2
Optimal diabetes composite	MNCM	N	N					2[6]	
Diabetes – smoking status	MNCM	N	N					2	
Diabetes – daily aspirin use	MNCM	N	N					2	
CARDIOVASCULAR CONDITIONS									
Cardiovascular – LDL testing	HEDIS–CMC[7]	Y	N	2	1	2	2		2
Cardiovascular – LDL control	HEDIS–CMC[2]	Y	N			2			2
CAD patients receiving lipid-lowering therapy	PCPI	Y	Y	2	2		2		
HEART ATTACK / HEART FAILURE									
Beta blocker treatment after heart attack[8]	HEDIS®– BBH	Y	Y		2	2	2		
Persistence of beta blocker therapy – Post MI	HEDIS®– PBH	Y	Y	2	2	2	2		
HF: LVF testing	PCPI	Y	Y		2				
Heart failure: ACE inhibitor/ARB therapy	PCPI	Y	Y	2					
Warfarin for patients w/ CHF & atrial fibrillation	PCPI	Y	N	2	2				

[1] MNCM is calculating HbA1C control, LDL control, and BP control based on the direct data submission process. For Cycle 3, MNCM will only calculate the 6 diabetes measures (Composite + components)
[2] WCHQ will use HEDIS® specifications for Medicare FFS data analysis, and WCHQ specifications for aggregated data.
[3] MN will collect this measure via Hybrid data collection method.
[4] Administrative and EMR data only, no survey.
[5] CDC = Comprehensive Diabetes Care. This measure is one of those included in this category.
[6] MNCM Optimal Diabetes Composite: HbA1C control <7.0, LDL – control <100, BP control (collect <130/80 and <140/90), Daily Aspirin Use, Documented Tobacco free
[7] CMC = Cholesterol Management for Patients with Cardiovascular Conditions. This measure is one of those included in this category.
[8] 2008 HEDIS® specifications for public comment indicate potential retirement of this measure (http://www.ncqa.org/Programs/PublicComment/HEDIS2008/index.htm)

Measure	Specifications	NQF endorsed	AQA Starter Set	AZ	CA	IN	MA	MN[1]	WI[2]
MEDICATION MANAGEMENT									
Antidepression Medication Management – Acute Phase	HEDIS®– AMM	Y	Y	2					
Annual monitoring for patients on persistent medications: Ace inhibitors, Digoxin, Diuretics, Anti-convulsants, Statins	HEDIS®– MPM	Y	N		2[9]		2[10]		
OTHER									
Hypertension – blood pressure control	HEDIS®– CBP	Y	N						2
Use of Spirometry Testing in Assessment/Diagnosis of COPD	HEDIS®– SPR	N	N				2		

[9] Annual monitoring for patients on persistent medications: CCHRI:5- Components: Diuretics, Ace inhibitors, Digoxin, Anti-Convulsants, Statins Cycle 1 based on 2005 data, HEDIS® 2006; Cycle 2 based on 2006 data, HEDIS® 2007

[10] Annual monitoring for patients on persistent medications: MHQP: ARBs & Ace inhibitors, calculated separately but reported together.

Appendix B:
BQI Cycle 2 Measures, by Pilot

	Nephropathy Testing	Retinal Eye Exam	Diabetes LDL control	Diabetes HbA1c control	Diabetes Blood Pressure Control	Optimal diabetes composite	Diabetes – smoking status	Diabetes – daily aspirin use	Cardiovascular LDL testing
Center for Health Information & Research (CHIR)									
Age Range	<70 OR >15	x	x	x	x	x	x	x	18-75
Measurement period	1 year (2 tests)	x	x	x	x	x	x	x	1 year
Denominator period	24 months	x	x	x	x	x	x	x	24 months
Specifications	Symmetry EBM Connect (http://www.symmetry-health.com/products/product_SEBM.php), possible algorithm modification								
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care, Medicaid FFS; Medicaid Managed Care								
Specialties / Medical Professionals	Attribution to Primary Care Physicians (Doctors of Medicine, Doctors of Osteopathy) in Internal Medicine, Family Practice, Pediatrics, Geriatrics, and OB/GYN								
California Cooperative Healthcare Reporting Initiative (CCHRI)									
Age Range	18-75	x	x	x	x	x	x	x	x
Measurement period	1 year	x	x	x	x	x	x	x	x
Denominator period	2 years	x	x	x	x	x	x	x	x
Specifications	Administrative. HEDIS® 2007, Medicare Care Management Performance Demonstration Quality Measurement & Reporting Specifications (HF, CAD measures), PCPI (Osteoporosis Management)								
Population covered	Commercial PPO (3 plans); Medicare FFS								
Specialties / Medical Professionals	Attribution to MDs (PCPs, and excluding such MDs such as radiology, etc), DOs, Nurse Practitioners and Physician Assistants. Specialties as per measure specifications, including primary care (family practice, internal medicine), OB/GYN, endocrinology, cardiology, gastroenterology.								
Indiana Health Information Exchange (IHIE)									
Age Range	x	18-75	18-75	18-75	x	x	x	x	18-75
Measurement period	x	1 year	1 year	1 year	x	x	x	x	1 year
Denominator period	x	2 years	2 years	2 years	x	x	x	x	2 years
Specifications	HEDIS® 2007								
Population covered	Commercial Managed Care and FFS; Medicare FFS; Medicare Managed Care								
Specialties / Medical Professionals	Attribution to MDs, DOs, PAs, and NPs.								

Appendix B:
BQI Cycle 2 Measures, by Pilot

	Cardiovascular LDL control	CAD patients receiving lipid-lowering therapy	Beta blocker treatment after heart attack	Persistence of beta blocker treatment post-MI	HF: LVF testing	HF: ACEI/ARB therapy	Warfarin for patient w/CHF and atrial fibrillation	Antidepression medication management	Colorectal cancer screening
Center for Health Information & Research (CHIR)									
Age Range	x	>19	x	>35	x	>35	>19	> 18 2/3 yrs	x
Measurement period	x	1 year	x	1 year	x	1 year	1 year	20 months	x
Denominator period	x	1 year	x	1 year	x	1 year	1 year	1 year	x
Specifications	Symmetry EBM Connect (http://www.symmetry-health.com/products/product_SEBM.php), possible algorithm modification								
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care, Medicaid FFS; Medicaid Managed Care								
Specialties / Medical Professionals	Attribution to Primary Care Physicians (Doctors of Medicine, Doctors of Osteopathy) in Internal Medicine, Family Practice, Pediatrics, Geriatrics, and OB/GYN								
California Cooperative Healthcare Reporting Initiative (CCHRI)									
Age Range	x	>=18	x	>= 35	>=18	x	>=18	x	x
Measurement period	x	1 year	x	1 year	1 year	x	1 year	x	x
Denominator period	x	1 year	x	1 year	1 year	x	1 year	x	x
Specifications	Administrative. HEDIS® 2007, Medicare Care Management Performance Demonstration Quality Measurement & Reporting Specifications (HF, CAD measures), PCPI (Osteoporosis Management)								
Population covered	Commercial PPO (3 plans); Medicare FFS								
Specialties / Medical Professionals	Attribution to MDs (PCPs, and excluding such MDs such as radiology, etc), DOs, Nurse Practitioners and Physician Assistants. Specialties as per measure specifications, including primary care (family practice, internal medicine), OB/GYN, endocrinology, cardiology, gastroenterology.								
Indiana Health Information Exchange (IHIE)									
Age Range	18-75	x	>=35	>=35	x	x	x	x	x
Measurement period	1 year	x	1 year	1 year	x	x	x	x	x
Denominator period	2 years	x	18 months	18 months	x	x	x	x	x
Specifications	HEDIS® 2007								
Population covered	Commercial Managed Care and FFS; Medicare FFS; Medicare Managed Care								
Specialties / Medical Professionals	Attribution to MDs, DOs, PAs, and NPs.								

Appendix B:
BQI Cycle 2 Measures, by Pilot

	Annual monitoring for patient on persistent meds	Hypertension – blood pressure control	Use of spirometry testing	Osteoporosis management	Pneumococcal vaccination
Center for Health Information & Research (CHIR)					
Age Range	x	x	x	x	x
Measurement period					
	x	x	x	x	x
Denominator period	x	x	x	x	x
Specifications	Symmetry EBM Connect (http://www.symmetry-health.com/products/product_SEBM.php), possible algorithm modification				
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care, Medicaid FFS; Medicaid Managed Care				
Specialties / Medical Professionals	Attribution to Primary Care Physicians (Doctors of Medicine, Doctors of Osteopathy) in Internal Medicine, Family Practice, Pediatrics, Geriatrics, and OB/GYN				
California Cooperative Healthcare Reporting Initiative (CCHRI)					
Age Range	>=18	x	x	>=67	x
Measurement period	1 year	x	x	1 year	x
Denominator period	1 year	x	x	18 months	x
Specifications	Administrative. HEDIS® 2007, Medicare Care Management Performance Demonstration Quality Measurement & Reporting Specifications (HF, CAD measures), PCPI (Osteoporosis Management)				
Population covered	Commercial PPO (3 plans); Medicare FFS				
Specialties / Medical Professionals	Attribution to MDs (PCPs, and excluding such MDs such as radiology, etc), DOs, Nurse Practitioners and Physician Assistants. Specialties as per measure specifications, including primary care (family practice, internal medicine), OB/GYN, endocrinology, cardiology, gastroenterology.				
Indiana Health Information Exchange (IHIE)					
Age Range	x	x	x	x	x
Measurement period	x	x	x	x	x
Denominator period	x	x	x	x	x
Specifications	HEDIS® 2007				
Population covered	Commercial Managed Care and FFS; Medicare FFS; Medicare Managed Care				
Specialties / Medical Professionals	Attribution to MDs, DOs, PAs, and NPs.				

Appendix B:
BQI Cycle 2 Measures, by Pilot

	Nephropathy Testing	Retinal Eye Exam	Diabetes LDL control	Diabetes HbA1c control	Diabetes Blood Pressure Control	Optimal diabetes composite	Diabetes – smoking status	Diabetes – daily aspirin use	Cardiovascular LDL testing
Massachusetts Health Quality Partners (MHQP)									
Age Range	18-75	x	x	x	x	x	x	x	18-75
Measurement period	1 year	x	x	x	x	x	x	x	1 year
Denominator period	2 years	x	x	x	x	x	x	x	2 years
Specifications	HEDIS® 2007, with claims adjustment factor as necessary								
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care								
Specialties / Medical Professionals	Attribution to MDs and DOs (PCPs), except where the patient did not see any primary care physician for a preventive or routine visit in the measurement period. Endocrinologists are included as PCPs where they are the only physician that has been seen for routine care. Other specialists to be included as PCPs if a plan had so designated them.								
Minnesota Community Measurement (MNCM)									
Age Range	x	x	18-75	18-75	18-75	18-75	18-75	18-75	x
Measurement period	x	x	1 year	1 year	1 year	1 year	1 year	1 year	x
Denominator period	x	x	2 years	2 years	2 years	2 years	2 years	2 years	x
Specifications	HEDIS® 2007 (Hybrid for Colorectal Cancer); MNCM								
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care; Medicaid FFS								
Specialties / Medical Professionals	All as appropriate.								
Wisconsin Collaborative for Healthcare Quality (WCHQ)									
Age Range	18-85	x	18-85	x	18-85	x	x	x	18-85
Measurement period	1 year	x	1 year	x	1 year	x	x	x	1 year
Denominator period	2 years	x	2 years	x	2 years	x	x	x	2 years
Specifications	WCHQ. Medical groups/providers determine denominators as per WCHQ methodology, and then provide numerator information.								
Population covered	All payers; Medicare FFS.								
Specialties / Medical Professionals	PAs, NPs, DOs, MDs.								

Appendix B: BQI Cycle 2 Measures, by Pilot

	Cardiovascular LDL control	CAD patients receiving lipid-lowering therapy	Beta blocker treatment after heart attack	Persistence of beta blocker treatment post-MI	HF: LVF testing	HF: ACEI/ARB therapy	Warfarin for patient w/CHF and atrial fibrillation	Antidepression medication management	Colorectal cancer screening
Massachusetts Health Quality Partners (MHQP)									
Age Range	x	>=18	>=35	>=35	x	x	x	x	x
Measurement period	x	1 year	1 year	1 year	x	x	x	x	x
Denominator period	x	1 year	18 months	18 months	x	x	x	x	x
Specifications	HEDIS® 2007, with claims adjustment factor as necessary								
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care								
Specialties / Medical Professionals	Attribution to MDs and DOs (PCPs), except where the patient did not see any primary care physician for a preventive or routine visit in the measurement period. Endocrinologists are included as PCPs where they are the only physician that has been seen for routine care. Other specialists to be included as PCPs if a plan had so designated them.								
Minnesota Community Measurement (MNCM)									
Age Range	x	x	x	x	x	x	x	x	50-80
Measurement period	x	x	x	x	x	x	x	x	10 years if possible
Denominator period	x	x	x	x	x	x	x	x	10 years if possible
Specifications	HEDIS® 2007 (Hybrid for Colorectal Cancer); MNCM								
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care; Medicaid FFS								
Specialties / Medical Professionals	All as appropriate.								
Wisconsin Collaborative for Healthcare Quality (WCHQ)									
Age Range	18-85	x	x	x	x	x	x	x	x
Measurement period	1 year	x	x	x	x	x	x	x	x
Denominator period	2 years	x	x	x	x	x	x	x	x
Specifications	WCHQ. Medical groups/providers determine denominators as per WCHQ methodology, and then provide numerator information.								
Population covered	All payers; Medicare FFS.								
Specialties / Medical Professionals	PAs, NPs, DOs, MDs.								

Appendix B:
BQI Cycle 2 Measures, by Pilot

	Annual monitoring for patient on persistent meds	Hypertension – blood pressure control	Use of spirometry testing	Osteoporosis management	Pneumococcal vaccination
Massachusetts Health Quality Partners (MHQP)					
Age Range	>=18	x	>=40	x	x
Measurement period	1 year	x	1 year	x	x
Denominator period	1 year	x	2 years	x	x
Specifications	HEDIS® 2007, with claims adjustment factor as necessary				
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care				
Specialties / Medical Professionals	Attribution to MDs and DOs (PCPs), except where the patient did not see any primary care physician for a preventive or routine visit in the measurement period. Endocrinologists are included as PCPs where they are the only physician that has been seen for routine care. Other specialists to be included as PCPs if a plan had so designated them.				
Minnesota Community Measurement (MNCM)					
Age Range	x	x	x	x	x
Measurement period	x	x	x	x	x
Denominator period	x	x	x	x	x
Specifications	HEDIS® 2007 (Hybrid for Colorectal Cancer); MNCM				
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care; Medicaid FFS				
Specialties / Medical Professionals	All as appropriate.				
Wisconsin Collaborative for Healthcare Quality (WCHQ)					
Age Range	x	18-85	x	x	>=65
Measurement period	x	1 year	x	x	1 year
Denominator period	x	2 years	x	x	3 years
Specifications	WCHQ. Medical groups/providers determine denominators as per WCHQ methodology, and then provide numerator information.				
Population covered	All payers; Medicare FFS.				
Specialties / Medical Professionals	PAs, NPs, DOs, MDs.				

Appendix C: Pilot Physician-Physician Group Attribution Methods

Pilot Site	Attribution Method
Arizona CHIR	CHIR assigns its own unique physician IDs to individual practitioners. CHIR has found that differing payer databases have differing provider IDs. Only Medicare uses a TIN. Others use UPIN +6, +4, and more. Many of the UPINs are out of date, specifically regarding previous groups that are not current. This has been very problematic. CHIR is exploring using State Licenture IDs for individual and possibly group identification.
California CCHRI	Data is received from both payers and providers at the individual practitioner level only. CCHRI has explored a probability-based linking method. Many (perhaps 50%?) of providers are individual. Group assignment is very problematic.
Indiana IHIE	IHIE built a database attributing physicians to groups including UPINs, TINs and EINs. The database was validated by provider self-identification. The database is updated ad hoc through an electronic Report Delivery System, using voluntary input from providers and payers. Groups are motivated to be current in order to achieve best P4P reward.
Massachusetts MHQP	MHQP started many years ago assigning physicians to groups using payer databases. Found frequent disagreement among payers, and the provider groups felt all payer results were inaccurate. Therefore MHQP changed to current system that uses provider-defined groups. Updated at least annually through provider-available web portal. Reliable, well-accepted by patients and providers, validated, engages groups actively, helps relationships. MHQP has a tiered system of reporting data to public. Can report on medical groups of minimum 3 physicians, practice sites, or larger networks. MHQP found physicians associate into groups differently for different payers. Groups are asked to review and correct the following data elements: contact information, practices listed within their group, physicians and specialties within each practice, and responsible contact person in each group. In the future MHQP intends to add data elements such as role of the physician, role of the contact person, and multiple contacts.
Minnesota MNCM	MNCM aggregates all practice locations with the same corporate name into a unique Group ID#. In the MNCM market each practice location address has its own TIN, but in some cases up to 10-25 locations have the same corporate name and are therefore aggregated. Data is therefore reported by the corporate name Group ID, not the separate practice location.
Wisconsin WCHQ #1	WCHQ's existing system was based on data *collected and reported by provider groups* for ambulatory care. Therefore, providers were automatically grouped, and WCHQ only reported aggregated data for each group. No data from other the provider group's EMR was used. Data from groups included ALL PATIENTS AND ALL PAYERS non-separated. Accepted and relevant to providers and patients. This method requires physicians to measure and self-report numerators/denominators.
Wisconsin WCHQ #2	For Medicare FFS data in BQI, WCHQ has created groups similar to MN using corporate name. When WCHQ has measures that require hospital and other data, they have been trying to link hospital to outpatient data through complex linkage of UPIN+6 in hosps and TIN in offices, using intermediate databases of members and carriers.

Characteristics of Group Definitions By Exclusion Criteria

Physician Group Characteristics	TAX ID Method (CLAIMS) - INDIANA								
	(A) Greater than or equal to 1 physician /group & Business Location= 'Any' (Comparison Group)		(B) Greater than or equal to 1 physicians /group & Business Location= Indiana		(C) Greater than or equal to 2 physicians /group & Business Location= 'Any'		(D) Greater than or equal to 2 physicians /group & Business Location= Indiana		Overall Direction of % change from A to (B, C, & D)
	Number/ Average	%	Number/ Average	% Change from A	Number/ Average	% Change from A	Number/ Average	% Change from A	
# of Groups	7,329	100%	3,694	-50%	2,102	-71%	1,565	-79%	↓
# of physicians Attributed	22,180	100%	14,966	-33%	16,912	-24%	12,837	-42%	↓
# of Unique Physicians	18,180	100%	11,733	-35%	14,105	-22%	10,454	-42%	↓
# of beneficiaries attributed	4,168,084	100%	3,057,151	-27%	3,334,961	-20%	2,708,855	-35%	↓
# Unique beneficiaries	813,398	100%	768,245	-6%	781,254	-4%	746,580	-8%	↓
Mean physicians / Group	3.02	100%	4.05	34%	8.05	167%	8.20	172%	↑
Mean beneficiaries / Group	568.71	100%	827.60	46%	1,586.57	179%	1,730.90	204%	↑

* Selected for future comparison

Characteristics of Group Definitions By Exclusion Criteria

Physician Group Characteristics	TAX ID METHOD (MEMBER) - INDIANA								
	(A) Greater than or equal to 1 physician /group & Business Location= 'Any' (Comparison Group)		(B) Greater than or equal to 1 physicians /group & Business Location= Indiana		(C) Greater than or equal to 2 physicians /group & Business Location= 'Any'		(D) Greater than or equal to 2 physicians /group & Business Location= Indiana		Overall Direction of % change from A to (B, C, & D)
	Number/ Average	%	Number/ Average	% Change from A	Number/ Average	% Change from A	Number/ Average	% Change from A	
# of Groups	3,238	100%	3,078	-5%	2,104	-35%	2,040	-37%	↓
# of physicians Attributed	24,013	100%	23,419	-2%	22,878	-5%	22,381	-7%	↓
# of Unique Physicians	15,769	100%	15,333	-3%	15,051	-5%	14,687	-7%	↓
# of beneficiaries attributed	2,512,362	100%	2,467,669	-2%	2,367,554	-6%	2,323,403	-8%	↓
# Unique beneficiaries	733,735	100%	729,741	-1%	722,841	-1%	718,511	-2%	↓
Mean physicians / Group	7.42	100%	7.61	3%	10.87	46%	10.97	48%	↑
Mean beneficiaries / Group	775.9	100%	801.72	3%	1,125.26	45%	1,138.92	47%	↑

* Selected for future comparison

Characteristics of Group Definitions By Exclusion Criteria

Physician Group Characteristics	GROUP UPIN6 METHOD – INDIANA								
	(A) Greater than or equal to 1 physician /group & Business Location= 'Any' (Comparison Group)		(B) Greater than or equal to 1 physicians /group & Business Location= Indiana		(C) Greater than or equal to 2 physicians /group & Business Location= 'Any'		(D) Greater than or equal to 2 physicians /group & Business Location= Indiana		Overall Direction of % change from A to (B, C, & D)
	Number/ Average	%	Number/ Average	% Change from A	Number/ Average	% Change from A	Number/ Average	% Change from A	
# of Groups	8,435	100%	2,909	-66%	3,329	-61%	1,860	-78%	↓
# of physicians Attributed	30,082	100%	19,704	-34%	24,976	-17%	18,655	-38%	↓
# of Unique Physicians	15,488	100%	13,028	-16%	14,031	-9%	12,420	-20%	↓
# of beneficiaries attributed	3,860,565	100%	3,559,114	-8%	3,459,091	-10%	3,331,154	-14%	↓
# Unique beneficiaries	780,624	100%	774,924	-1%	767,899	-2%	764,041	-2%	↓
Mean physicians / Group	3.57	100%	6.77	90%	7.5	110%	10.02	181%	↑
Mean beneficiaries / Group	457.68	100%	1,223.48	167%	1,039.08	127%	1,790.94	291%	↑

* Selected for future comparison

Characteristics of Group Definitions By Exclusion Criteria

	GROUP (UPIN6 + UPIN4) METHOD - INDIANA								
	(A) Greater than or equal to 1 physician /group & Business Location= 'Any' (Comparison Group)		(B) Greater than or equal to 1 physicians /group & Business Location= Indiana		(C) Greater than or equal to 2 physicians /group & Business Location= 'Any'		(D) Greater than or equal to 2 physicians /group & Business Location= Indiana		Overall Direction of % change from A to (B, C, & D)
	Number/ Average	%	Number/ Average	% Change from A	Number/ Average	% Change from A	Number/ Average	% Change from A	
Physician Group Characteristics	12,868	100%	4,432	-66%	4,929	-62%	2,839	-78%	↓
# of Groups	38,607	100%	23,958	-38%	30,668	-21%	22,365	-42%	↓
# of physicians Attributed	15,488	100%	13,028	-16%	12,041	-22%	12,200	-21%	↓
# of Unique Physicians	4,115,130	100%	3,804,221	-8%	3,578,119	-13%	3,499,098	-15%	↓
# of beneficiaries attributed	780,624	100%	774924	-1%	752,074	-4%	761,787	-2%	↓
# Unique beneficiaries	3.00	100%	5.4	80%	6.22	107%	7.88	163%	↑
Mean physicians / Group	319.81	100%	858.35	168%	725.93	127%	1,232.51	285%	↑

* Selected for future comparison

Characteristics of Group Definitions By Exclusion Criteria

TAX ID METHOD (CLAIMS) - ARIZONA

Physician Group Characteristics	(A) Greater than or equal to 1 physician /group & Business Location= 'Any' (Comparison Group)		(B) Greater than or equal to 1 physicians /group & Business Location= Arizona		(C) Greater than or equal to 2 physicians /group & Business Location= 'Any'		(D) Greater than or equal to 2 physicians /group & Business Location= Arizona		Overall Direction of % change from A to (B, C, & D)
	Number/ Average	%	Number/ Average	% Change from A	Number/ Average	% Change from A	Number/ Average	% Change from A	
# of Groups	7,112	100%	3,414	-52%	1,632	-77%	1,094	-85%	↓
# of physicians Attributed	18,452	100%	11,037	-40%	12,967	-30%	8,717	-53%	↓
# of Unique Physicians	15,079	100%	8,314	-45%	10,458	-31%	6,643	-56%	↓
# of beneficiaries attributed	2,538,367	100%	1,741,779	-31%	1,790,379	-29%	1,354,619	-47%	↓
# Unique beneficiaries	507,320	100%	464,768.00	-8%	469,654	-7%	433,647	-15%	↓
Mean physicians / Group	2.59	100%	3.23	25%	7.95	207%	7.97	208%	↑
Mean beneficiaries / Group	356.66	100%	510.19	43%	1,097.05	208%	1,238.23	247%	↑

* Selected for future comparison

Characteristics of Group Definitions By Exclusion Criteria

TAX ID METHOD (MEMBER) -ARIZONA

Physician Group Characteristics	(A) Greater than or equal to 1 physician /group & Business Location= 'Any' (Comparison Group)		(B) Greater than or equal to 1 physicians /group & Business Location= Arizona		(C) Greater than or equal to 2 physicians /group & Business Location= 'Any'		(D) Greater than or equal to 2 physicians /group & Business Location= Arizona		Overall Direction of % change from A to (B, C, & D)
	Number/ Average	%	Number/ Average	% Change from A	Number/ Average	% Change from A	Number/ Average	% Change from A	
# of Groups	3,120	100%	2,996	-4%	1,662	-47%	1,638	-48%	↓
# of physicians Attributed	17,651	100%	17,432	-1%	16,193	-8%	16,074	-8%	↓
# of Unique Physicians	11,359	100%	11,191	-1%	10,349	-9%	10,264	-10%	↓
# of beneficiaries attributed	1,185,921	100%	1,179,823	-1%	1,010,379	-17%	1,003,714	-18%	↓
# Unique beneficiaries	418,696	100%	417,920	0%	396,903	-5%	395,741	-5%	↓
Mean physicians / Group	5.66	100%	5.82	3%	9.74	72%	9.81	73%	↑
Mean beneficiaries / Group	380.10	100%	393.8	4%	607.93	60%	612.77	61%	↑

* Selected for future comparison

Characteristics of Group Definitions By Exclusion Criteria

GROUP UPIN6 METHOD - ARIZONA

Physician Group Characteristics	(A) Greater than or equal to 1 physician /group & Business Location= 'Any' (Comparison Group)		(B) Greater than or equal to 1 physicians /group & Business Location= Arizona		(C) Greater than or equal to 2 physicians /group & Business Location= 'Any'		(D) Greater than or equal to 2 physicians /group & Business Location= Arizona		Overall Direction of % change from A to (B, C, & D)
	Number/ Average	%	Number/ Average	% Change from A	Number/ Average	% Change from A	Number/ Average	% Change from A	
# of Groups	7,336	100%	2,732	-63%	2,435	-67%	1,386	-81%	↓
# of physicians Attributed	22,012	100%	14,494	-34%	17,111	-22%	13,148	-40%	↓
# of Unique Physicians	11,707	100%	9,537	-19%	9,924	-15%	8,650	-26%	↓
# of beneficiaries attributed	2,510,660	100%	2,193,530	-13%	2,057,536	-18%	1,937,845	-23%	↓
# Unique beneficiaries	493,938	100%	487,413	-1%	477,528	-3%	474,740	-4%	↓
Mean physicians / Group	3.00	100%	5.3	77%	7.03	134%	9.49	216%	↑
Mean beneficiaries / Group	342.24	100%	802.9	135%	844.98	147%	1,398	308%	↑

* Selected for future comparison

Characteristics of Group Definitions By Exclusion Criteria

GROUP (UPIN6 + UPIN4) METHOD - ARIZONA

Physician Group Characteristics	(A) Greater than or equal to 1 physician /group & Business Location= 'Any' (Comparison Group)		(B) Greater than or equal to 1 physicians /group & Business Location= Arizona		(C) Greater than or equal to 2 physicians /group & Business Location= 'Any'		(D) Greater than or equal to 2 physicians /group & Business Location= Arizona		Overall Direction of % change from A to (B, C, & D)
	Number/ Average	%	Number/ Average	% Change from A	Number/ Average	% Change from A	Number/ Average	% Change from A	
# of Groups	9,918	100%	3,268	-67%	2,987	-70%	1,769	-82%	↓
# of physicians Attributed	25,774	100%	16,349	-37%	18,843	-27%	14,850	-42%	↓
# of Unique Physicians	11,707	100%	9,537	-19%	9,594	-18%	8,581	-27%	↓
# of beneficiaries attributed	2,574,544	100%	2,261,242	-12%	2,067,800	-20%	1,988,996	-23%	↓
# Unique beneficiaries	493,938	100%	487,413	-1%	475,353	-4%	473,863	-4%	↓
Mean physicians / Group	2.60	100%	5	92%	6.31	143%	8.39	223%	↑
Mean beneficiaries / Group	259.58	100%	691.93	167%	692.27	167%	1,124.36	333%	↑

* Selected for future comparison

Appendix E. List of Evaluation and Management Codes used to identify physicians eligible to be attributed to beneficiaries

Category / Subcategory	Codes	Category / Subcategory	Codes
Office or Other Outpatient Services	99201-99205	Subsequent Nursing Facility Care	99311-99313
New Patient	99211-99215	Nursing Facility Discharge Services	99315-99316
Established Patient	99217	Domiciliary, Rest Home or	
Hospital Observation Discharge Services	99218-99220	Custodial Care Services	
Hospital Observation Services		New Patient	99321-99323
Hospital Observation or Inpatient Care		Established Patient	99331-99333
Services (Including Admission and Discharge	99234-99236	Home Services	
Services)		New Patient	99341-99345
Hospital Inpatient Services	99221-99223	Established Patient	99347-99350
Initial Hospital Care	99231-99233	Prolonged Services	
Subsequent Hospital Care	99238-99239	With Direct Patient Contact	99354-99357
Hospital Discharge Services		Without Direct Patient Contact	99358-99359
Consultations	99241-99245	Standby Services	99360
Office Consultations	99251-99255	Case Management Services	
Initial Inpatient Consultations	99261-99263	Team Conferences	99361-99362
Follow-up Inpatient Consultations	99271-99275--	Telephone Calls	99371-99373
Confirmatory Consultations	99281-99288	Care Plan Oversight Services	99374-99380 —
Emergency Department Services	99289-99290	Preventive Medicine Services	
Pediatric Patient Transport		New Patient	99381 -99387 --
Critical Care Services	99291-99292	Established Patient	99391-99397-
Adult (over 24 months of age)	99293-99294	Individual Counseling	99401-99404
Pediatric	99295-99296	Group Counseling	99411-99412
Neonatal	99298-99299	Other	99420-99429
Intensive Care (Low Birth Weight)		Newborn Care	99431-99440
Nursing Facility Services		Special E/M Services	99450-99456
Comprehensive Nursing Facility Assessments	99301-99303	Other E/M Services	99499 —

Appendix F: Definition of PCP by Provider Specialty and Measure

Measure	Primary Care Provider Specialties
Breast Cancer Screening	Primary Care Physicians[3] including: General Practice Family Practice Internal Medicine Geriatrician Obstetrics and Gynecology Oncologist
LDL testing for diabetics	Primary Care Physicians including: General Practice Family Practice Internal Medicine Geriatrician Endocrinologist
HbA1c testing for diabetics	
Nephropathy testing for diabetics	
Colorectal cancer screening	Primary Care Physicians including: General Practice Family Practice Internal Medicine Geriatrician Gastroenterologist

Results using 1 touch rule

Measure	State	Statewide Rate	Denominator	Eligible Physicians*	1 touch rule			
					Mean Physician Rate	Standard Deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physcian‡
Breast Cancer Screening	AZ	59.8%	47,525	5,848	72.7%	19.2%	70.4%	84.3%
	CA		0	0				
	IN	57.9%	86,032	9,519	67.8%	18.4%	62.3%	89.9%
	MA	68.9%	70,979	8,519	79.7%	16.1%	77.5%	82.4%
	MN	68.5%	44,514	7,425	77.5%	17.2%	76.9%	82.5%
	WI	66.7%	62,685	6,862	74.1%	17.2%	77.2%	84.2%
Colorectal Cancer Screening	AZ		0	0				
	CA		0	0				
	IN	38.3%	463,760	7,360	50.5%	18.9%	96.5%	88.5%
	MA		38,986	3,776			85.2%	104.4%
	MN		0	0				
	WI		0	0				
HbA1c Testing	AZ	73.7%	48,155	5,848	82.9%	15.9%	73.3%	85.3%
	CA		0	0				
	IN		0	0				
	MA		0	0				
	MN		0	0				
	WI		0	0				
Retinal Eye Exam	AZ		0	0				
	CA		0	0				
	IN		0	0				
	MA		0	0				
	MN		0	0				
	WI		0	0				
LDL Testing	AZ	69.7%	48,155	5,848	80.0%	17.4%	73.3%	85.3%
	CA		0	0				
	IN		0	0				
	MA		0	0				
	MN		0	0				
	WI		0	0				
Nephropathy Testing	AZ	62.2%	48,155	5,848	74.8%	19.3%	73.3%	85.3%
	CA		0	0				
	IN		0	0				
	MA		0	0				
	MN		0	0				
	WI		0	0				

* "Eligible physicians" is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

† "% eligible physicians w/ rate" is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ "% beneficiaries in denominator w/ a physician" is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Results using 2 touch rule

Measure	State	Statewide Rate	Denominator	2 touch rule			
				Mean Physician Rate	Standard Deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physcian‡
Breast Cancer Screening	AZ	59.8%	47,525	73.8%	19.4%	62.4%	78.2%
	CA		0				
	IN	57.9%	86,032	68.8%	19.2%	55.8%	84.2%
	MA	68.9%	70,979	80.6%	17.4%	67.4%	75.2%
	MN	68.5%	44,514	79.2%	18.5%	65.9%	72.4%
	WI	66.7%	62,685	75.2%	18.4%	68.5%	74.9%
Colorectal Cancer Screening	AZ		0				
	CA		0				
	IN	38.3%	463,760	51.1%	19.2%	86.3%	83.0%
	MA		38,986			71.6%	91.0%
	MN		0				
	WI		0				
HbA1c Testing	AZ	73.7%	48,155	84.1%	15.8%	64.6%	81.0%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
Retinal Eye Exam	AZ		0				
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
LDL Testing	AZ	69.7%	48,155	81.3%	17.4%	64.6%	81.0%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
Nephropathy Testing	AZ	62.2%	48,155	75.2%	20.1%	64.6%	81.0%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				

* "Eligible physicians" is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

† "% eligible physicians w/ rate" is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ "% beneficiaries in denominator w/ a physician" is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Results using 30% rule

Measure	State	Statewide Rate	Denominator	30% rule			
				Mean Physician Rate	Standard Deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physcian‡
Breast Cancer Screening	AZ	59.8%	47,525	69.0%	21.6%	55.2%	83.3%
	CA		0				
	IN	57.9%	86,032	64.7%	21.0%	50.8%	89.0%
	MA	68.9%	70,979	77.7%	19.4%	63.0%	81.8%
	MN	68.5%	44,514	76.3%	20.7%	62.4%	80.9%
	WI	66.7%	62,685	73.0%	19.9%	67.1%	83.4%
Colorectal Cancer Screening	AZ		0				
	CA		0				
	IN	38.3%	463,760	44.5%	20.0%	80.0%	87.8%
	MA		38,986			78.8%	105.4%
	MN		0				
	WI		0				
HbA1c Testing	AZ	73.7%	48,155	80.8%	18.9%	55.0%	84.5%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
Retinal Eye Exam	AZ		0				
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
LDL Testing	AZ	69.7%	48,155	77.3%	20.1%	55.0%	84.5%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
Nephropathy Testing	AZ	62.2%	48,155	71.6%	22.6%	55.0%	84.5%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				

* "Eligible physicians" is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

† "% eligible physicians w/ rate" is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ "% beneficiaries in denominator w/ a physician" is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

Results using 50% rule

Measure	State	Statewide Rate	Denominator	50% rule			
				Mean Physician Rate	Standard Deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physician‡
Breast Cancer Screening	AZ	59.8%	47,525	69.2%	22.5%	47.1%	70.4%
	CA		0				
	IN	57.9%	86,032	64.7%	21.9%	43.4%	76.6%
	MA	68.9%	70,979	78.1%	20.0%	53.9%	71.2%
	MN	68.5%	44,514	77.2%	21.3%	51.7%	66.1%
	WI	66.7%	62,685	73.9%	21.0%	58.8%	71.5%
Colorectal Cancer Screening	AZ		0				
	CA		0				
	IN	38.3%	463,760	43.3%	21.0%	69.6%	77.2%
	MA		38,986			0.0%	0.0%
	MN		0				
	WI		0				
HbA1c Testing	AZ	73.7%	48,155	81.6%	18.7%	46.2%	73.7%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
Retinal Eye Exam	AZ		0				
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
LDL Testing	AZ	69.7%	48,155	78.4%	20.4%	46.2%	73.7%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
Nephropathy Testing	AZ	62.2%	48,155	71.5%	23.5%	46.2%	73.7%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				

* "Eligible physicians" is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

† "% eligible physicians w/ rate" is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ "% beneficiaries in denominator w/ a physician" is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

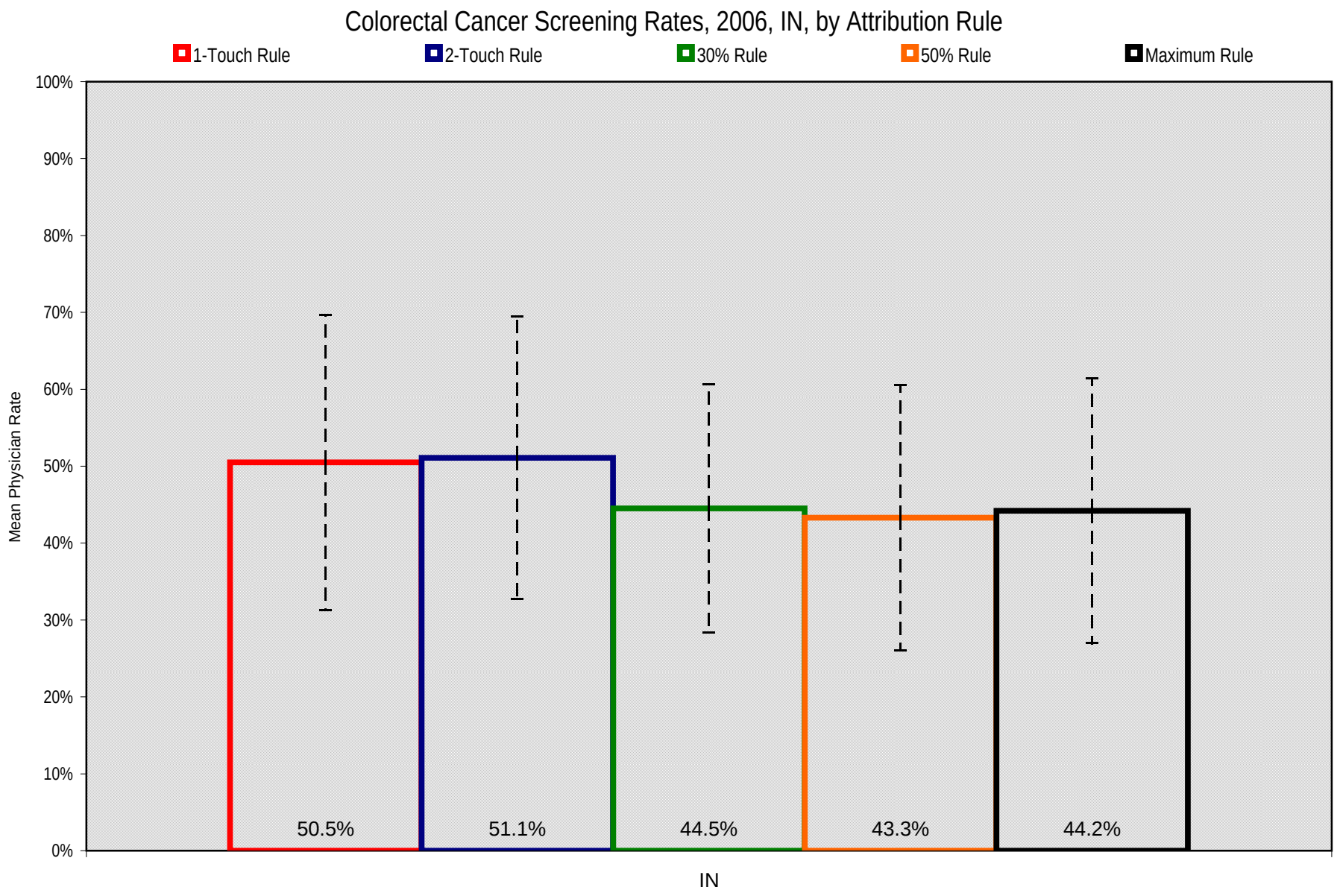
Results using maximum frequency rule

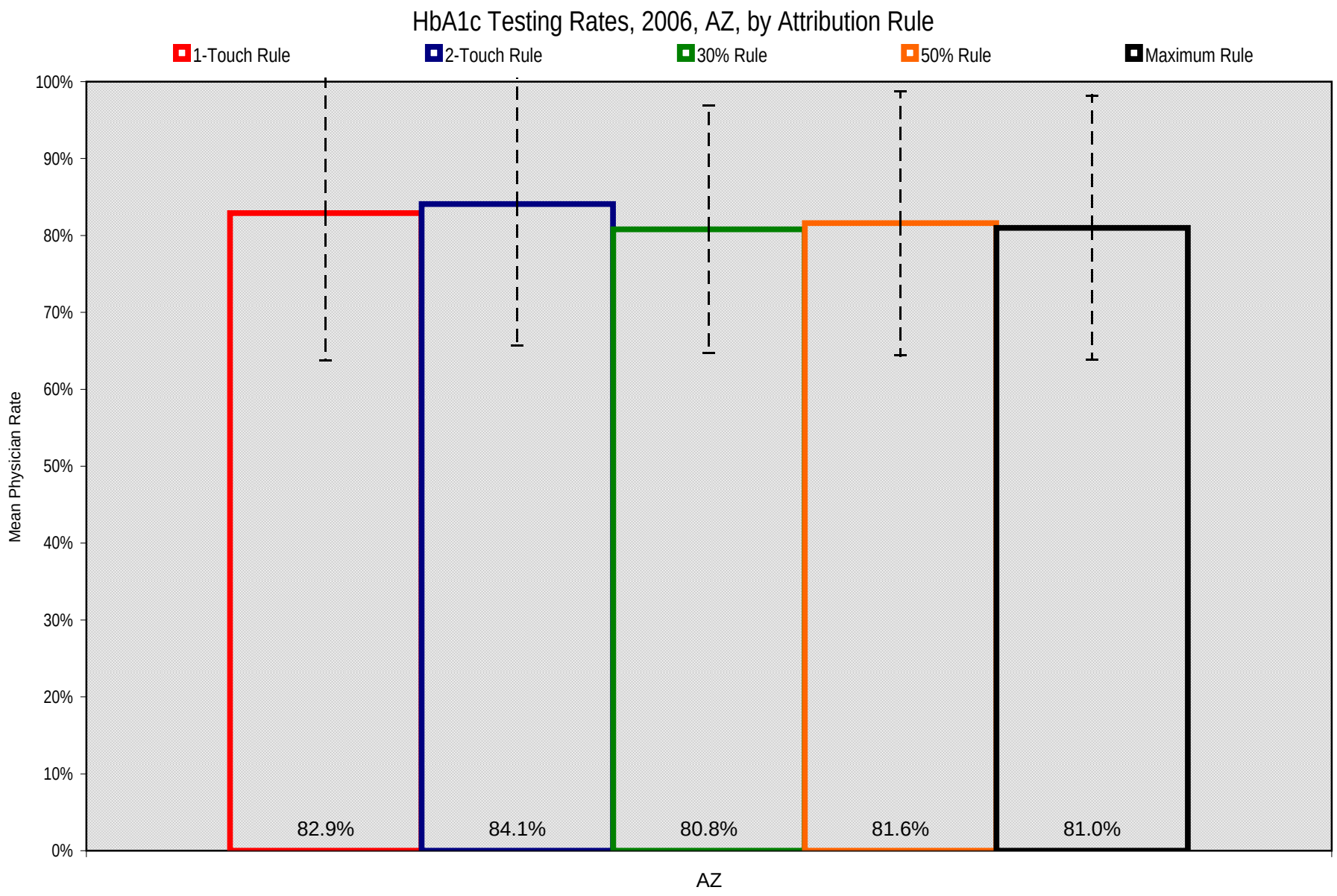
Measure	State	Statewide Rate	Denominator	Maximum Frequency rule			
				Mean Physician Rate	Standard Deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physician‡
Breast Cancer Screening	AZ	59.8%	47,525	69.2%	22.0%	51.8%	84.3%
	CA		0				
	IN	57.9%	86,032	65.0%	21.3%	47.7%	89.9%
	MA	68.9%	70,979	77.9%	19.7%	58.7%	82.4%
	MN	68.5%	44,514	76.7%	20.9%	58.5%	82.5%
	WI	66.7%	62,685	73.4%	20.5%	63.3%	84.2%
Colorectal Cancer Screening	AZ		0				
	CA		0				
	IN	38.3%	463,760	44.2%	20.4%	75.1%	88.5%
	MA		38,986				
	MN		0				
	WI		0				
HbA1c Testing	AZ	73.7%	48,155	81.0%	18.9%	50.9%	85.3%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
Retinal Eye Exam	AZ		0				
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
LDL Testing	AZ	69.7%	48,155	77.6%	20.4%	50.9%	85.3%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				
Nephropathy Testing	AZ	62.2%	48,155	71.6%	23.1%	50.9%	85.3%
	CA		0				
	IN		0				
	MA		0				
	MN		0				
	WI		0				

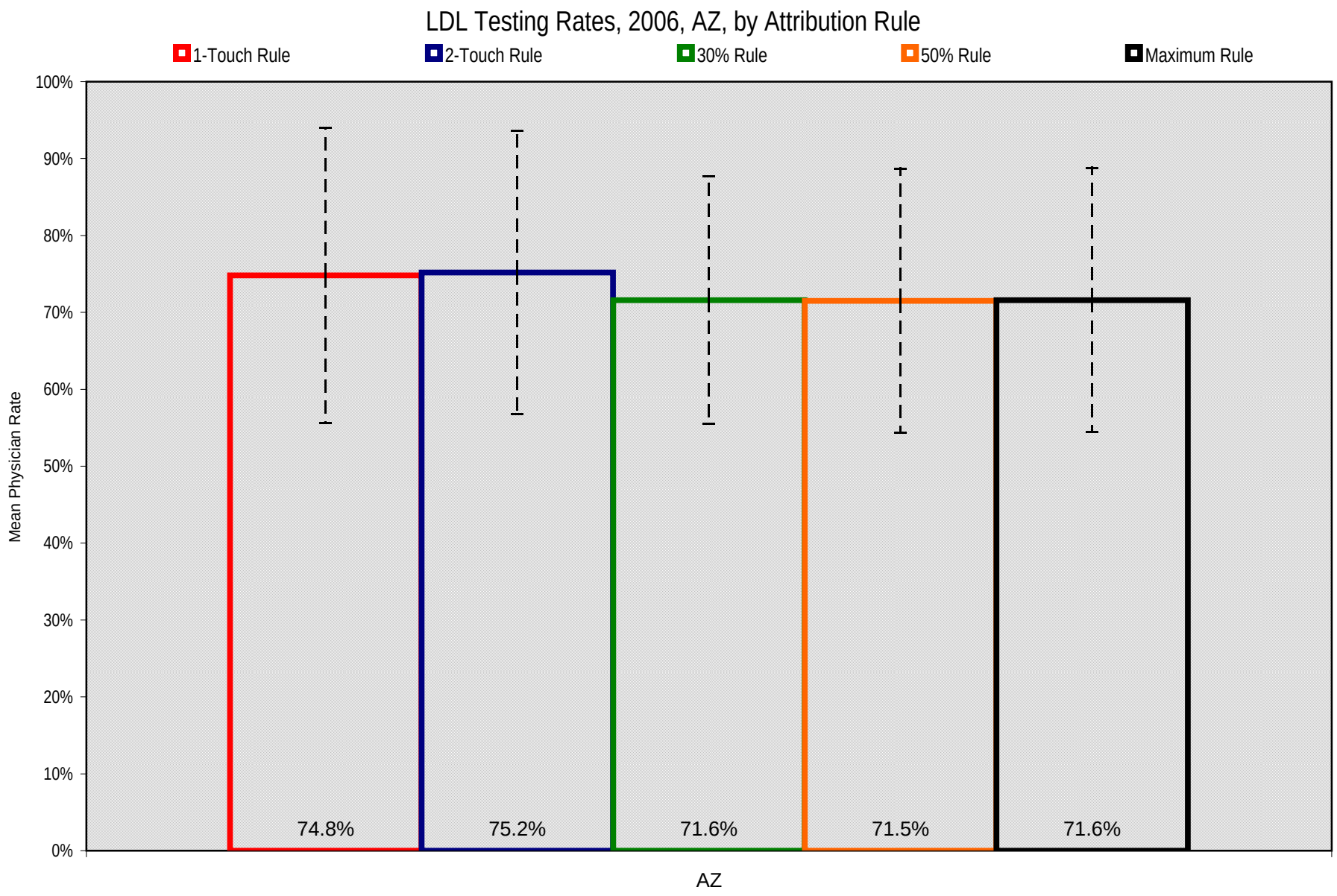
* "Eligible physicians" is the number of physicians with an E&M claim in an office setting who have a relevant specialty to the measure and are licensed in the state

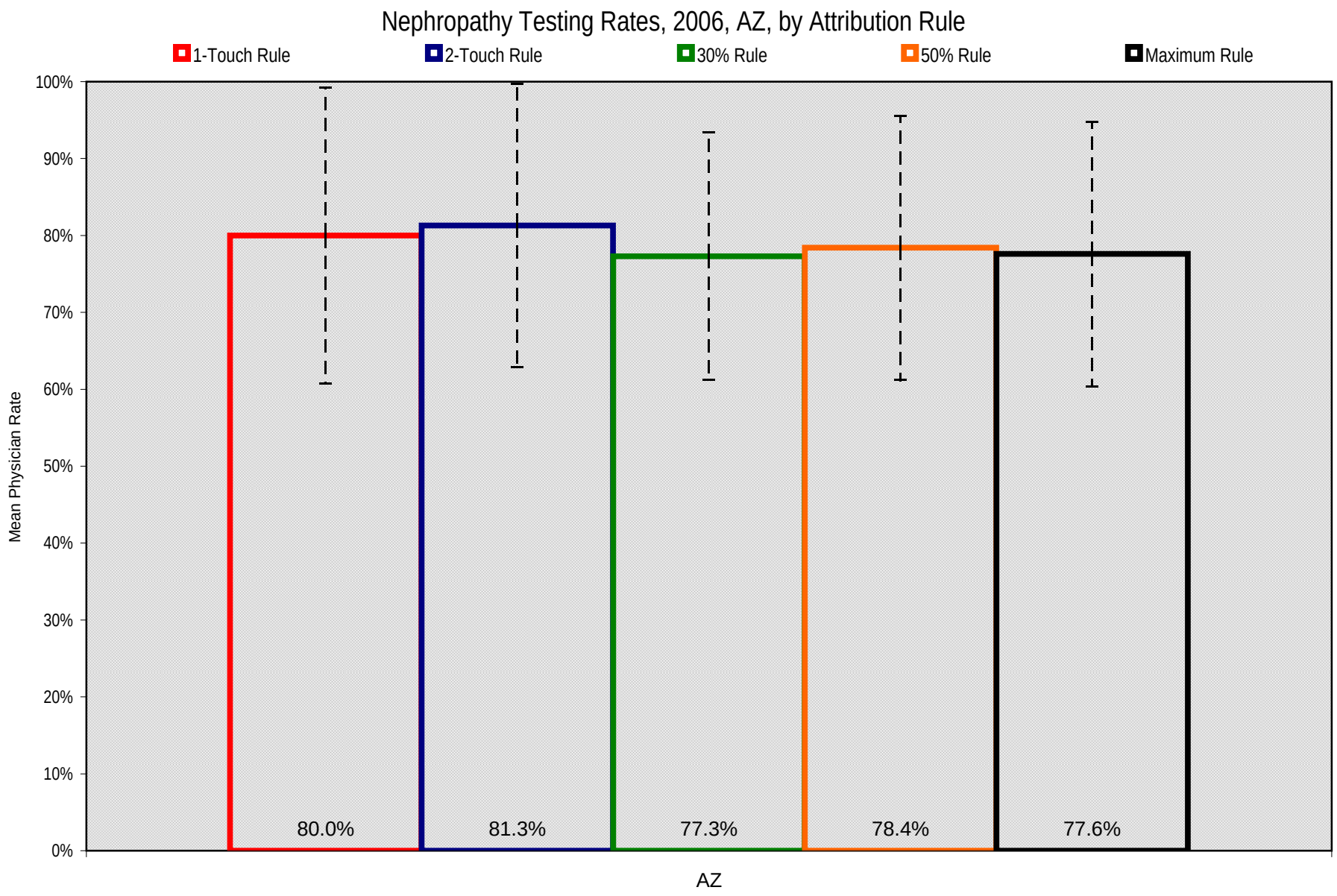
† "% eligible physicians w/ rate" is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

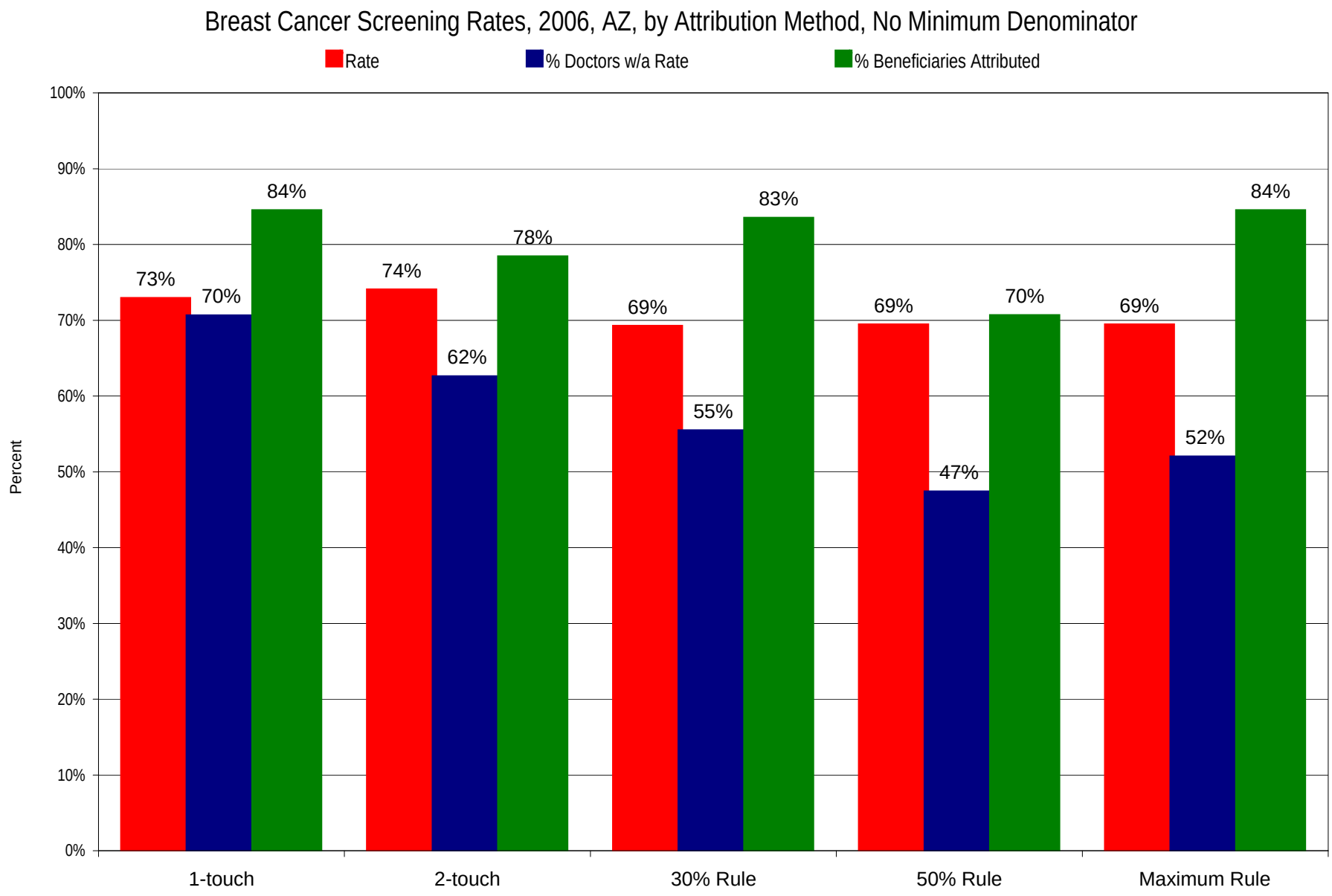
‡ "% beneficiaries in denominator w/ a physician" is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

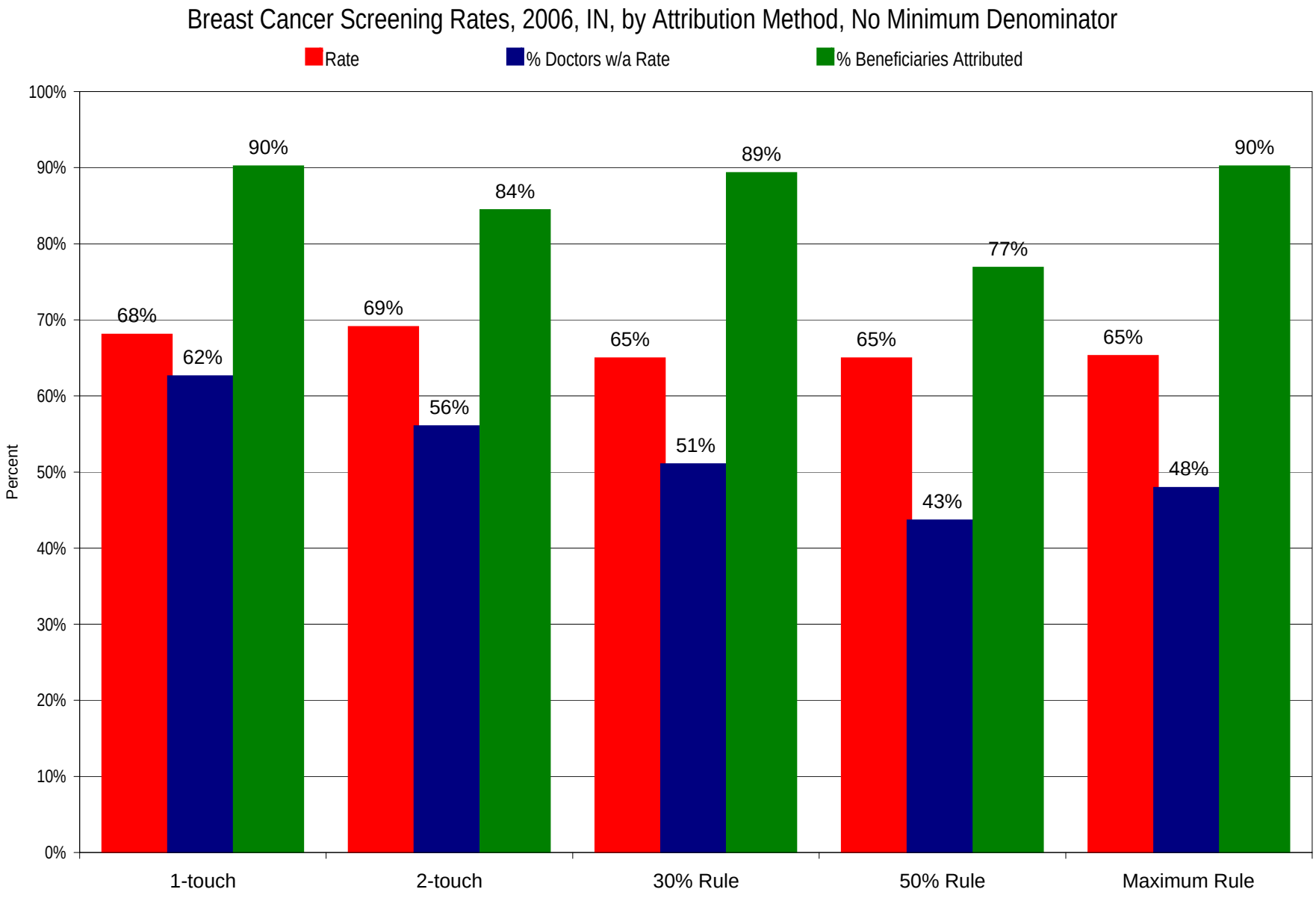

Colorectal Cancer Screening Rates, 2006, IN, by Attribution Rule
1-Touch Rule
2-Touch Rule
30% Rule
50% Rule
Maximum Rule
Mean Physician Rate
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
50.5%
51.1%
44.5%
43.3%
44.2%
IN

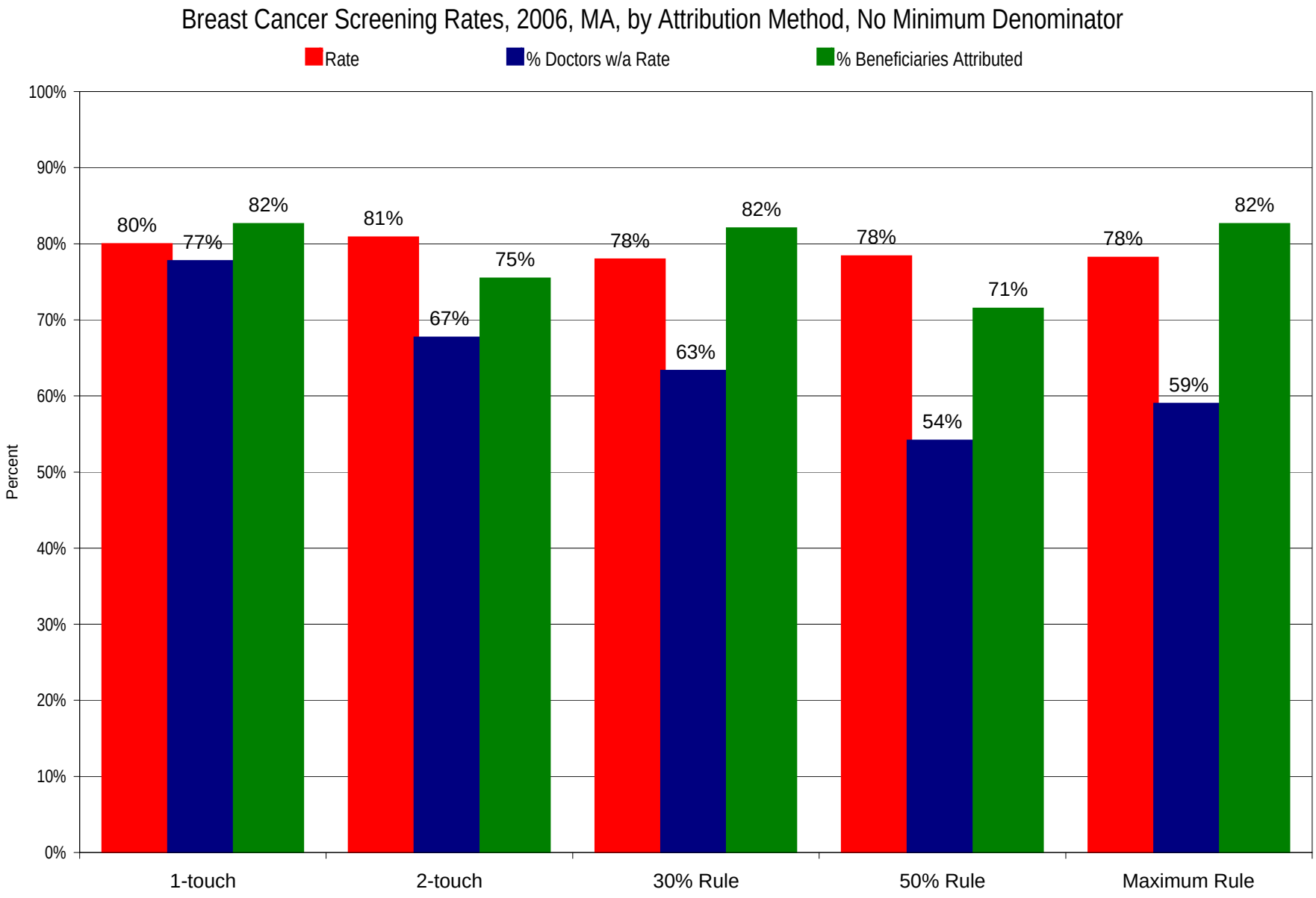

HbA1c Testing Rates, 2006, AZ, by Attribution Rule
1-Touch Rule
2-Touch Rule
30% Rule
50% Rule
Maximum Rule
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Mean Physician Rate
82.9%
84.1%
80.8%
81.6%
81.0%
AZ

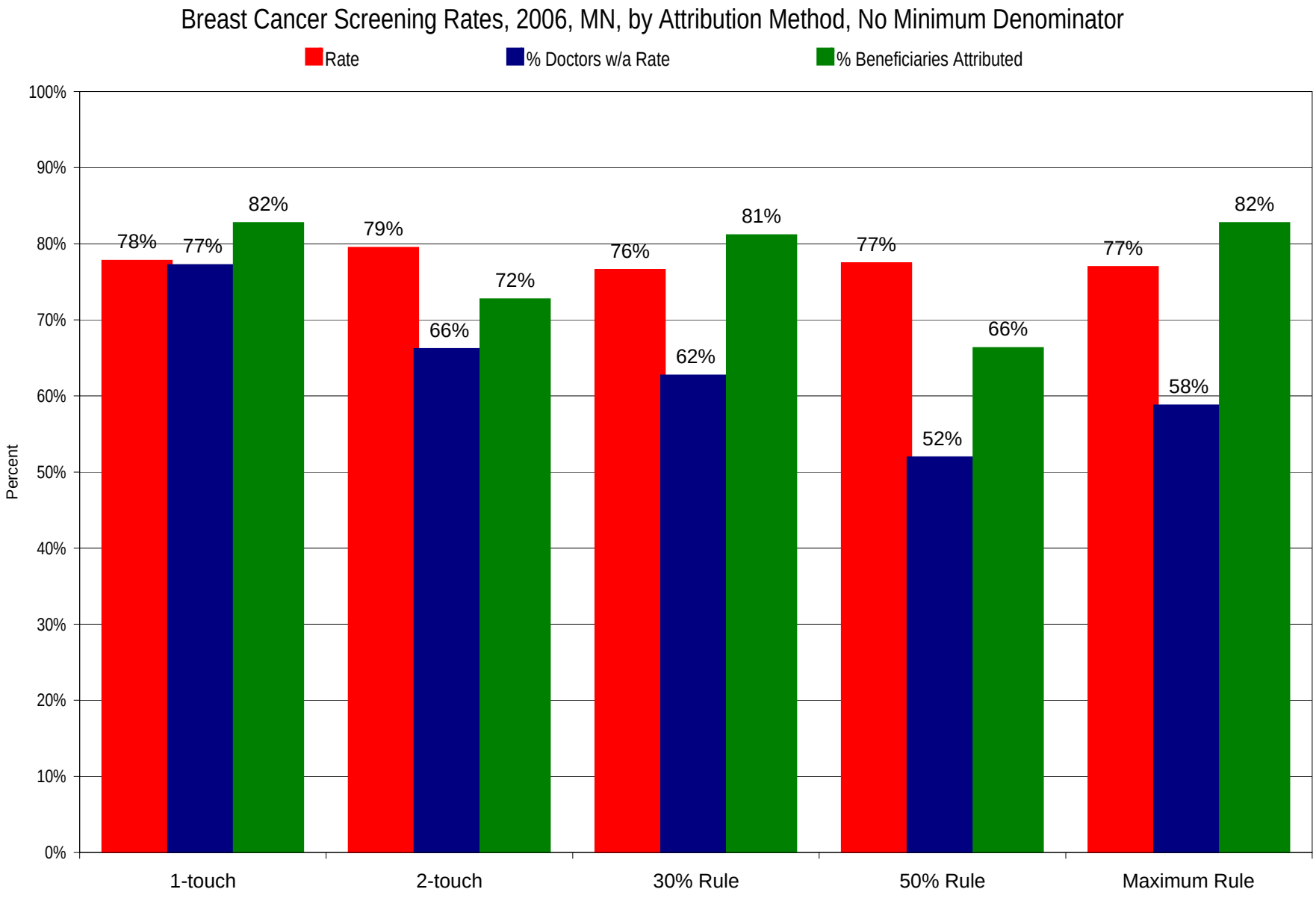

LDL Testing Rates, 2006, AZ, by Attribution Rule
1-Touch Rule
2-Touch Rule
30% Rule
50% Rule
Maximum Rule
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Mean Physician Rate
74.8%
75.2%
71.6%
71.5%
71.6%
AZ

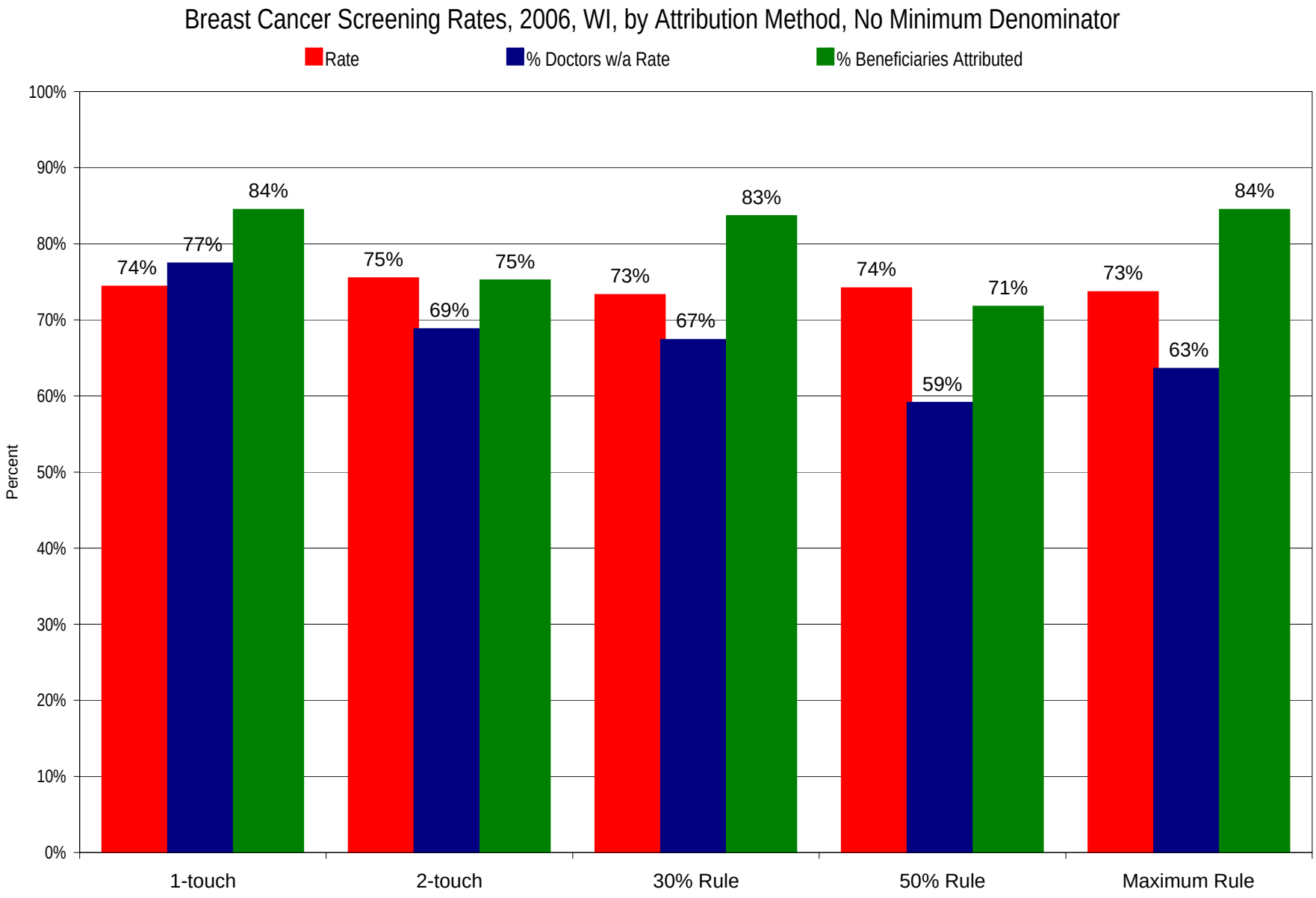

Nephropathy Testing Rates, 2006, AZ, by Attribution Rule
1-Touch Rule
2-Touch Rule
30% Rule
50% Rule
Maximum Rule
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Mean Physician Rate
80.0%
81.3%
77.3%
78.4%
77.6%
AZ

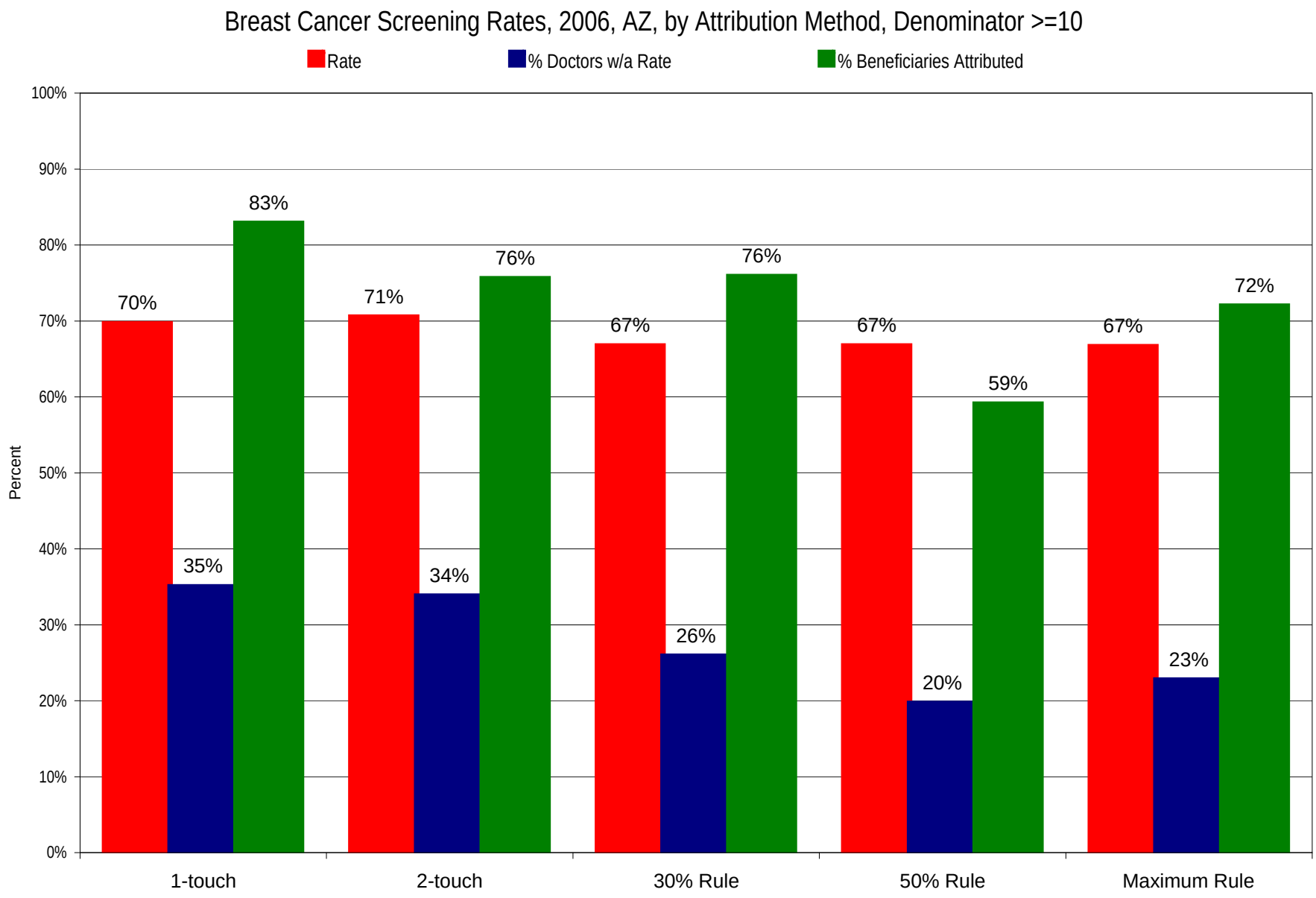

Breast Cancer Screening Rates, 2006, AZ, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
73%
70%
84%
74%
62%
78%
69%
55%
83%
69%
47%
70%
69%
52%
84%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

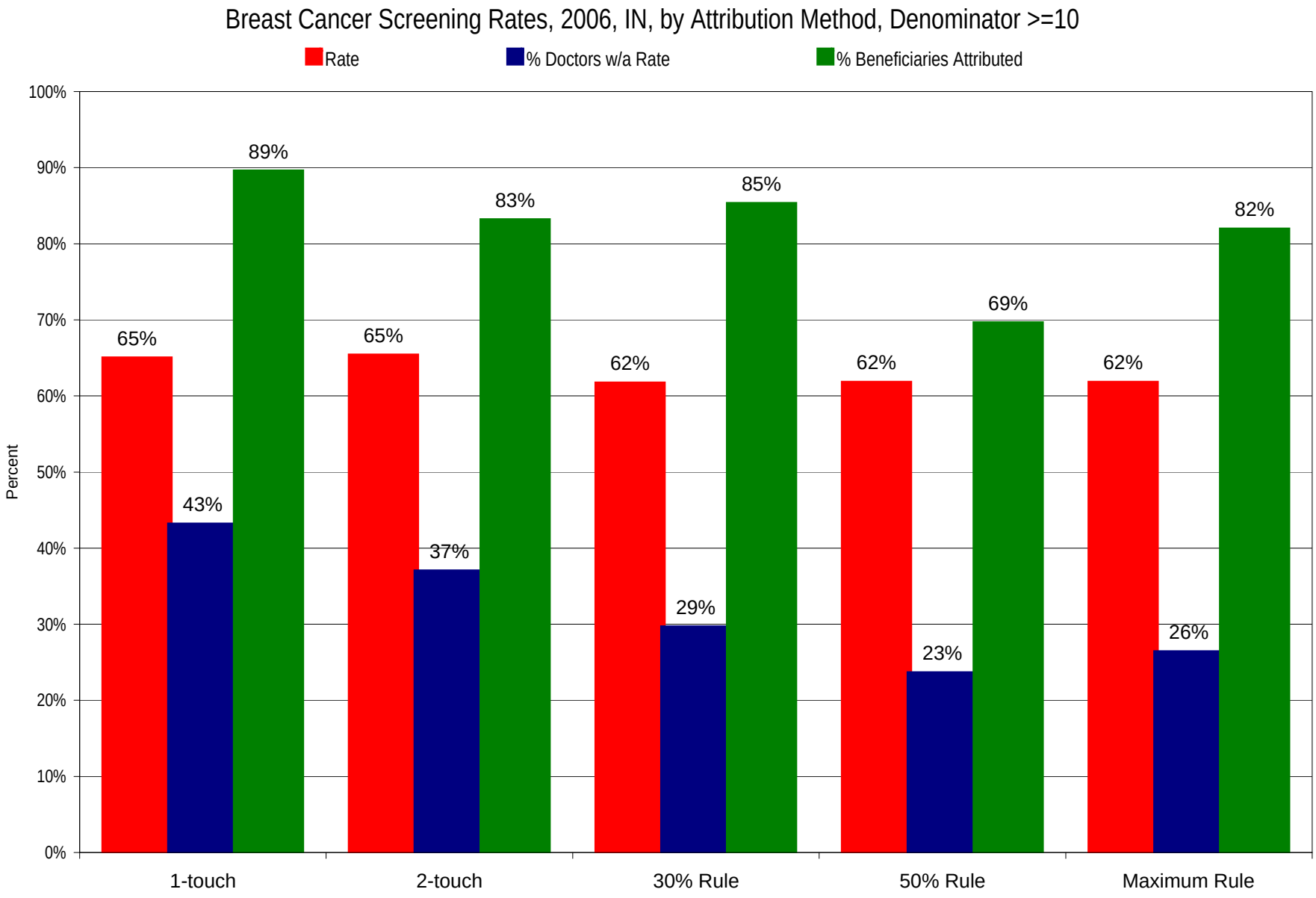

Breast Cancer Screening Rates, 2006, IN, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
68%
62%
90%
69%
56%
84%
65%
51%
89%
65%
43%
77%
65%
48%
90%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

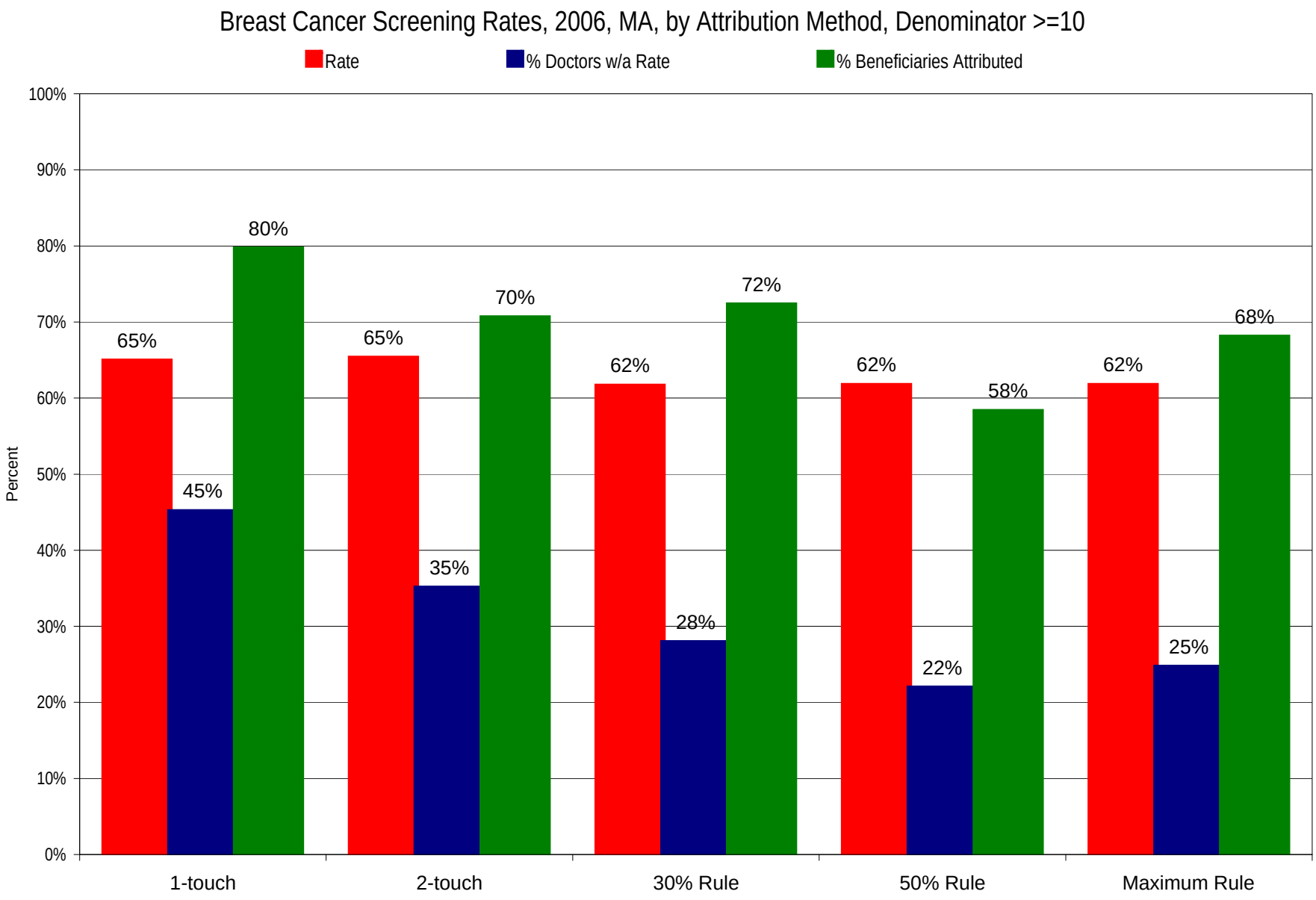

Breast Cancer Screening Rates, 2006, MA, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
80%
77%
82%
81%
67%
75%
78%
63%
82%
78%
54%
71%
78%
59%
82%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

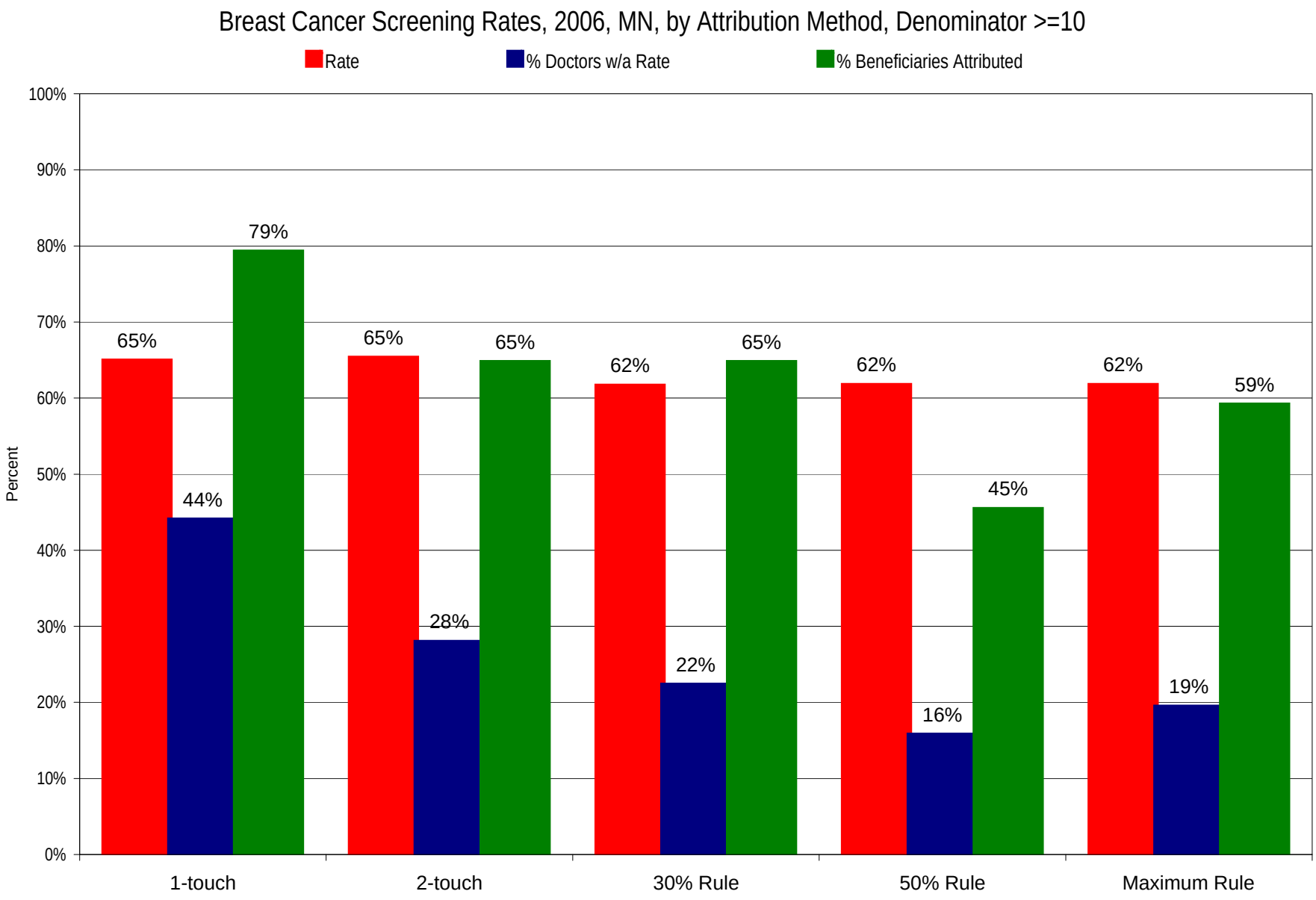

Breast Cancer Screening Rates, 2006, MN, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
78%
77%
82%
79%
66%
72%
76%
62%
81%
77%
52%
66%
77%
58%
82%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

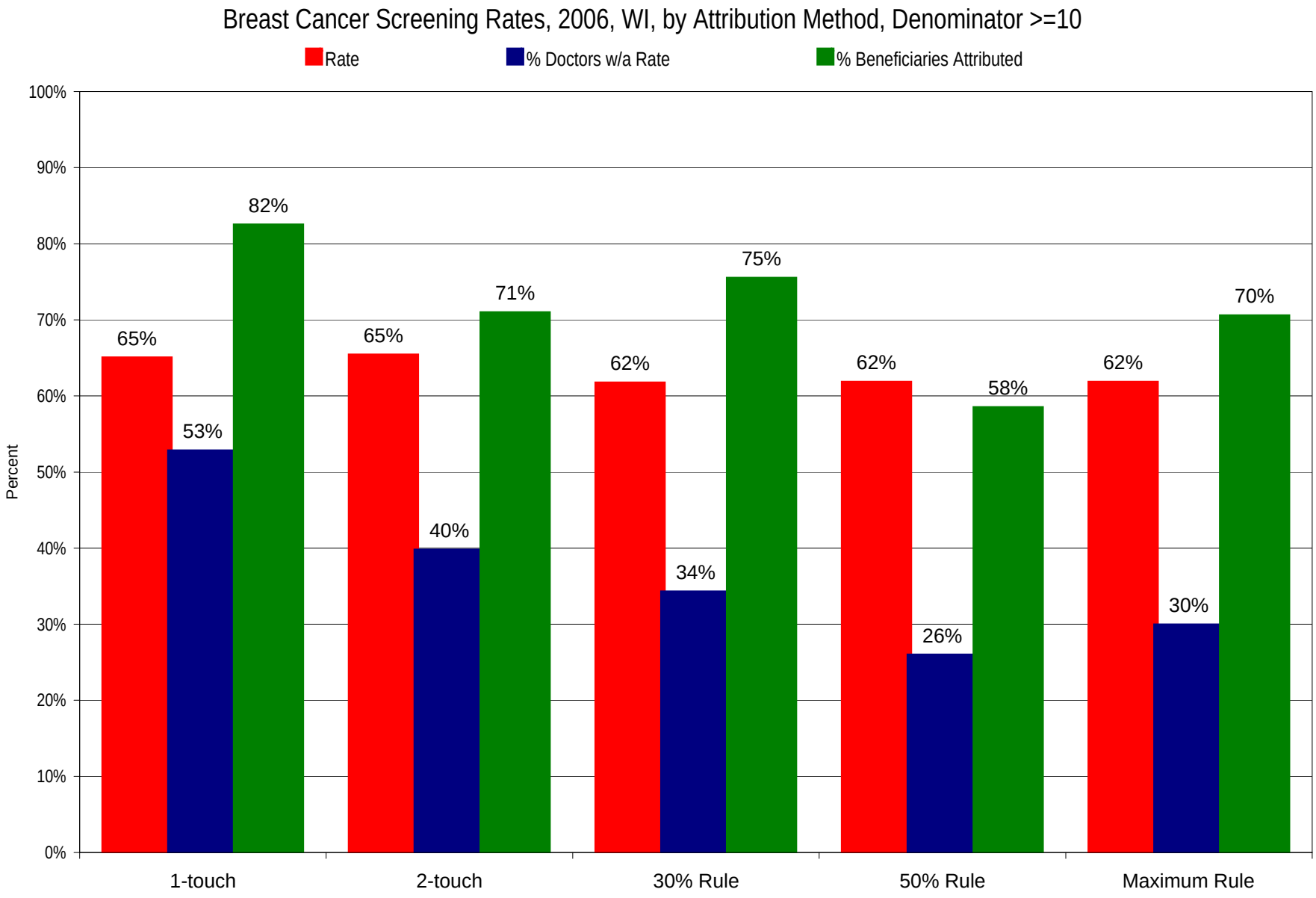

Breast Cancer Screening Rates, 2006, WI, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
74%
77%
84%
75%
69%
75%
73%
67%
83%
74%
59%
71%
73%
63%
84%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

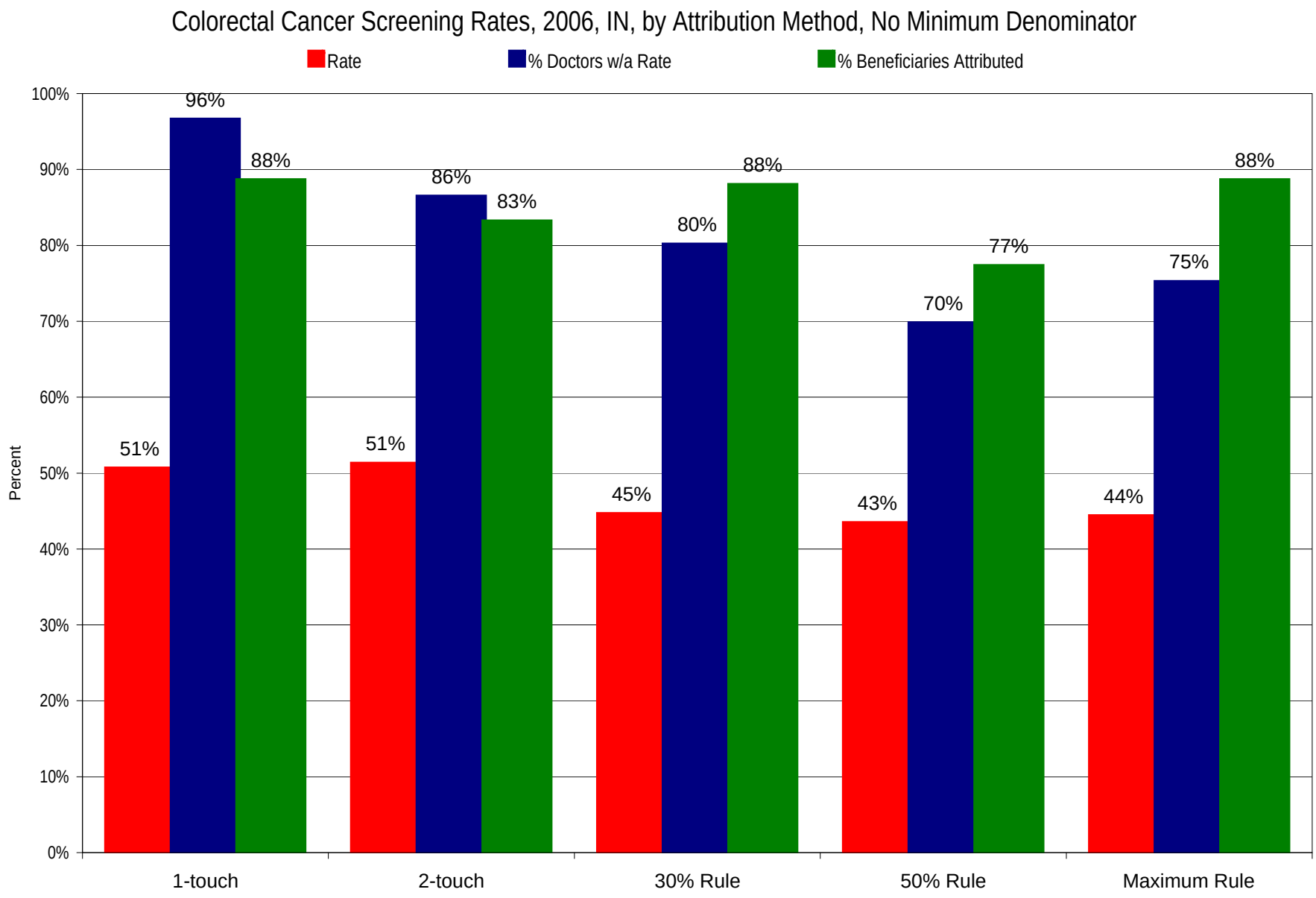

Breast Cancer Screening Rates, 2006, AZ, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
70%
35%
83%
71%
34%
76%
67%
26%
76%
67%
20%
59%
67%
23%
72%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

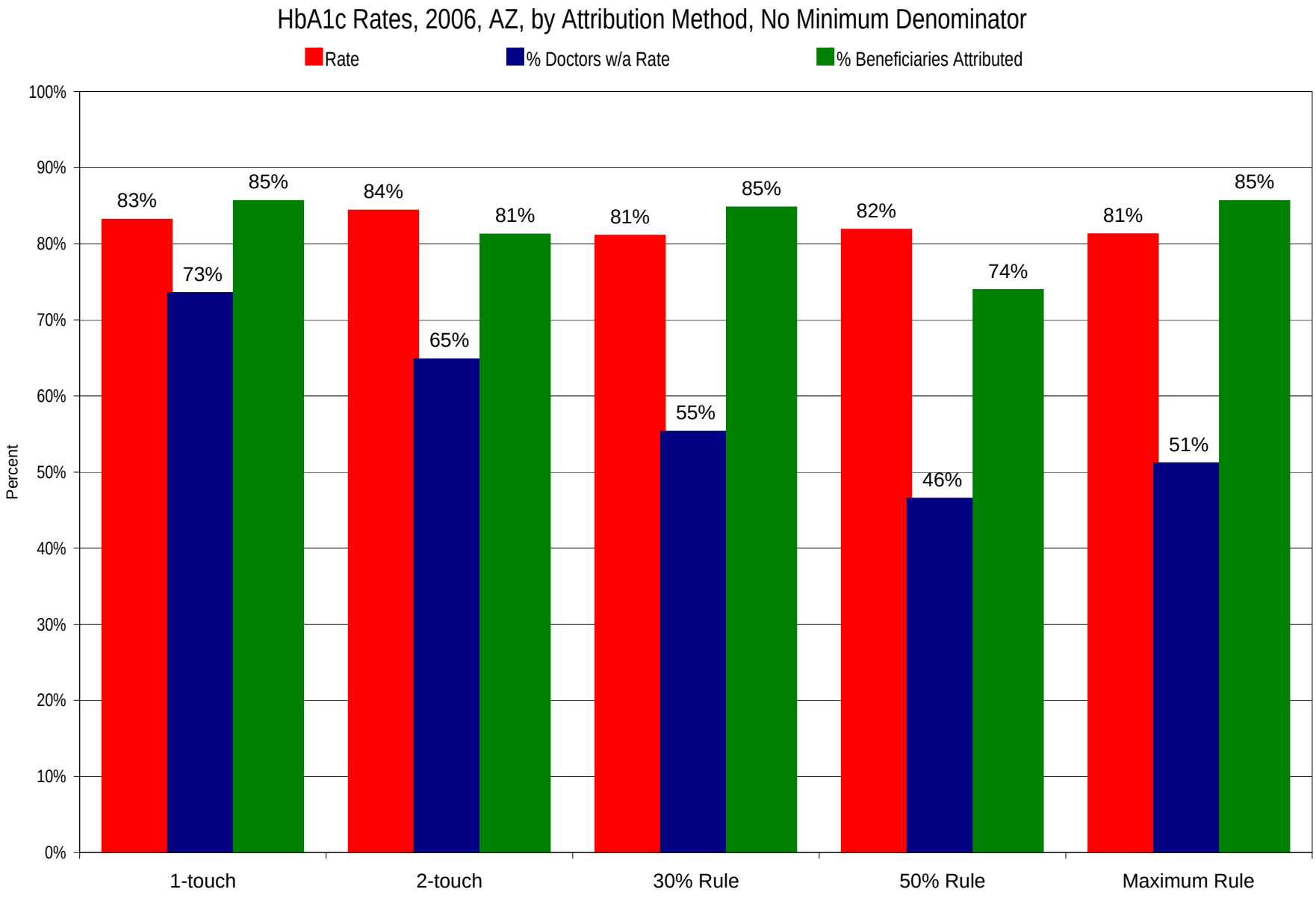

Breast Cancer Screening Rates, 2006, IN, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
65%
43%
89%
65%
37%
83%
62%
29%
85%
62%
23%
69%
62%
26%
82%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

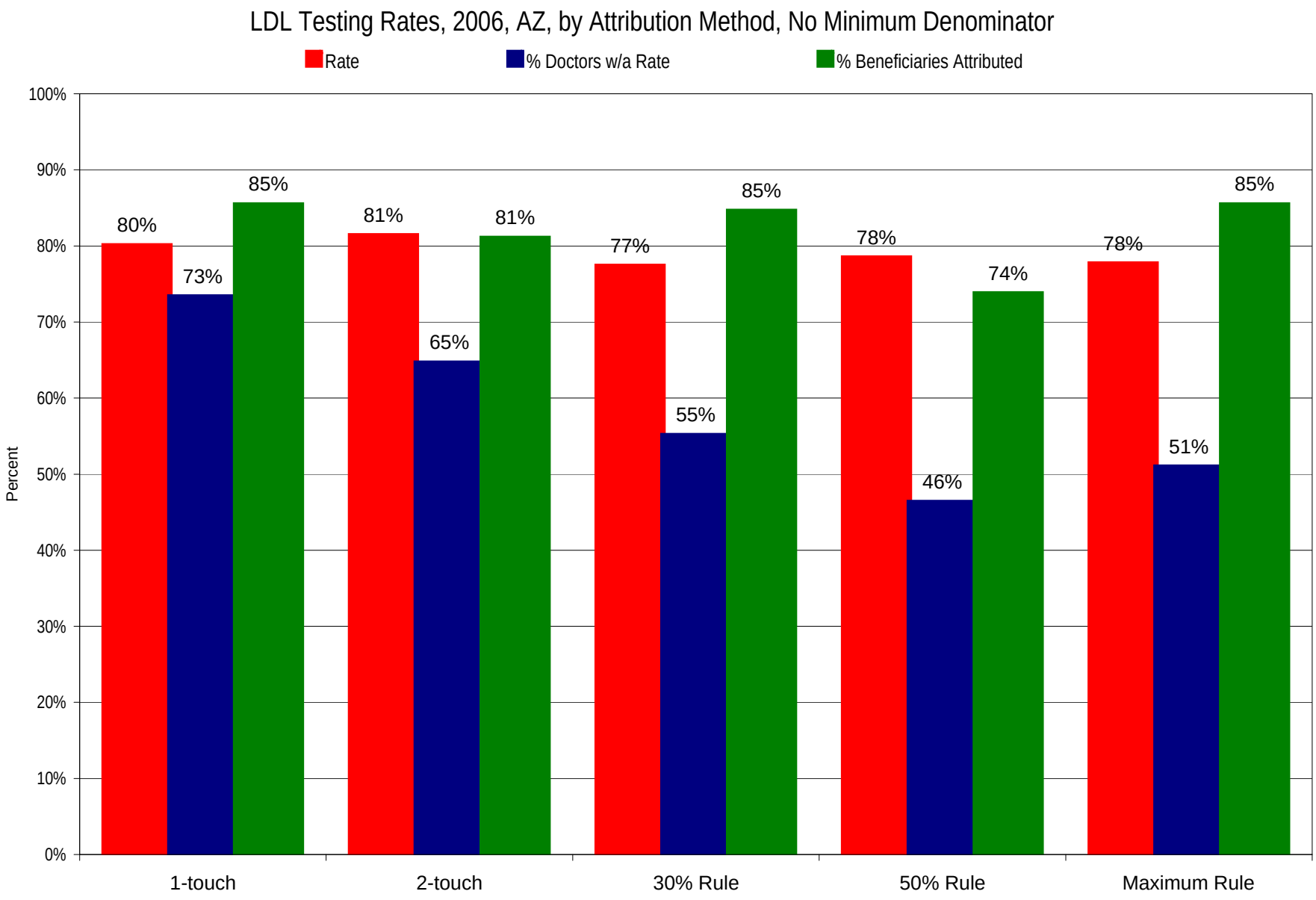

Breast Cancer Screening Rates, 2006, MA, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
65%
45%
80%
65%
35%
70%
62%
28%
72%
62%
22%
58%
62%
25%
68%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

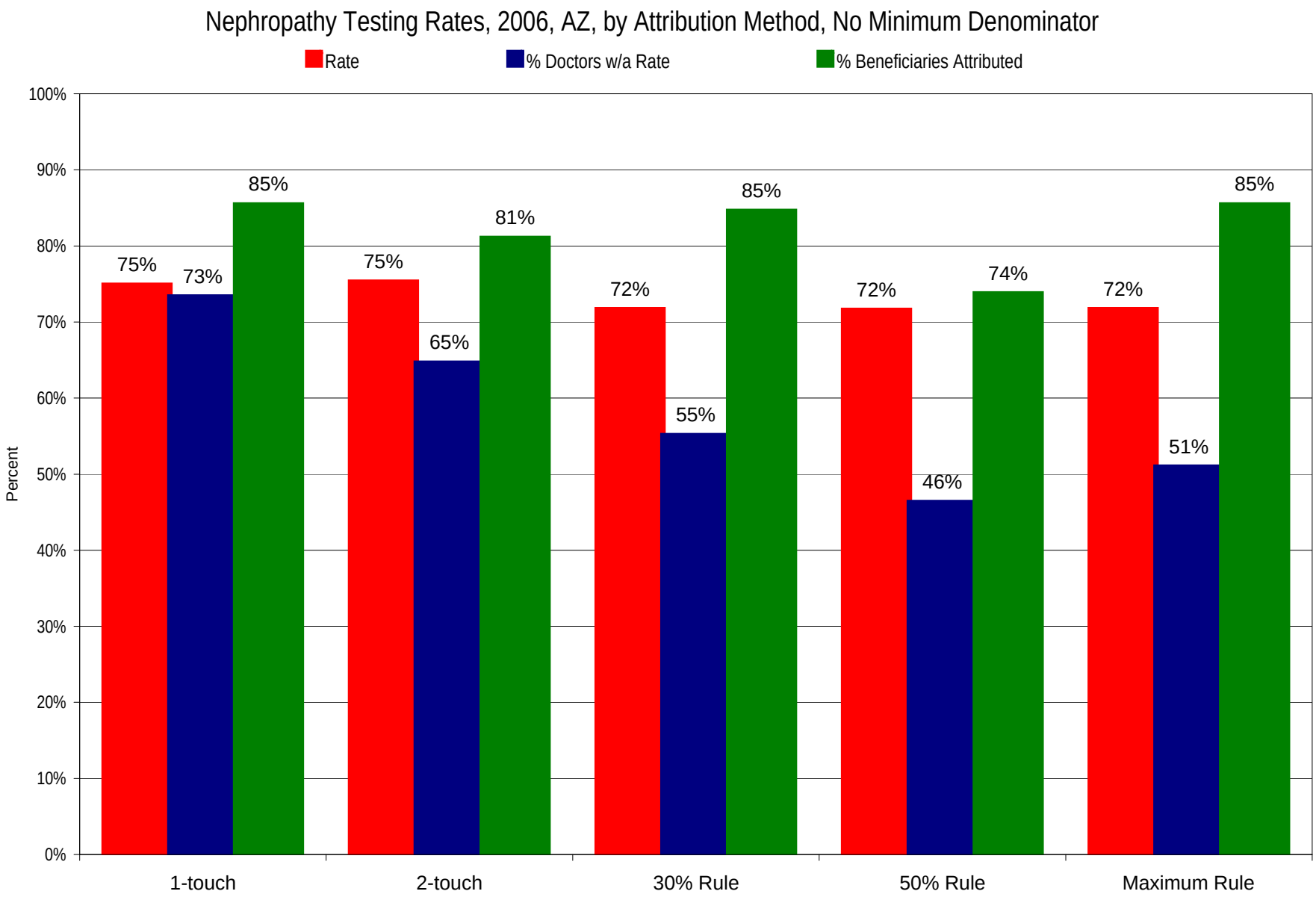

Breast Cancer Screening Rates, 2006, MN, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
65%
44%
79%
65%
28%
65%
62%
22%
65%
62%
16%
45%
62%
19%
59%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

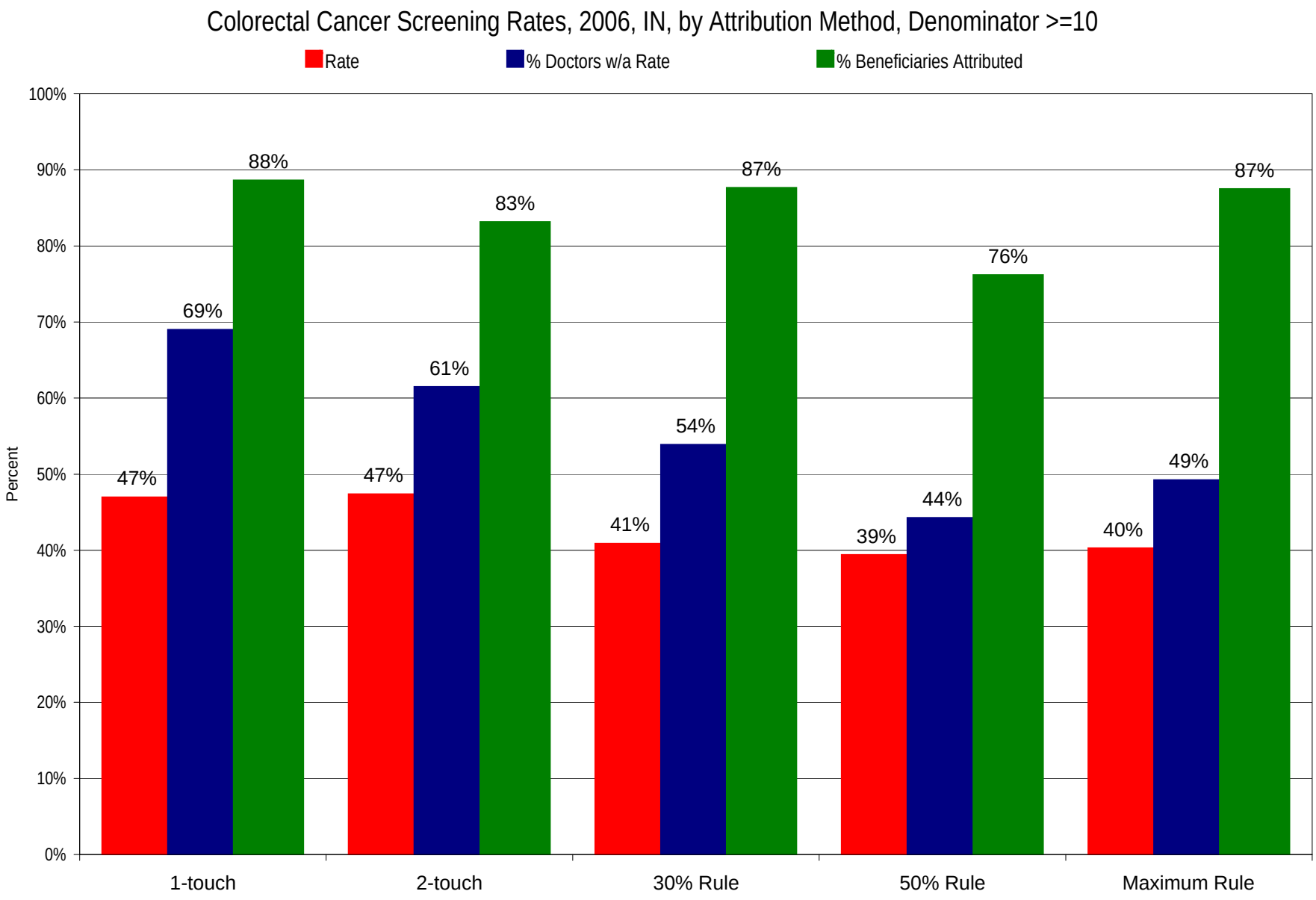

Breast Cancer Screening Rates, 2006, WI, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
65%
53%
82%
65%
40%
71%
62%
34%
75%
62%
26%
58%
62%
30%
70%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

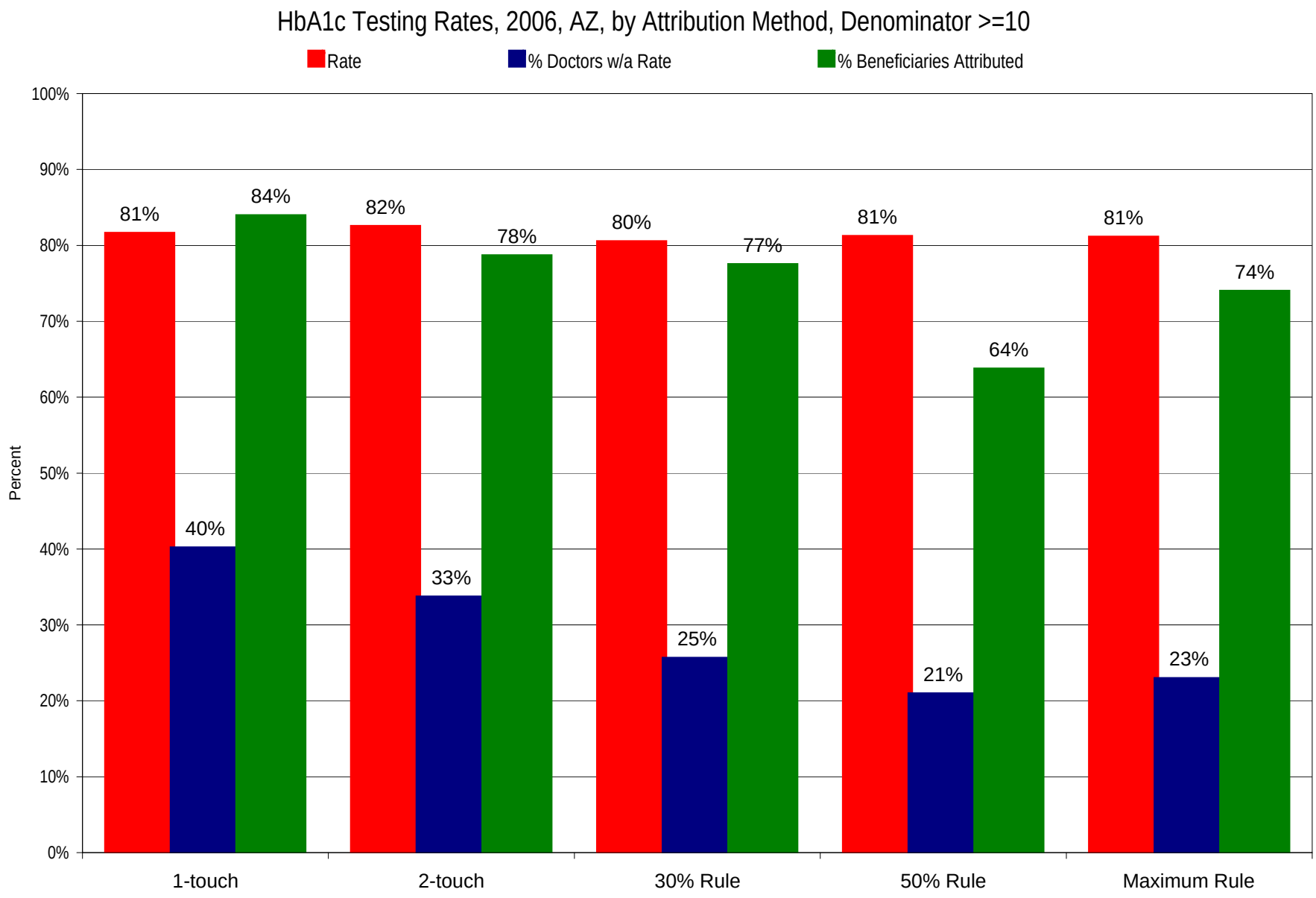

Colorectal Cancer Screening Rates, 2006, IN, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
51%
96%
88%
51%
86%
83%
45%
80%
88%
43%
70%
77%
44%
75%
88%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

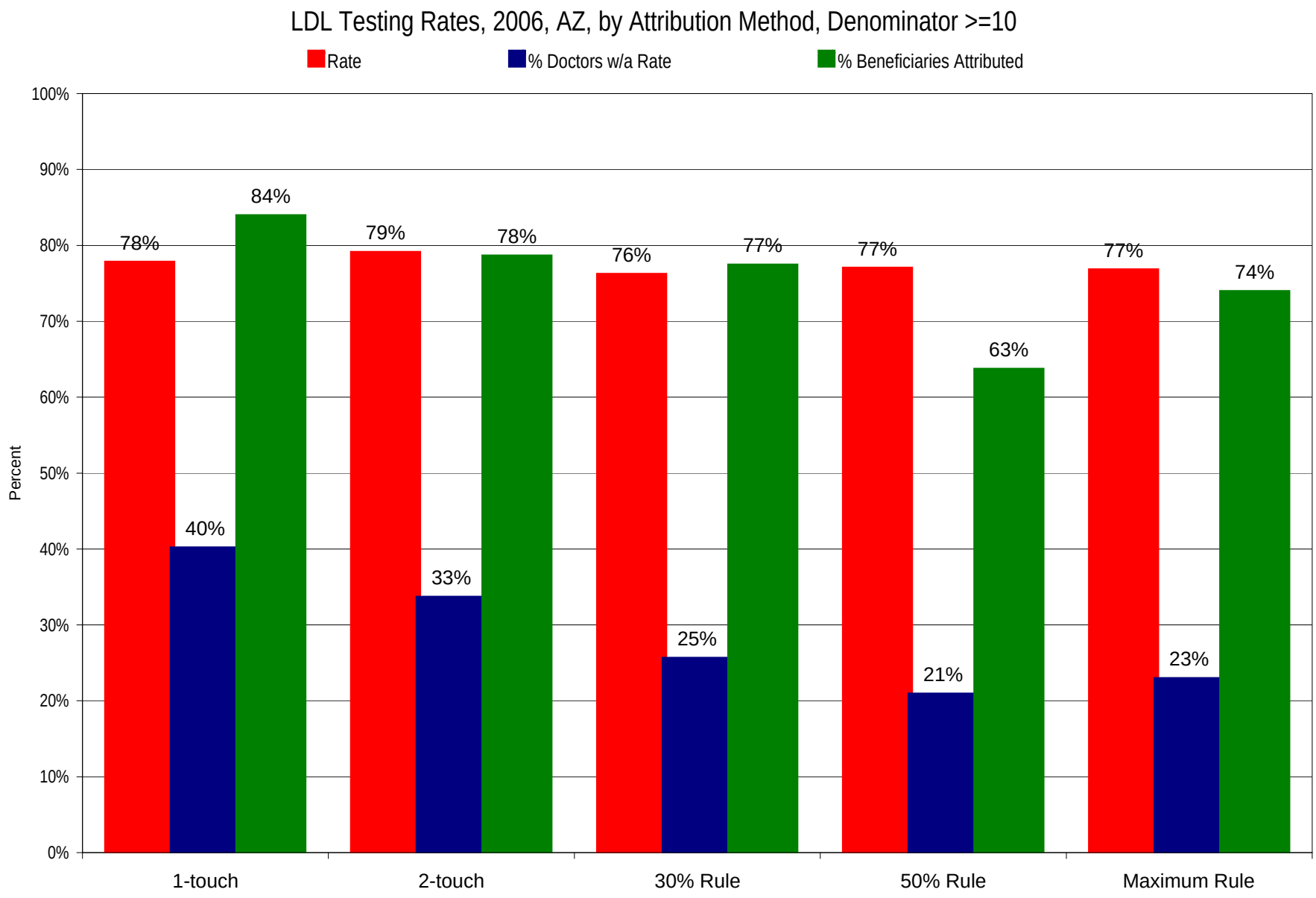

HbA1c Rates, 2006, AZ, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
83%
73%
85%
84%
65%
81%
81%
55%
85%
82%
46%
74%
81%
51%
85%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

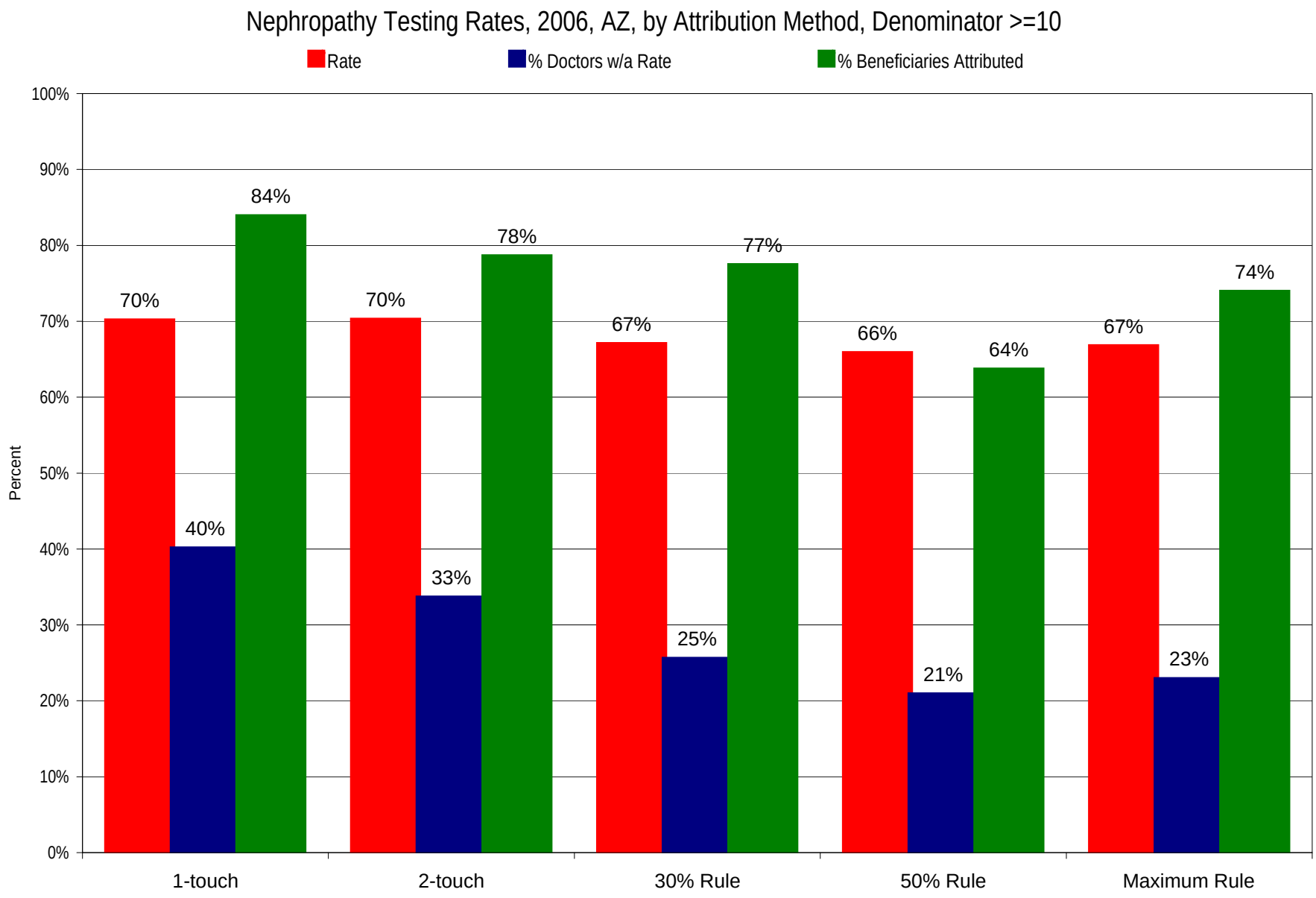

LDL Testing Rates, 2006, AZ, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
80%
73%
85%
81%
65%
81%
77%
55%
85%
78%
46%
74%
78%
51%
85%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

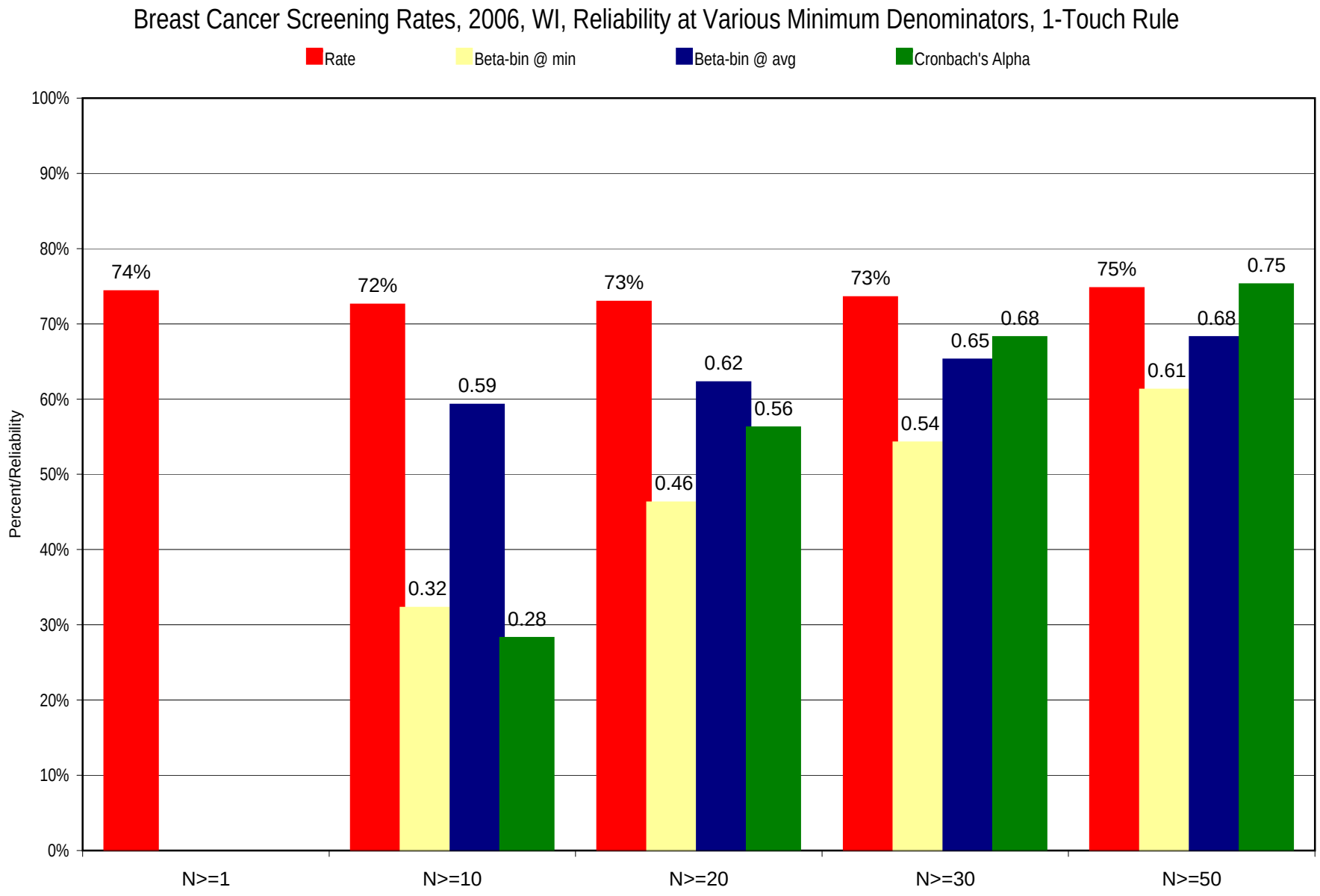

Nephropathy Testing Rates, 2006, AZ, by Attribution Method, No Minimum Denominator
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
75%
73%
85%
75%
65%
81%
72%
55%
85%
72%
46%
74%
72%
51%
85%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

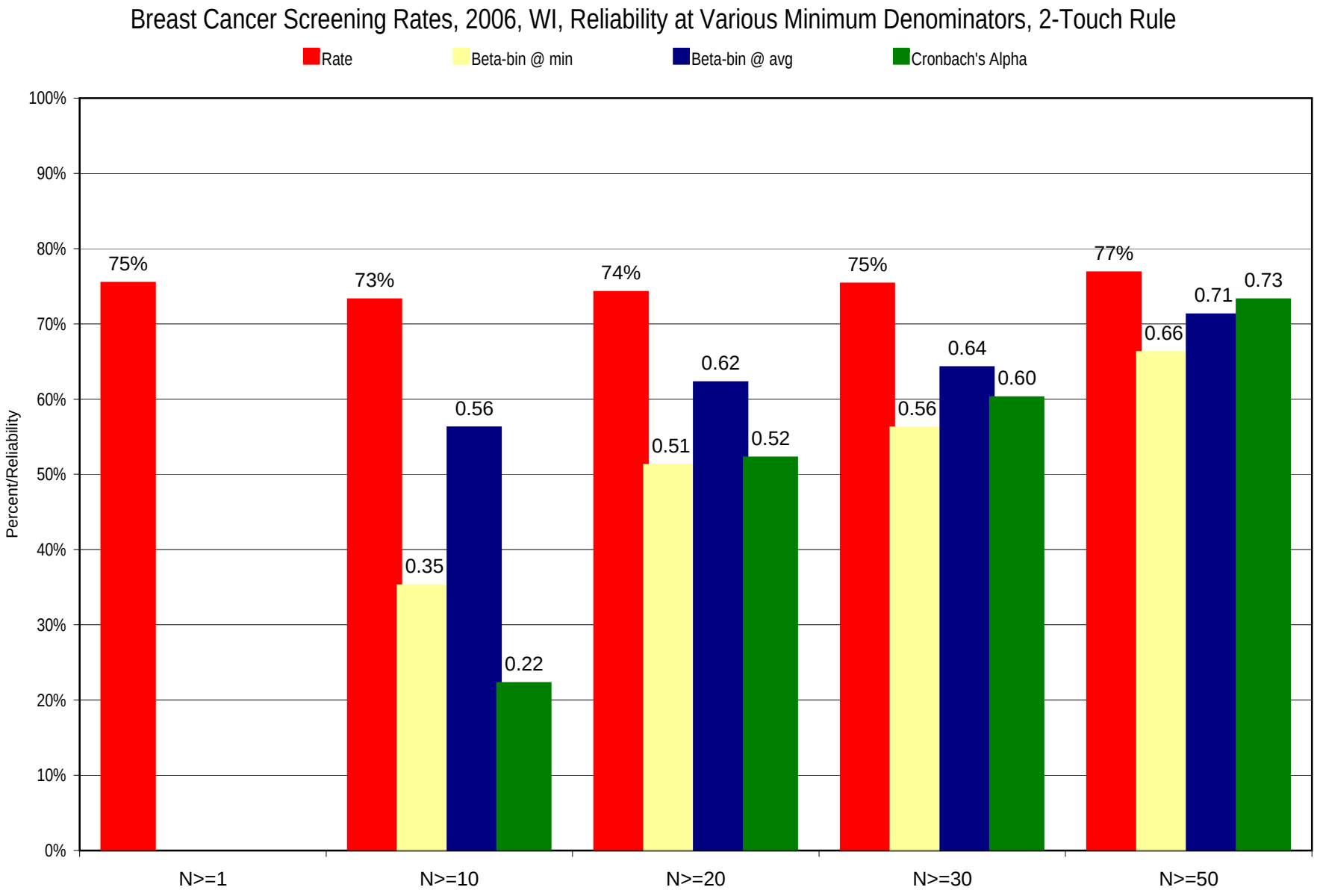

Colorectal Cancer Screening Rates, 2006, IN, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
47%
69%
88%
47%
61%
83%
41%
54%
87%
39%
44%
76%
40%
49%
87%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

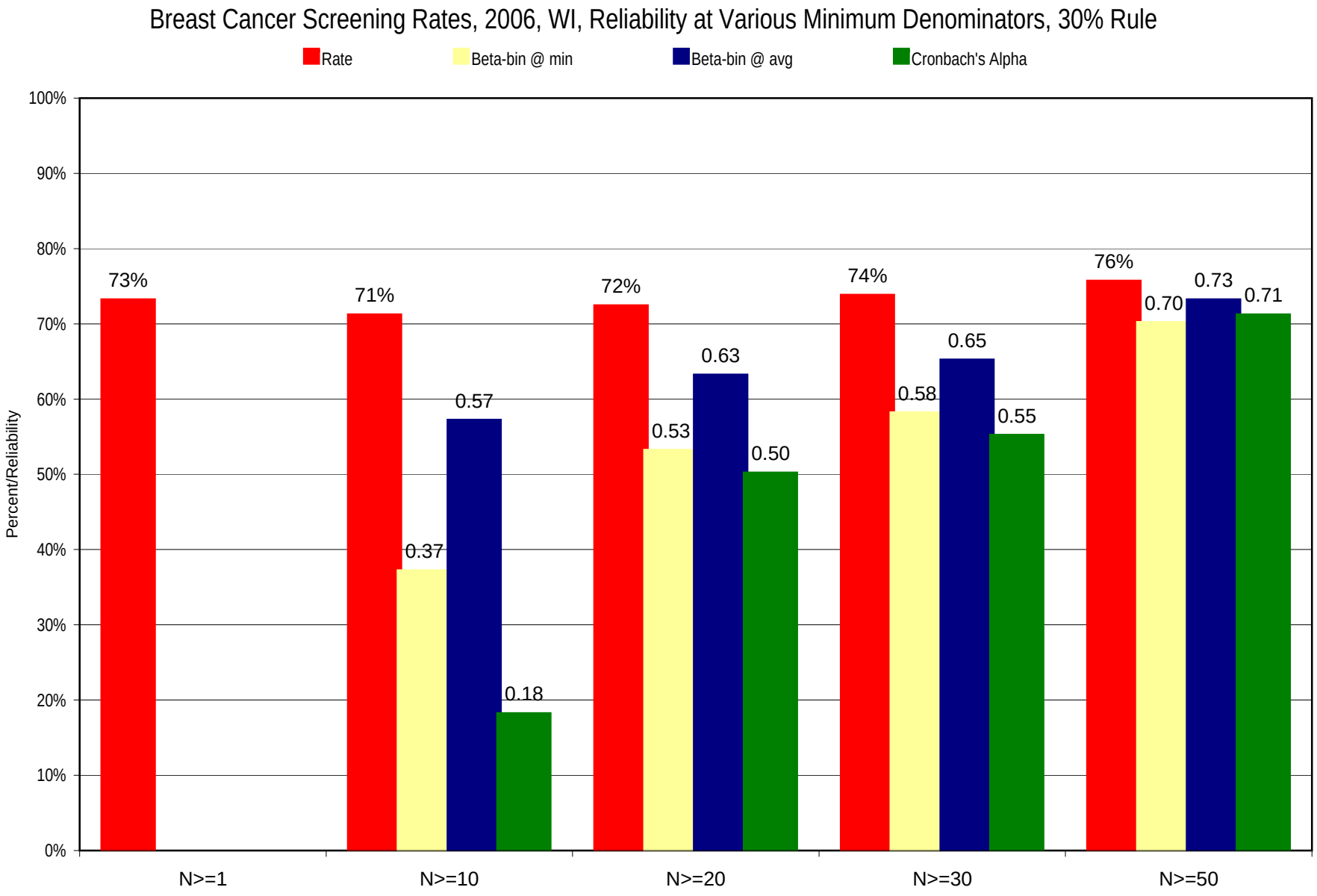

HbA1c Testing Rates, 2006, AZ, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
81%
40%
84%
82%
33%
78%
80%
25%
77%
81%
21%
64%
81%
23%
74%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

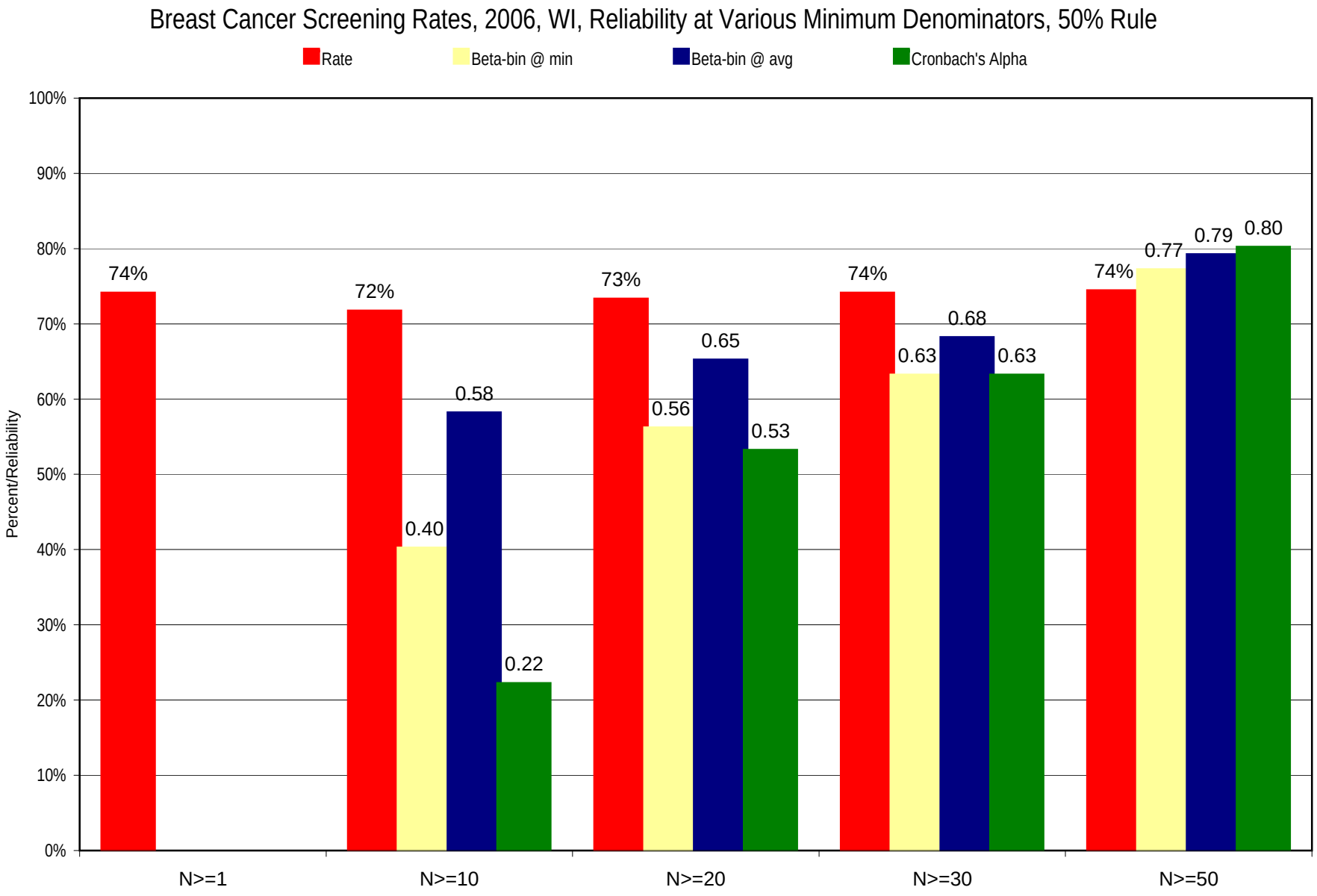

LDL Testing Rates, 2006, AZ, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
78%
40%
84%
79%
33%
78%
76%
25%
77%
77%
21%
63%
77%
23%
74%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

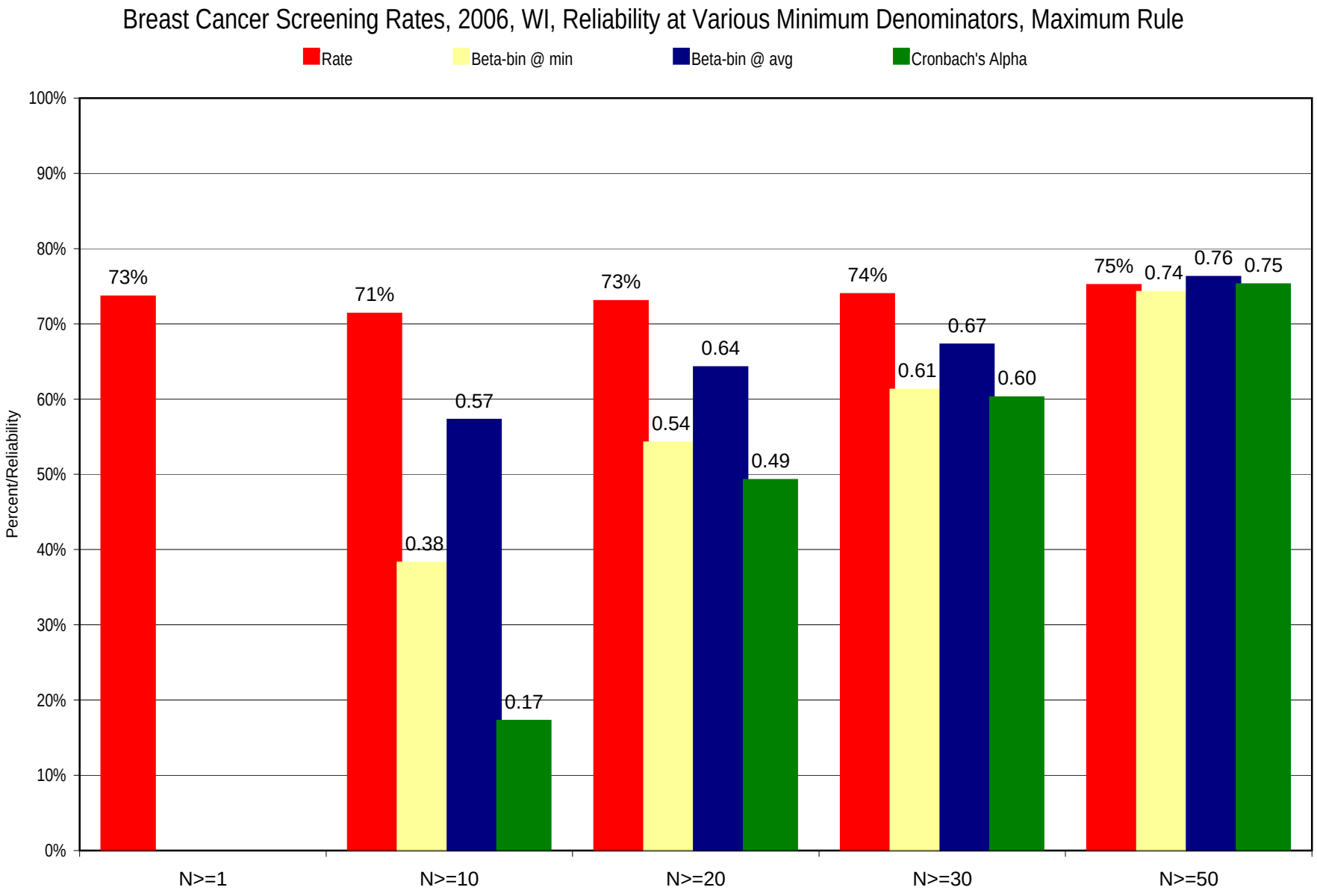

Nephropathy Testing Rates, 2006, AZ, by Attribution Method, Denominator >=10
Rate
% Doctors w/a Rate
% Beneficiaries Attributed
Percent
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
70%
40%
84%
70%
33%
78%
67%
25%
77%
66%
21%
64%
67%
23%
74%
1-touch
2-touch
30% Rule
50% Rule
Maximum Rule

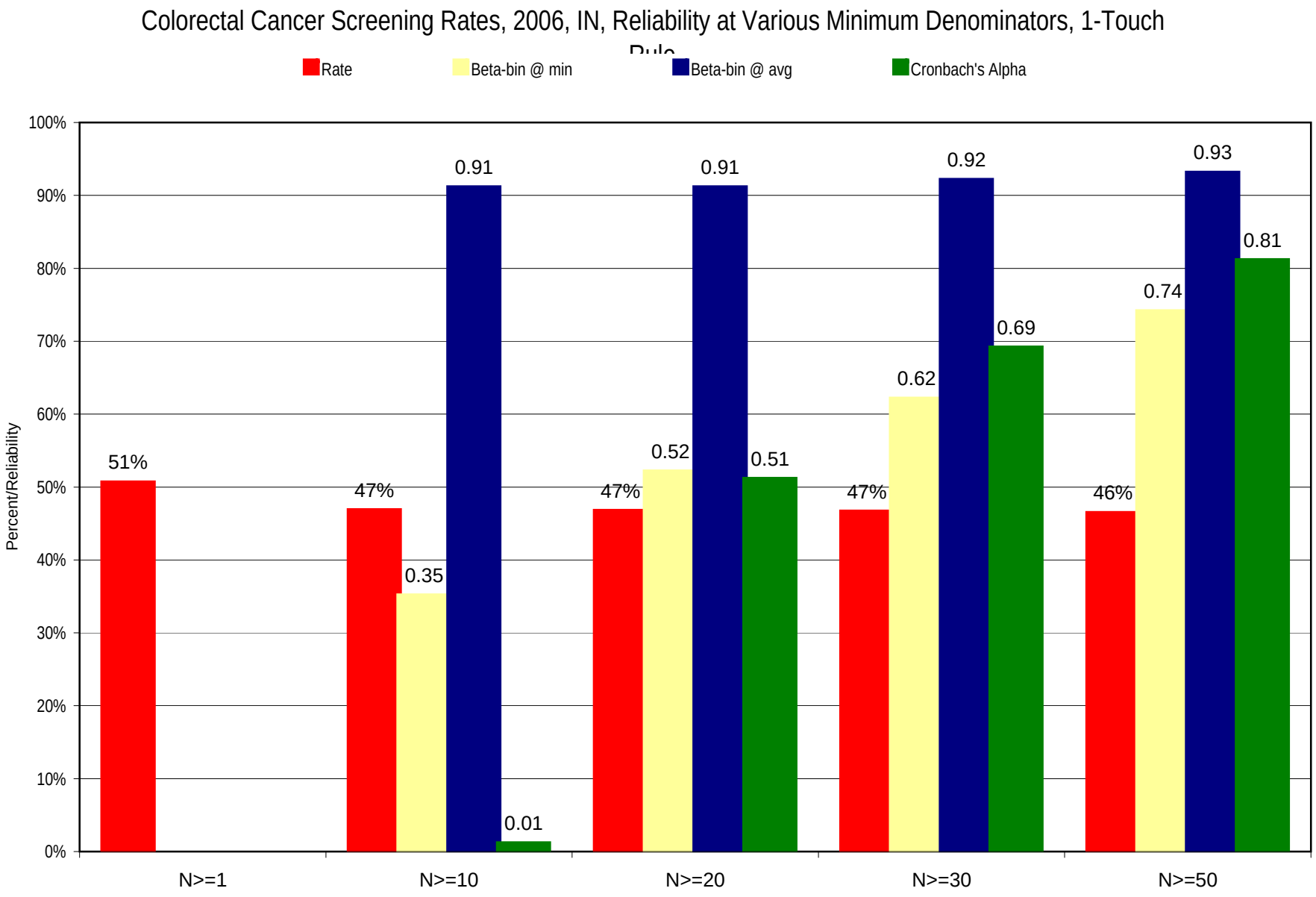

Breast Cancer Screening Rates, 2006, WI, Reliability at Various Minimum Denominators, 1-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
74%
72%
0.32
0.59
0.28
73%
0.46
0.62
0.56
73%
0.54
0.65
0.68
75%
0.61
0.68
0.75
N>=1
N>=10
N>=20
N>=30
N>=50

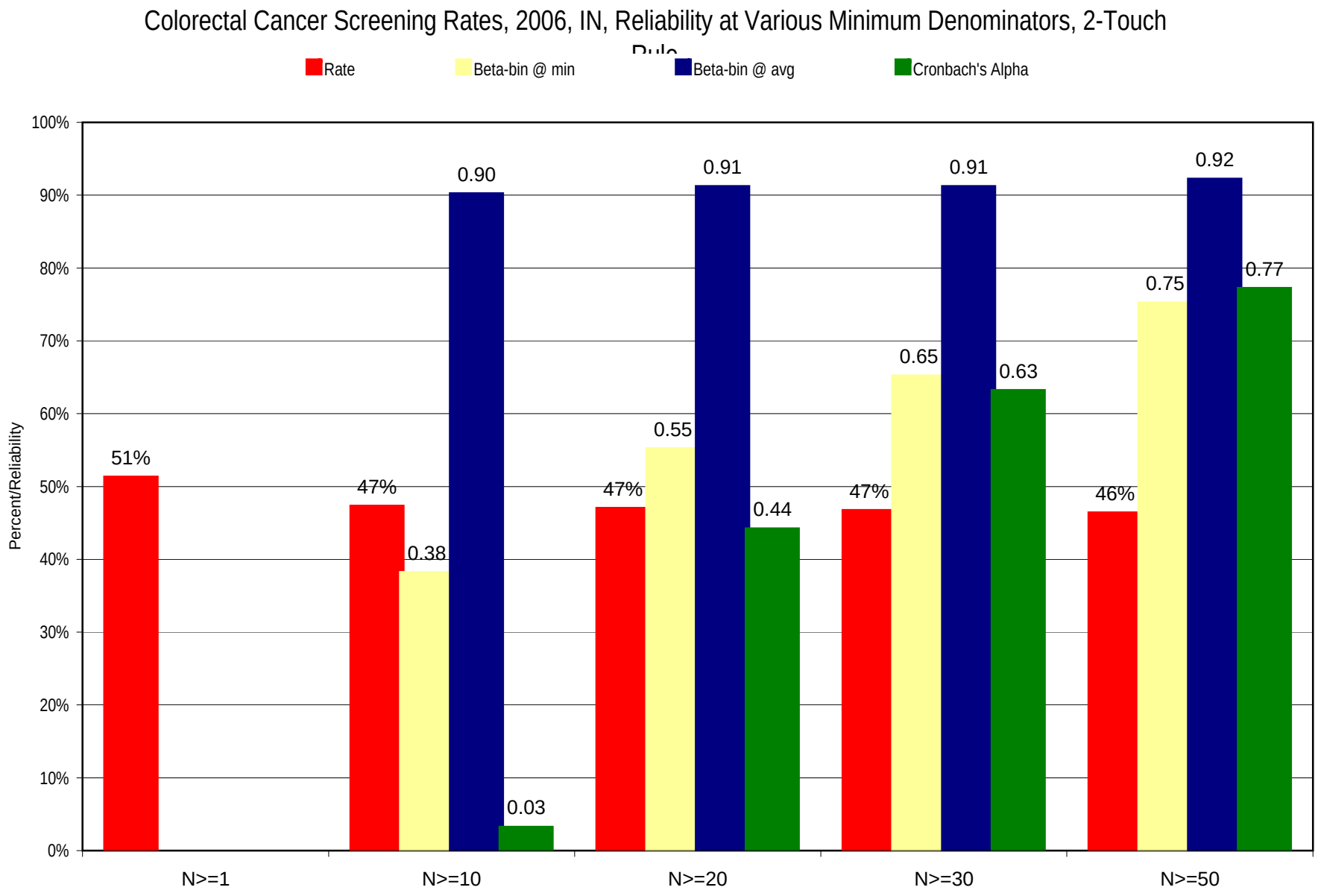

Breast Cancer Screening Rates, 2006, WI, Reliability at Various Minimum Denominators, 2-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
75%
73%
0.35
0.56
0.22
74%
0.51
0.62
0.52
75%
0.56
0.64
0.60
77%
0.66
0.71
0.73
N>=1
N>=10
N>=20
N>=30
N>=50

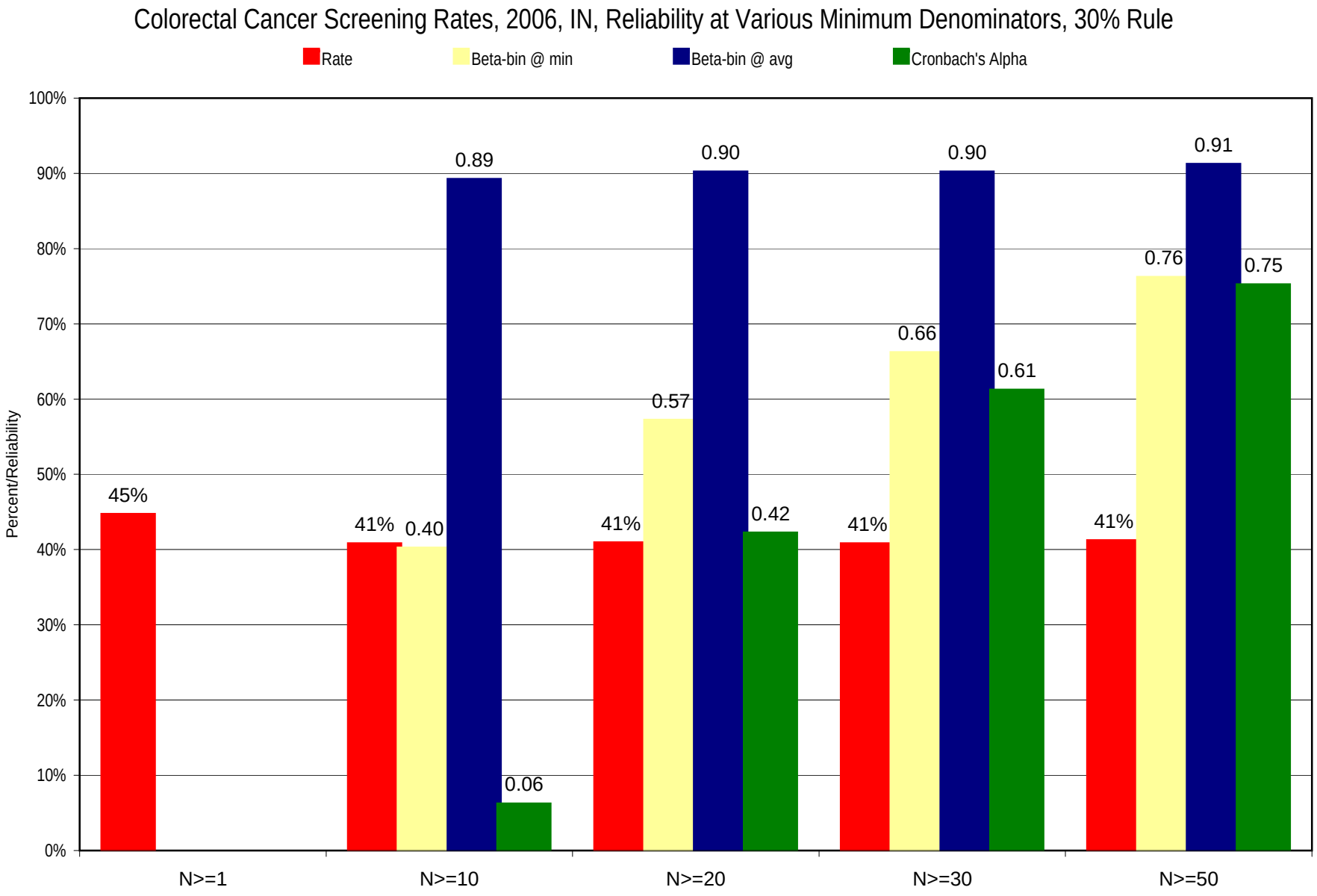

Breast Cancer Screening Rates, 2006, WI, Reliability at Various Minimum Denominators, 30% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
73%
71%
0.37
0.57
0.18
72%
0.53
0.63
0.50
74%
0.58
0.65
0.55
76%
0.70
0.73
0.71
N>=1
N>=10
N>=20
N>=30
N>=50

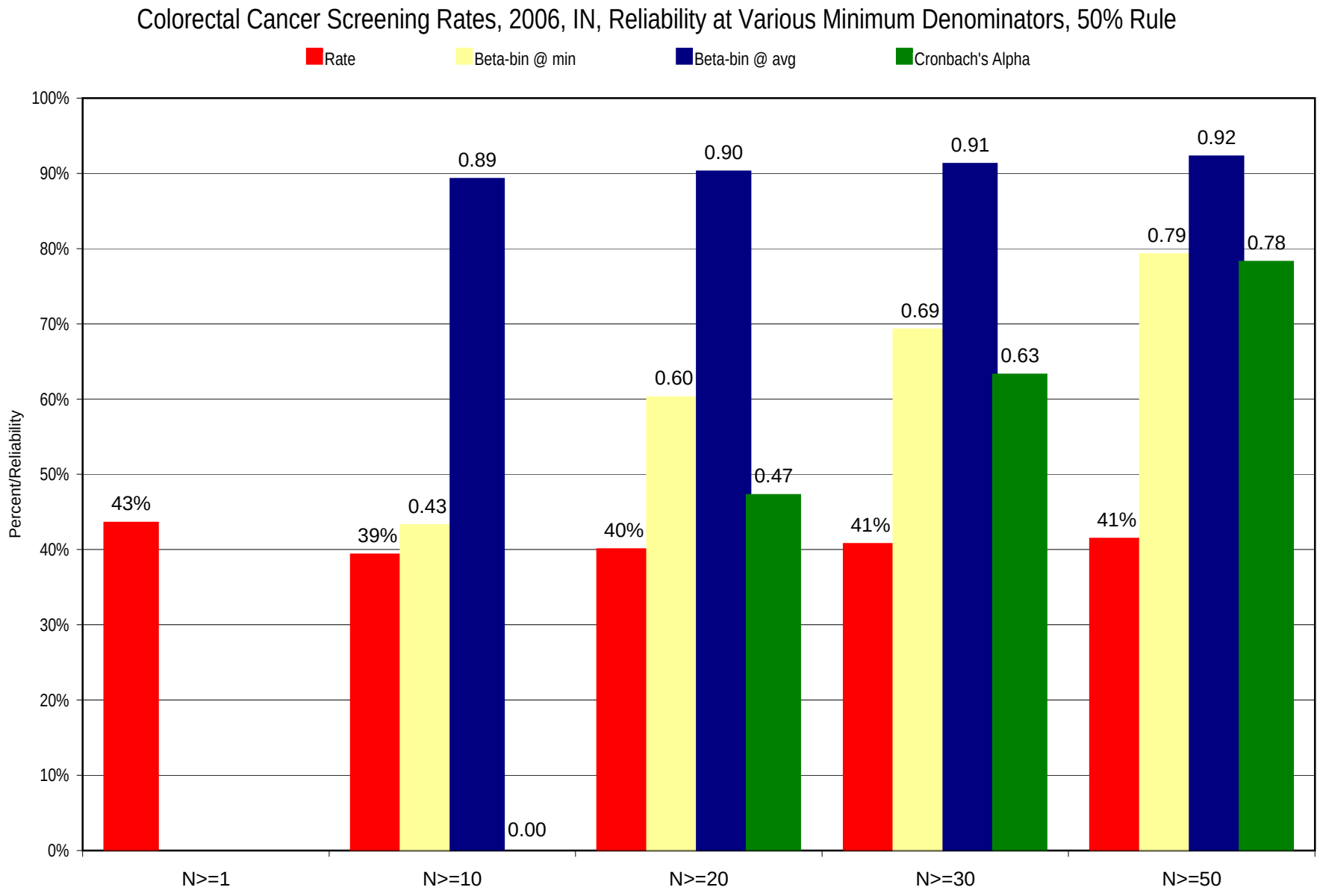

Breast Cancer Screening Rates, 2006, WI, Reliability at Various Minimum Denominators, 50% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
74%
72%
0.40
0.58
0.22
73%
0.56
0.65
0.53
74%
0.63
0.68
0.63
74%
0.77
0.79
0.80
N>=1
N>=10
N>=20
N>=30
N>=50

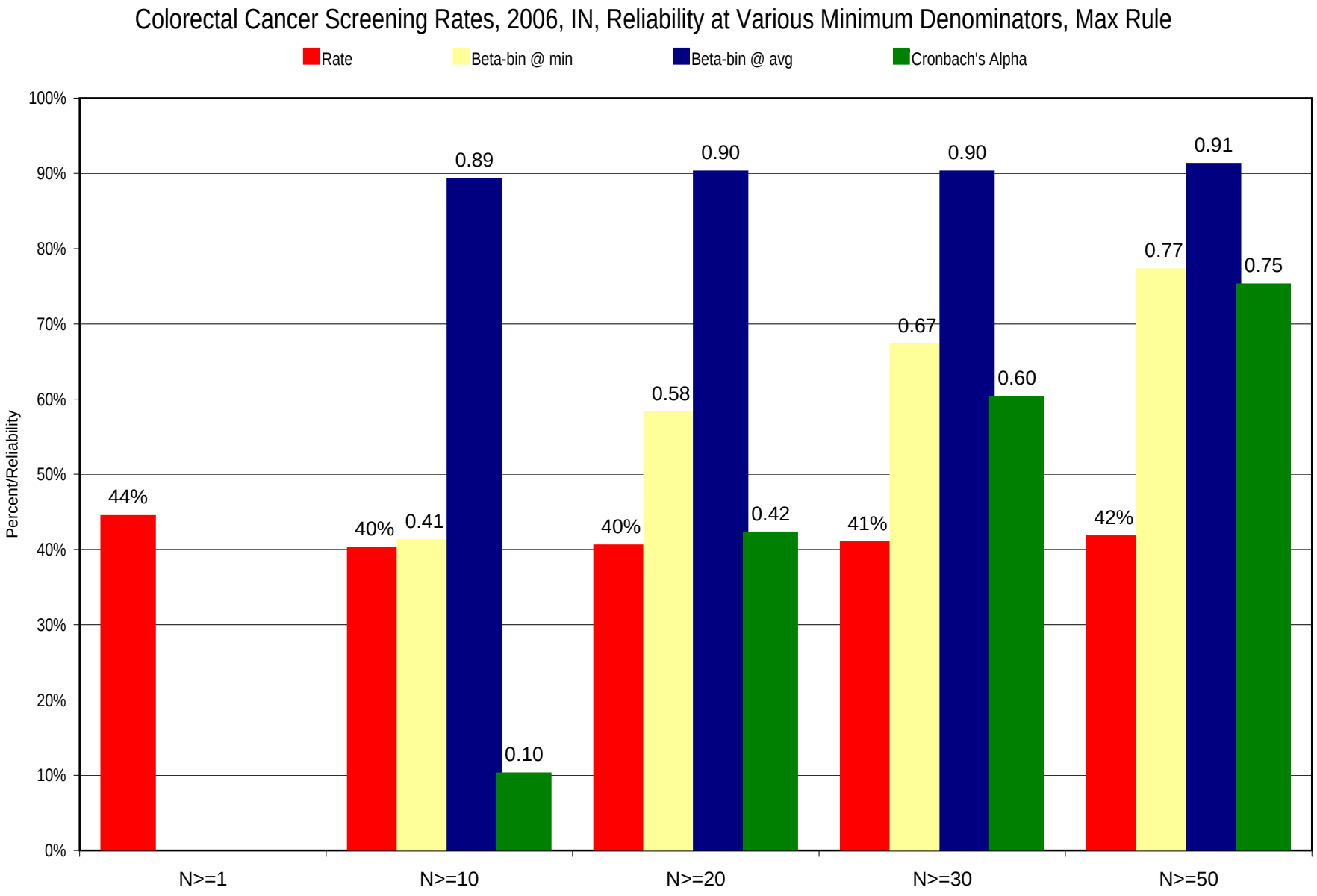

Breast Cancer Screening Rates, 2006, WI, Reliability at Various Minimum Denominators, Maximum Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
73%
71%
0.38
0.57
0.17
73%
0.54
0.64
0.49
74%
0.61
0.67
0.60
75%
0.74
0.76
0.75
N>=1
N>=10
N>=20
N>=30
N>=50

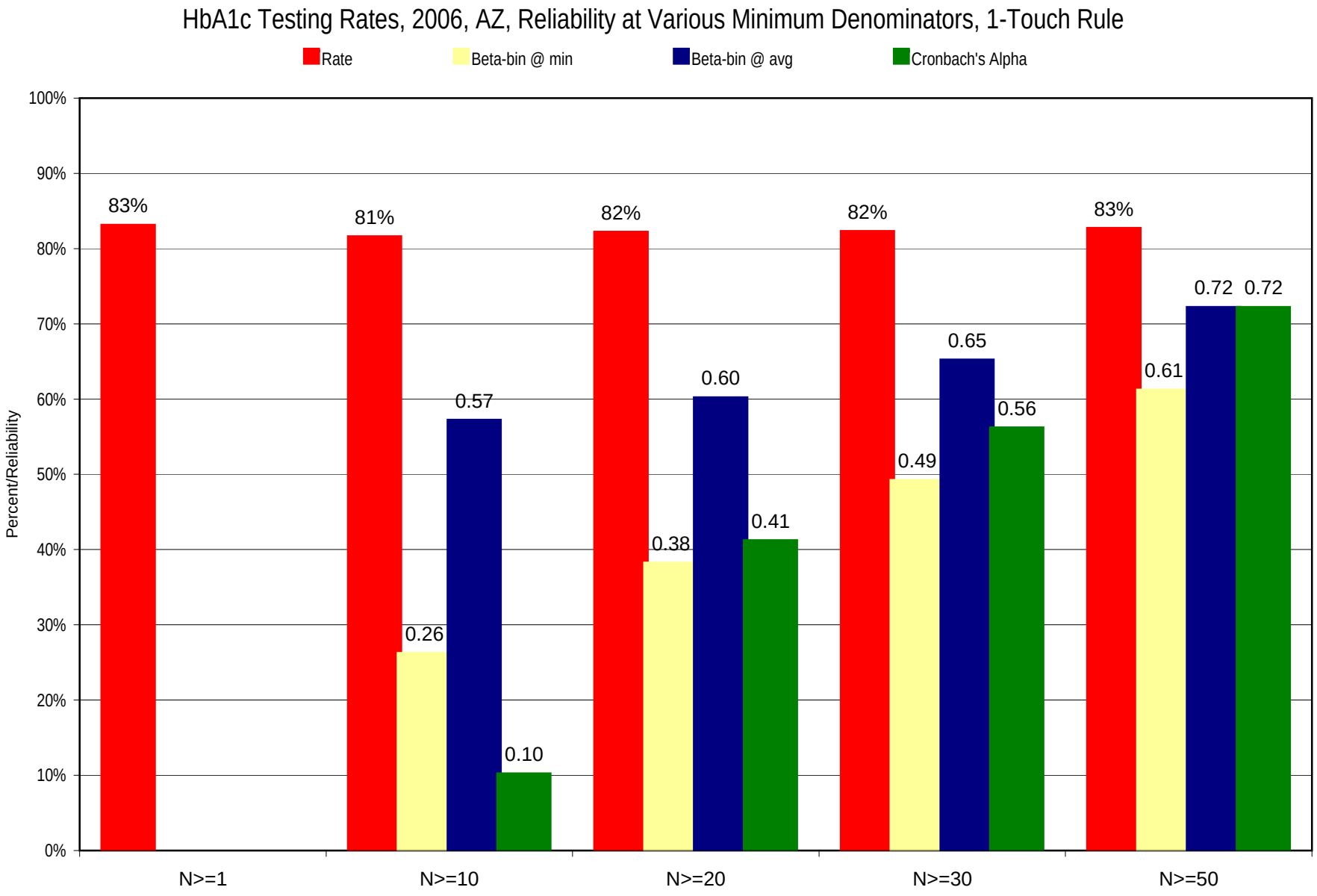

Colorectal Cancer Screening Rates, 2006, IN, Reliability at Various Minimum Denominators, 1-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
51%
47%
0.35
0.91
0.01
47%
0.52
0.91
0.51
47%
0.62
0.92
0.69
46%
0.74
0.93
0.81
N>=1
N>=10
N>=20
N>=30
N>=50

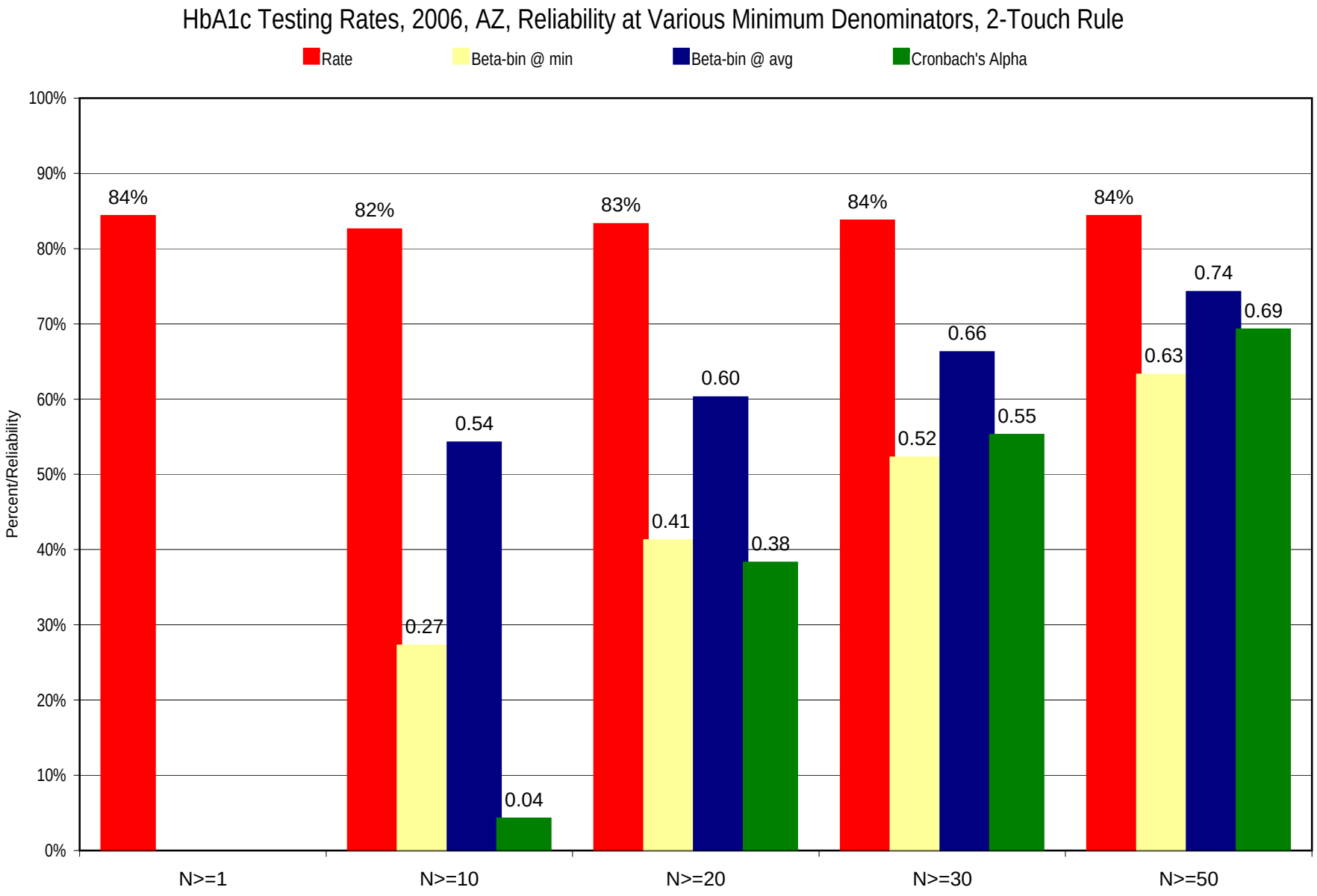

Colorectal Cancer Screening Rates, 2006, IN, Reliability at Various Minimum Denominators, 2-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
51%
47%
0.38
0.90
0.03
47%
0.55
0.91
0.44
47%
0.65
0.91
0.63
46%
0.75
0.92
0.77
N>=1
N>=10
N>=20
N>=30
N>=50

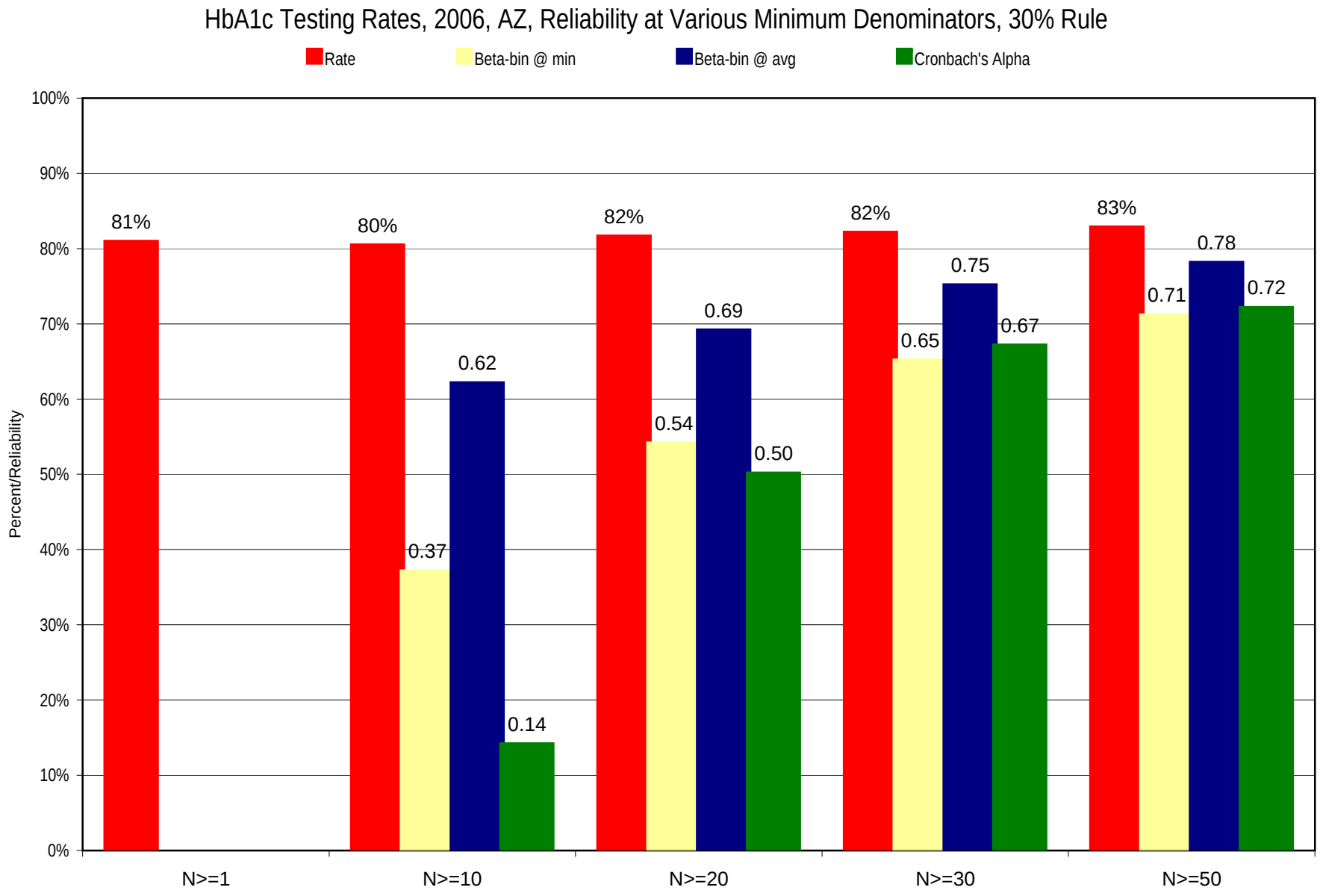

Colorectal Cancer Screening Rates, 2006, IN, Reliability at Various Minimum Denominators, 30% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
45%
41%
0.40
0.89
0.06
41%
0.57
0.90
0.42
41%
0.66
0.90
0.61
41%
0.76
0.91
0.75
N>=1
N>=10
N>=20
N>=30
N>=50

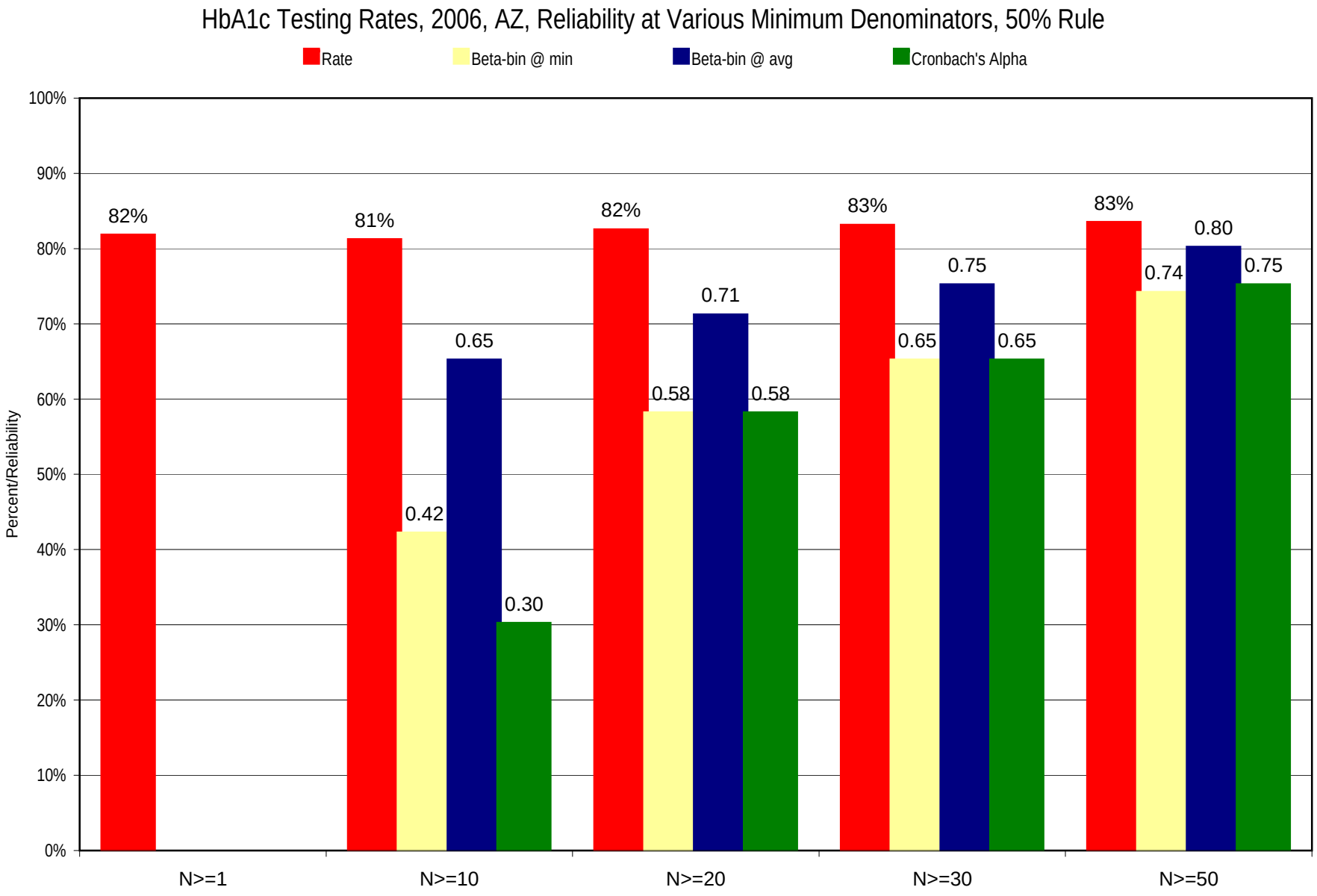

Colorectal Cancer Screening Rates, 2006, IN, Reliability at Various Minimum Denominators, 50% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
43%
39%
0.43
0.89
0.00
40%
0.60
0.90
0.47
41%
0.69
0.91
0.63
41%
0.79
0.92
0.78
N>=1
N>=10
N>=20
N>=30
N>=50

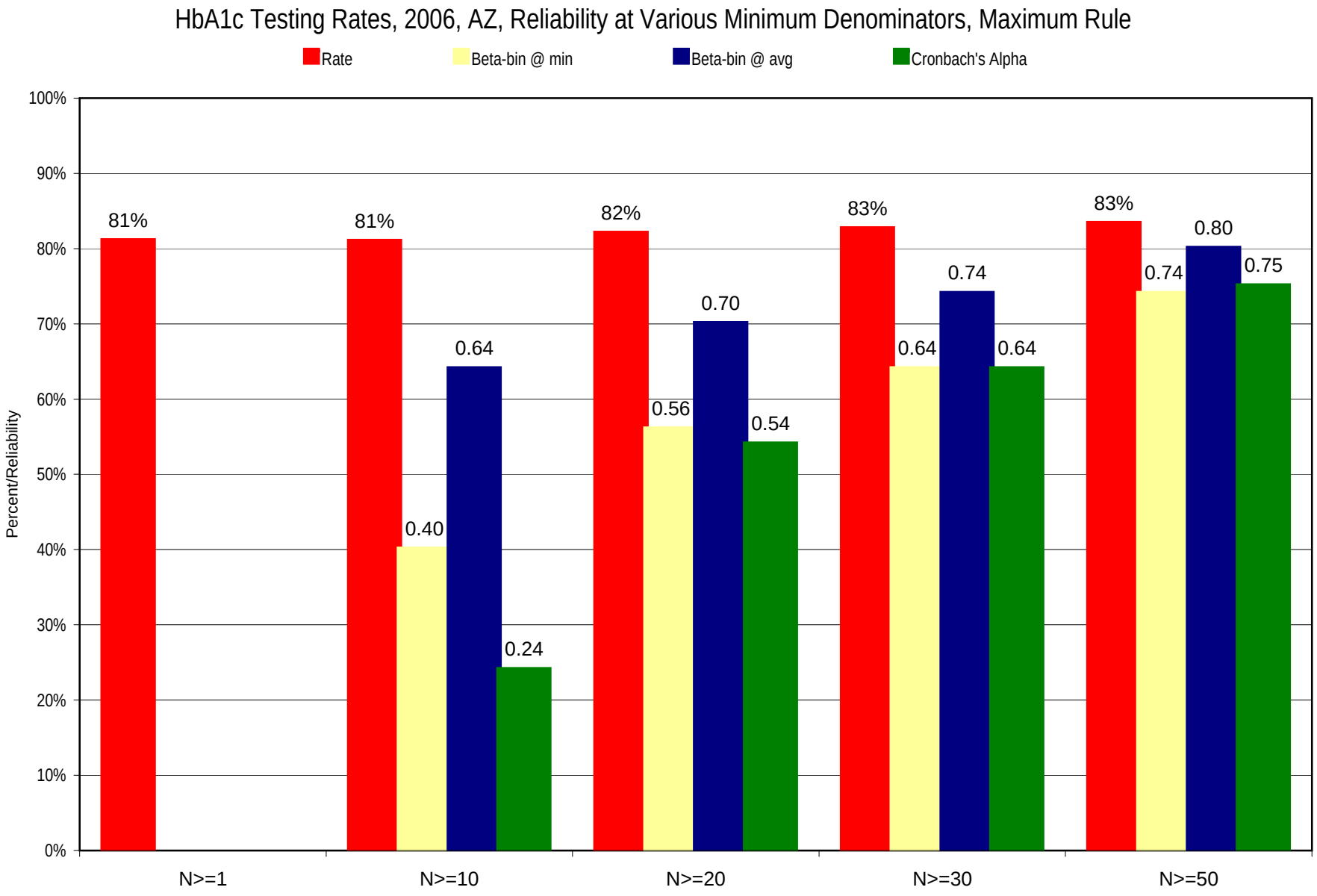

Colorectal Cancer Screening Rates, 2006, IN, Reliability at Various Minimum Denominators, Max Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
44%
40%
0.41
0.89
0.10
40%
0.58
0.90
0.42
41%
0.67
0.90
0.60
42%
0.77
0.91
0.75
N>=1
N>=10
N>=20
N>=30
N>=50

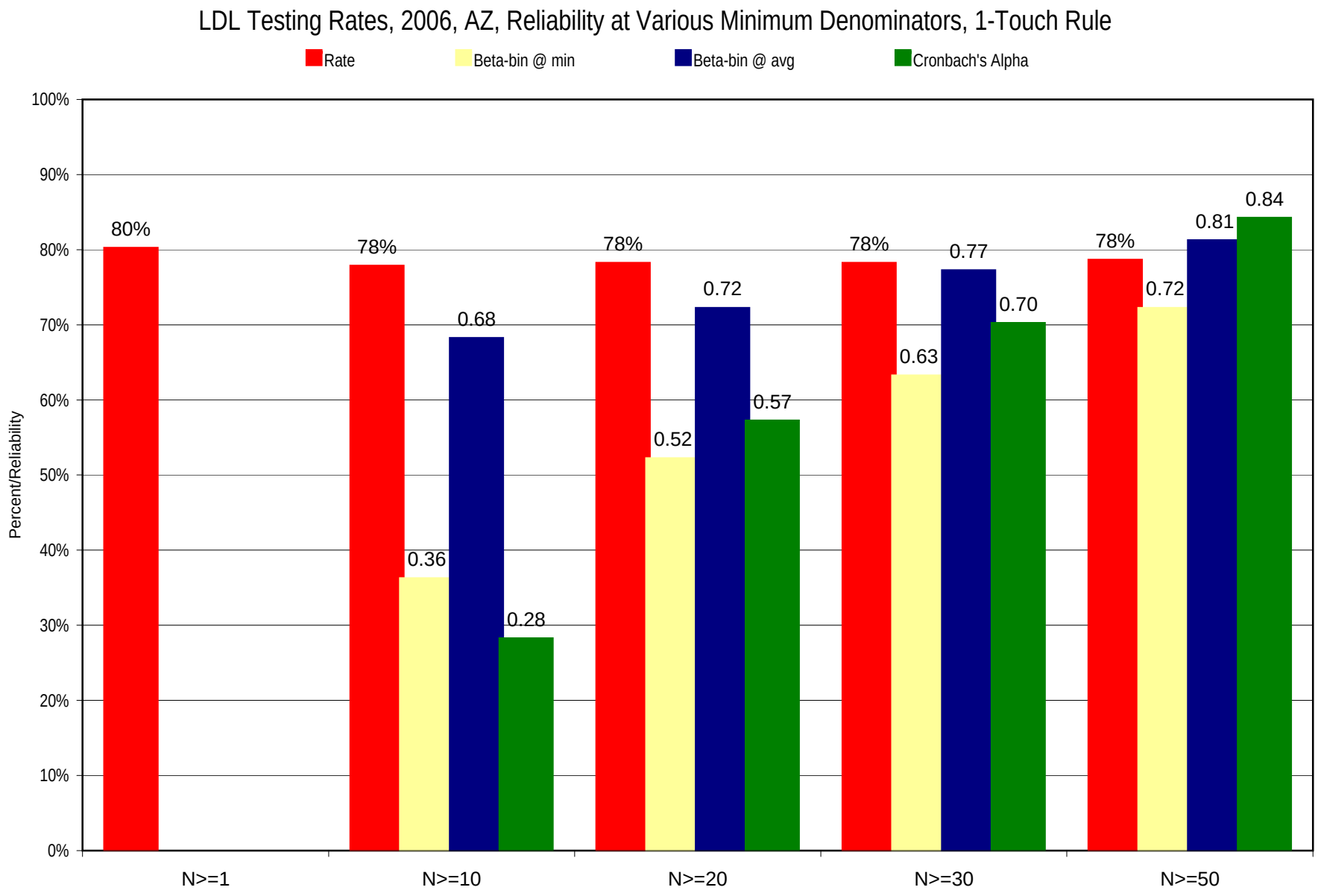

HbA1c Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 1-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
83%
81%
0.26
0.57
0.10
82%
0.38
0.60
0.41
82%
0.49
0.65
0.56
83%
0.61
0.72
0.72
N>=1
N>=10
N>=20
N>=30
N>=50

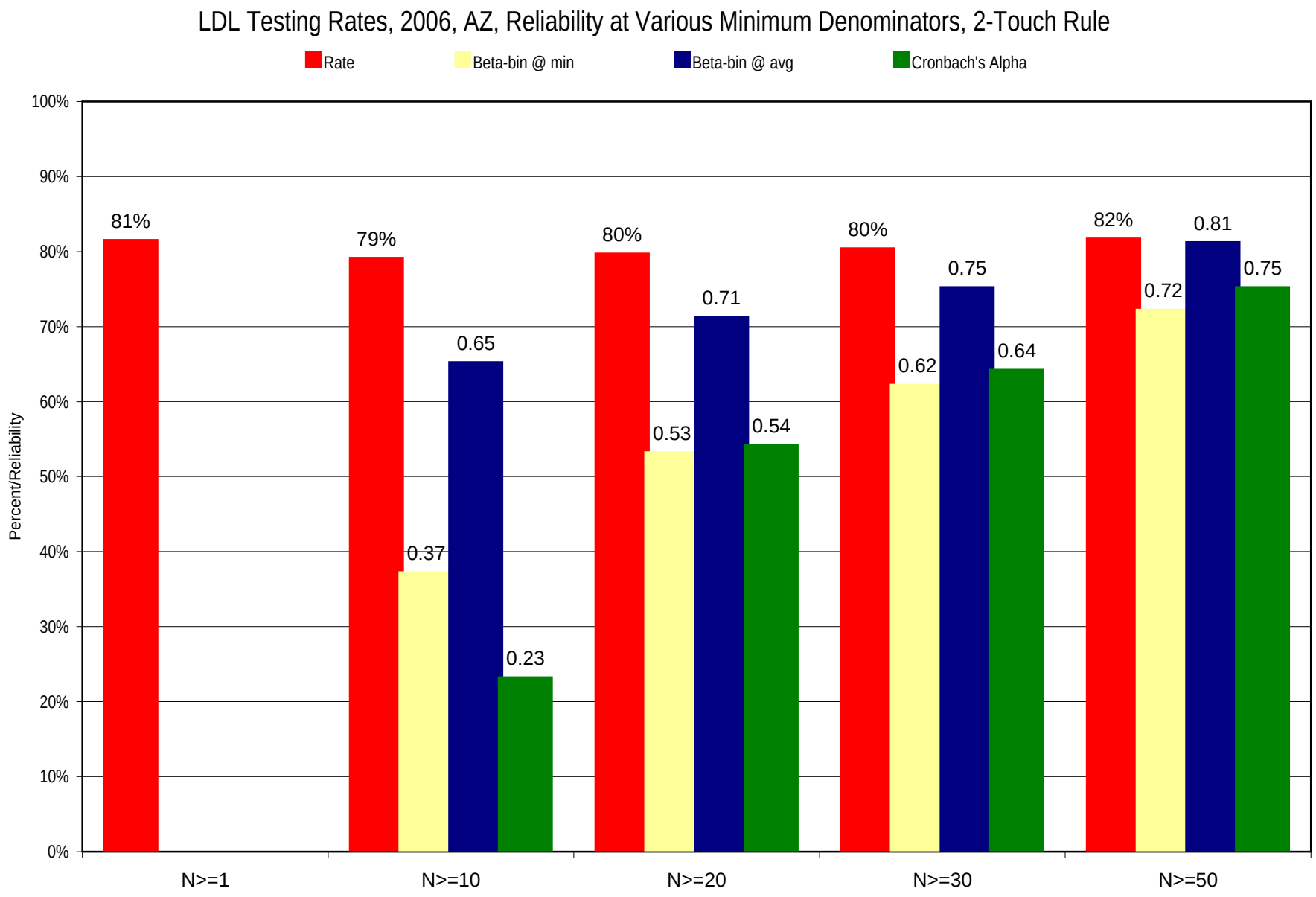

HbA1c Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 2-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
84%
82%
0.27
0.54
0.04
83%
0.41
0.60
0.38
84%
0.52
0.66
0.55
84%
0.63
0.74
0.69
N>=1
N>=10
N>=20
N>=30
N>=50

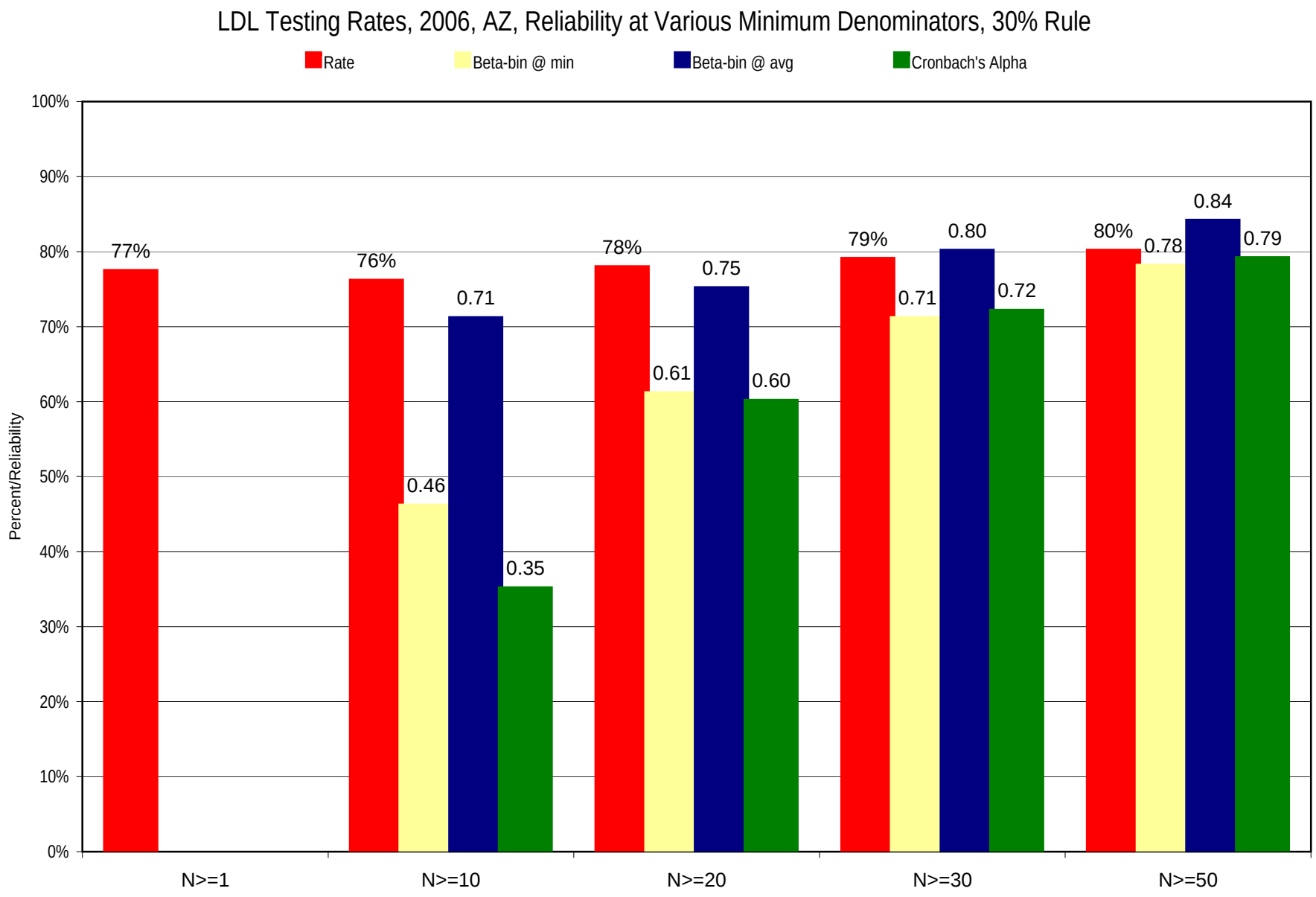

HbA1c Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 30% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
81%
80%
0.37
0.62
0.14
82%
0.54
0.69
0.50
82%
0.65
0.75
0.67
83%
0.71
0.78
0.72
N>=1
N>=10
N>=20
N>=30
N>=50

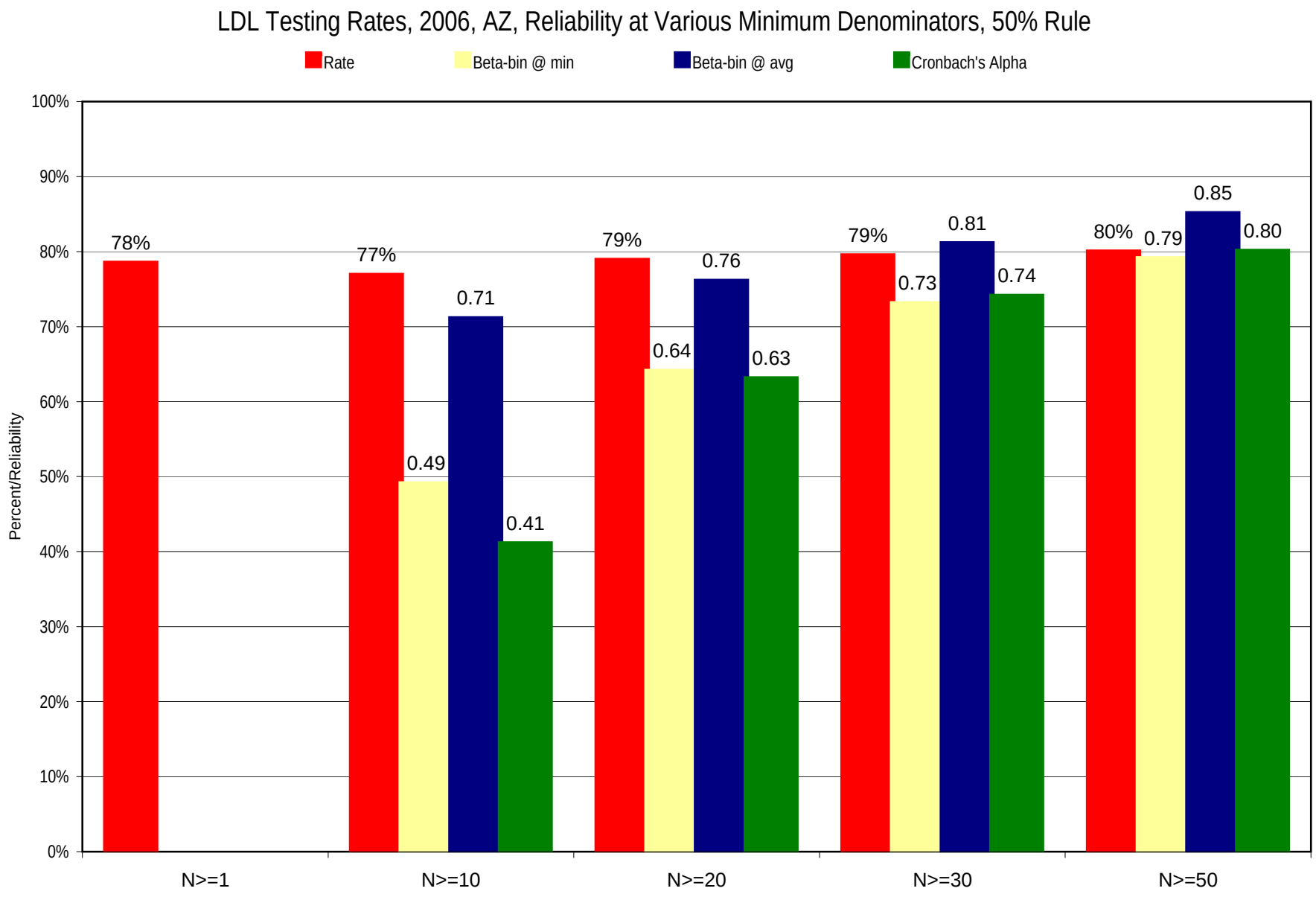

HbA1c Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 50% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
82%
81%
0.42
0.65
0.30
82%
0.58
0.71
0.58
83%
0.65
0.75
0.65
83%
0.74
0.80
0.75
N>=1
N>=10
N>=20
N>=30
N>=50

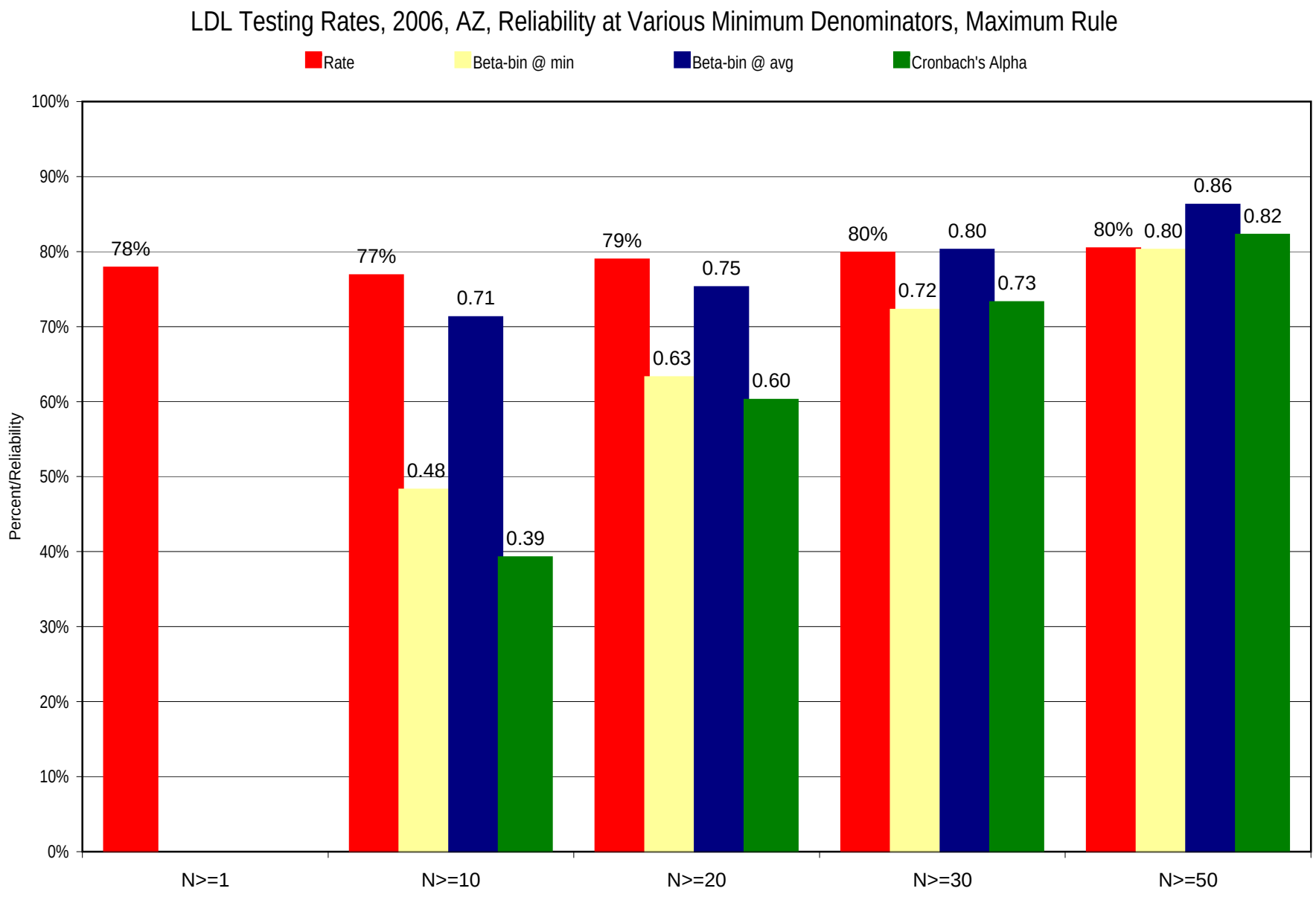

HbA1c Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, Maximum Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
81%
81%
0.40
0.64
0.24
82%
0.56
0.70
0.54
83%
0.64
0.74
0.64
83%
0.74
0.80
0.75
N>=1
N>=10
N>=20
N>=30
N>=50

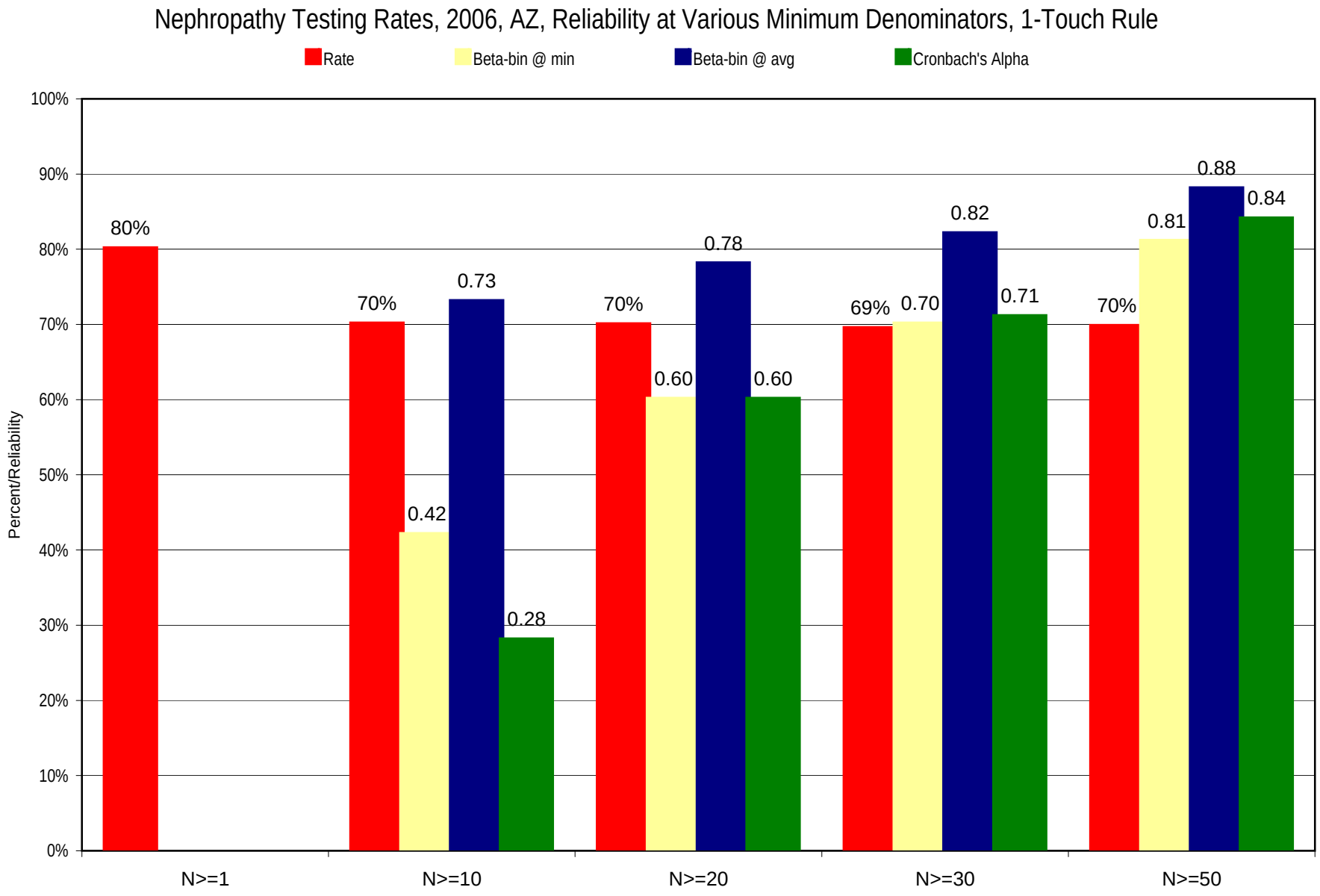

LDL Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 1-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
80%
78%
0.36
0.68
0.28
78%
0.52
0.72
0.57
78%
0.63
0.77
0.70
78%
0.72
0.81
0.84
N>=1
N>=10
N>=20
N>=30
N>=50

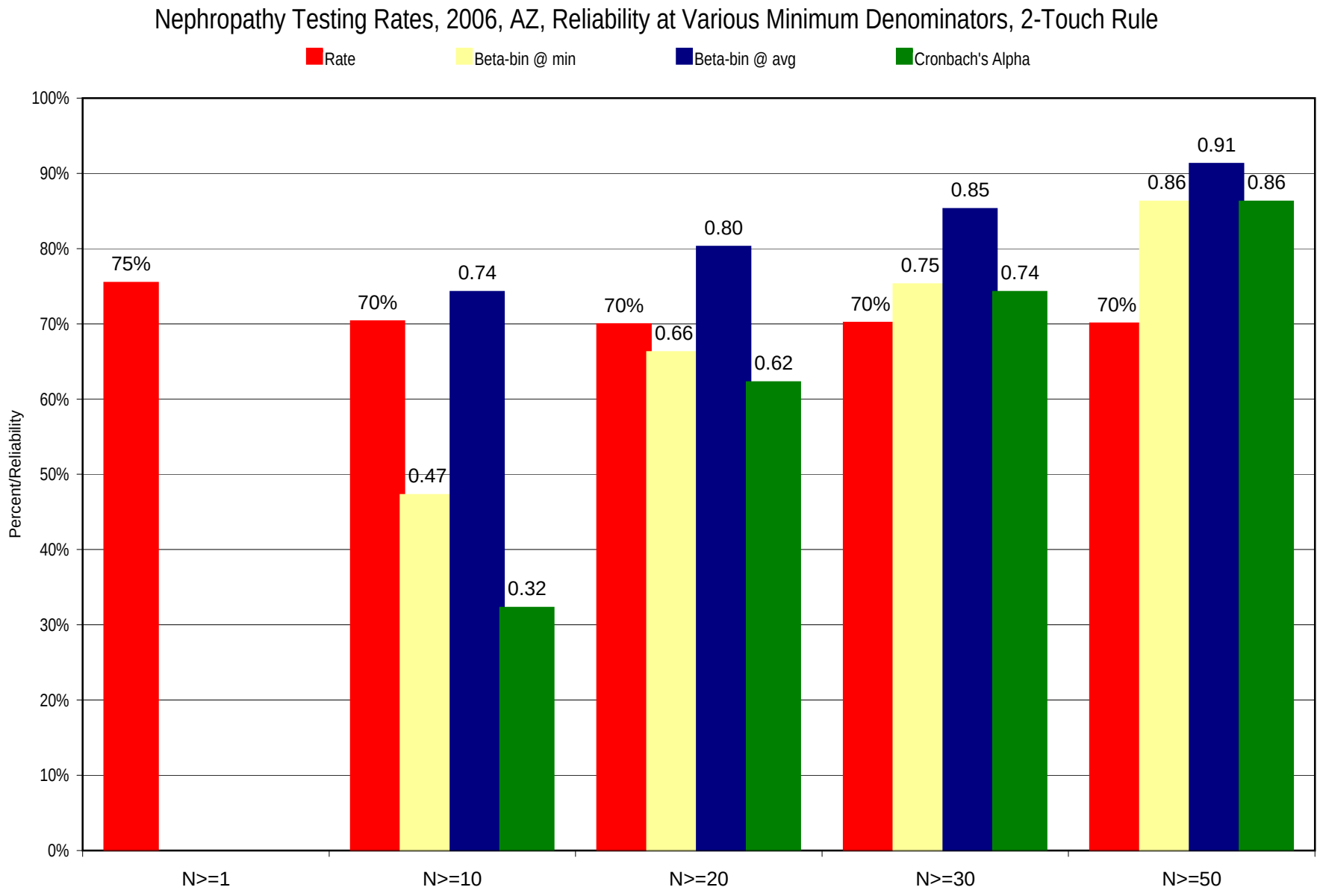

LDL Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 2-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
81%
79%
0.37
0.65
0.23
80%
0.53
0.71
0.54
80%
0.62
0.75
0.64
82%
0.72
0.81
0.75
N>=1
N>=10
N>=20
N>=30
N>=50

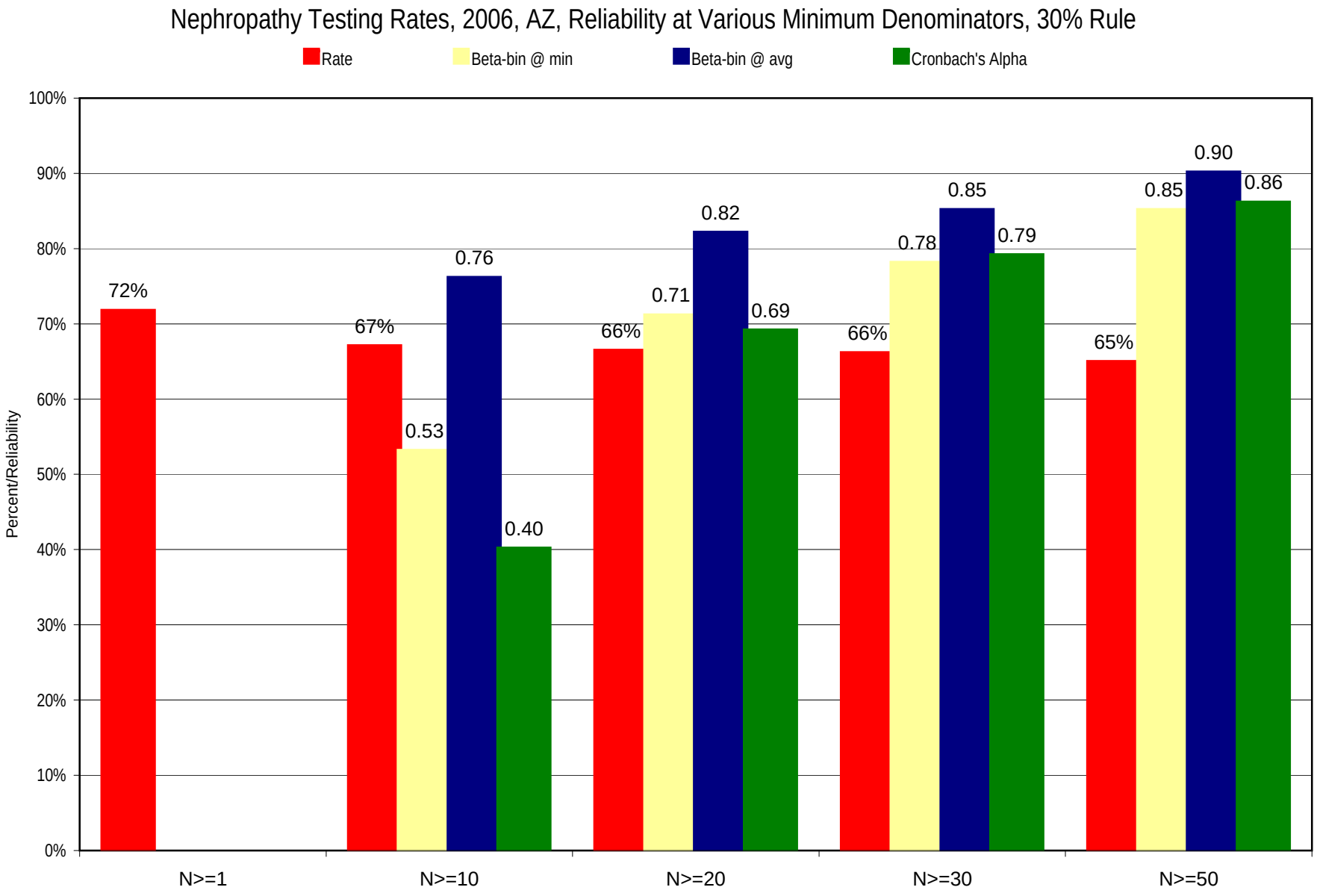

LDL Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 30% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
77%
76%
0.46
0.71
0.35
78%
0.61
0.75
0.60
79%
0.71
0.80
0.72
80%
0.78
0.84
0.79
N>=1
N>=10
N>=20
N>=30
N>=50

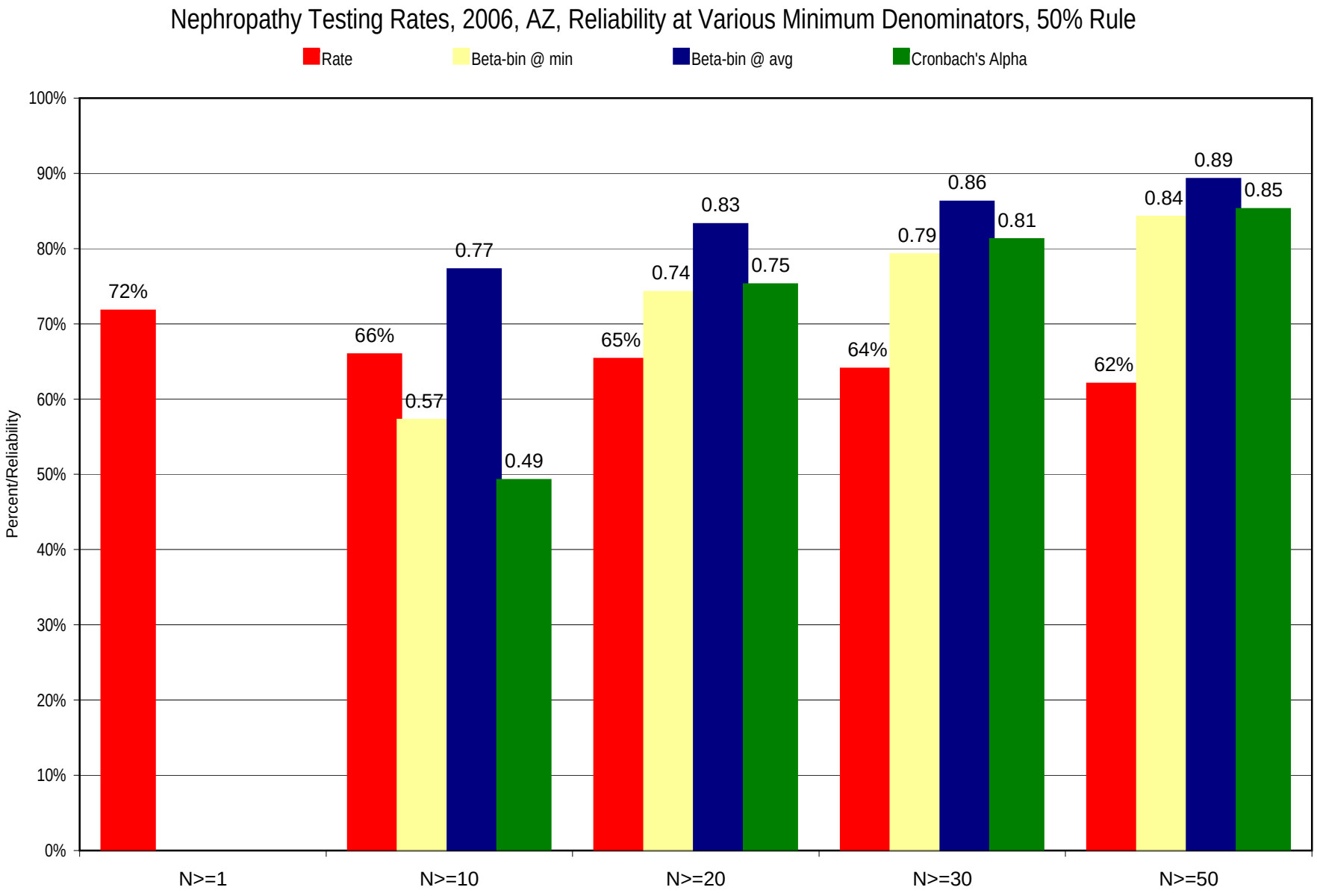

LDL Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 50% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
78%
77%
0.49
0.71
0.41
79%
0.64
0.76
0.63
79%
0.73
0.81
0.74
80%
0.79
0.85
0.80
N>=1
N>=10
N>=20
N>=30
N>=50

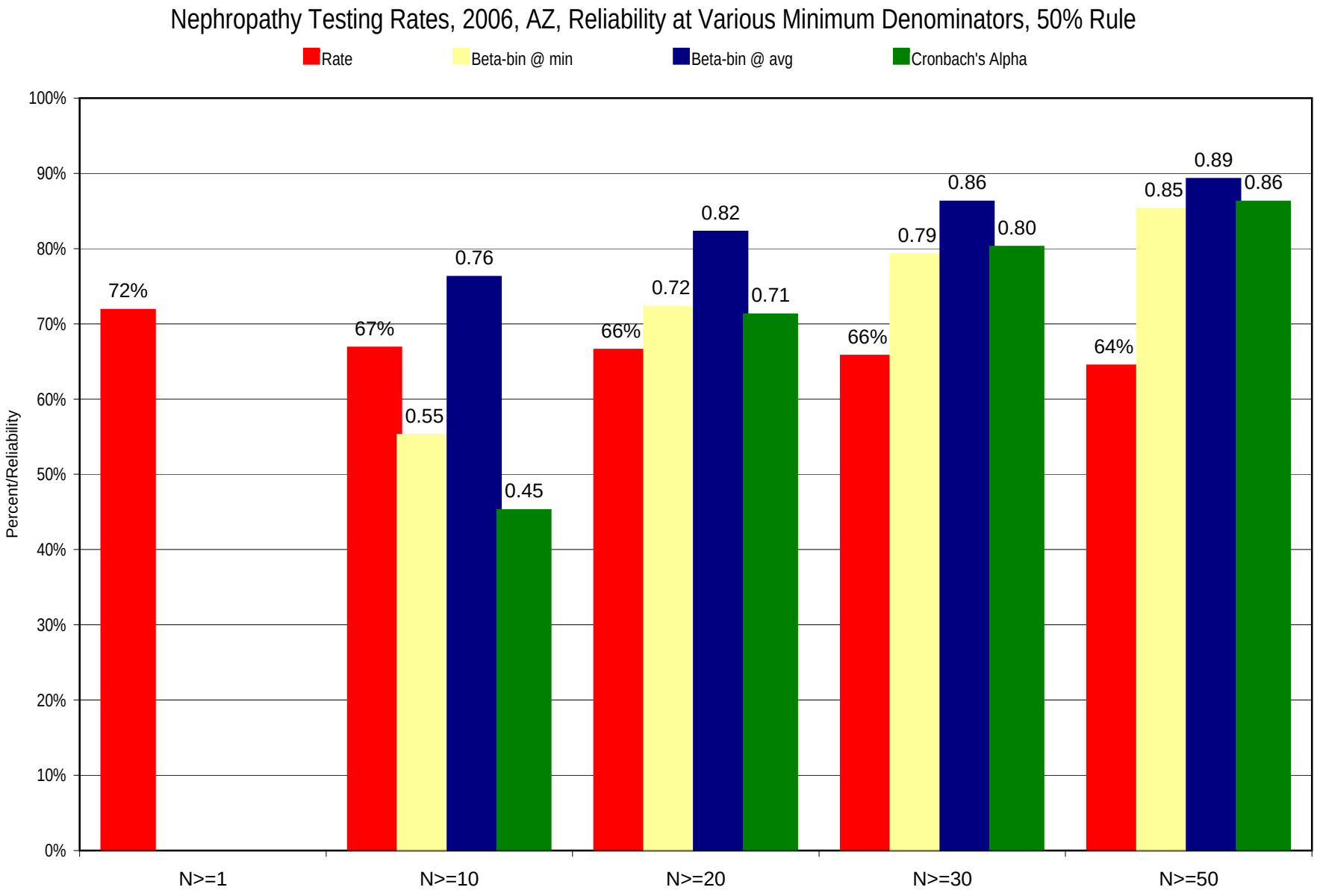

LDL Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, Maximum Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
78%
77%
0.48
0.71
0.39
79%
0.63
0.75
0.60
80%
0.72
0.80
0.73
80%
0.80
0.86
0.82
N>=1
N>=10
N>=20
N>=30
N>=50

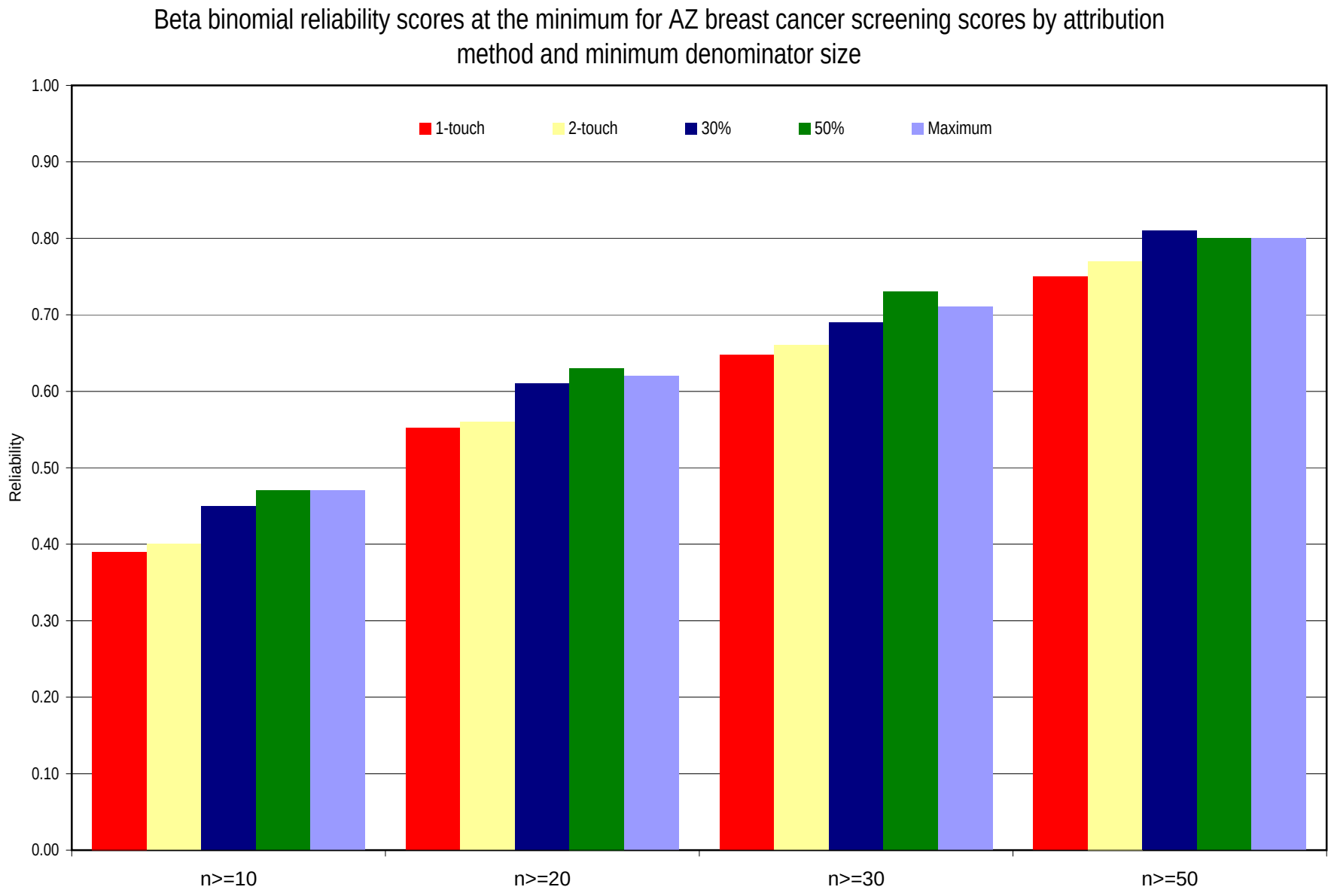

Nephropathy Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 1-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
80%
70%
0.42
0.73
0.28
70%
0.60
0.78
0.60
69%
0.70
0.82
0.71
70%
0.81
0.88
0.84
N>=1
N>=10
N>=20
N>=30
N>=50

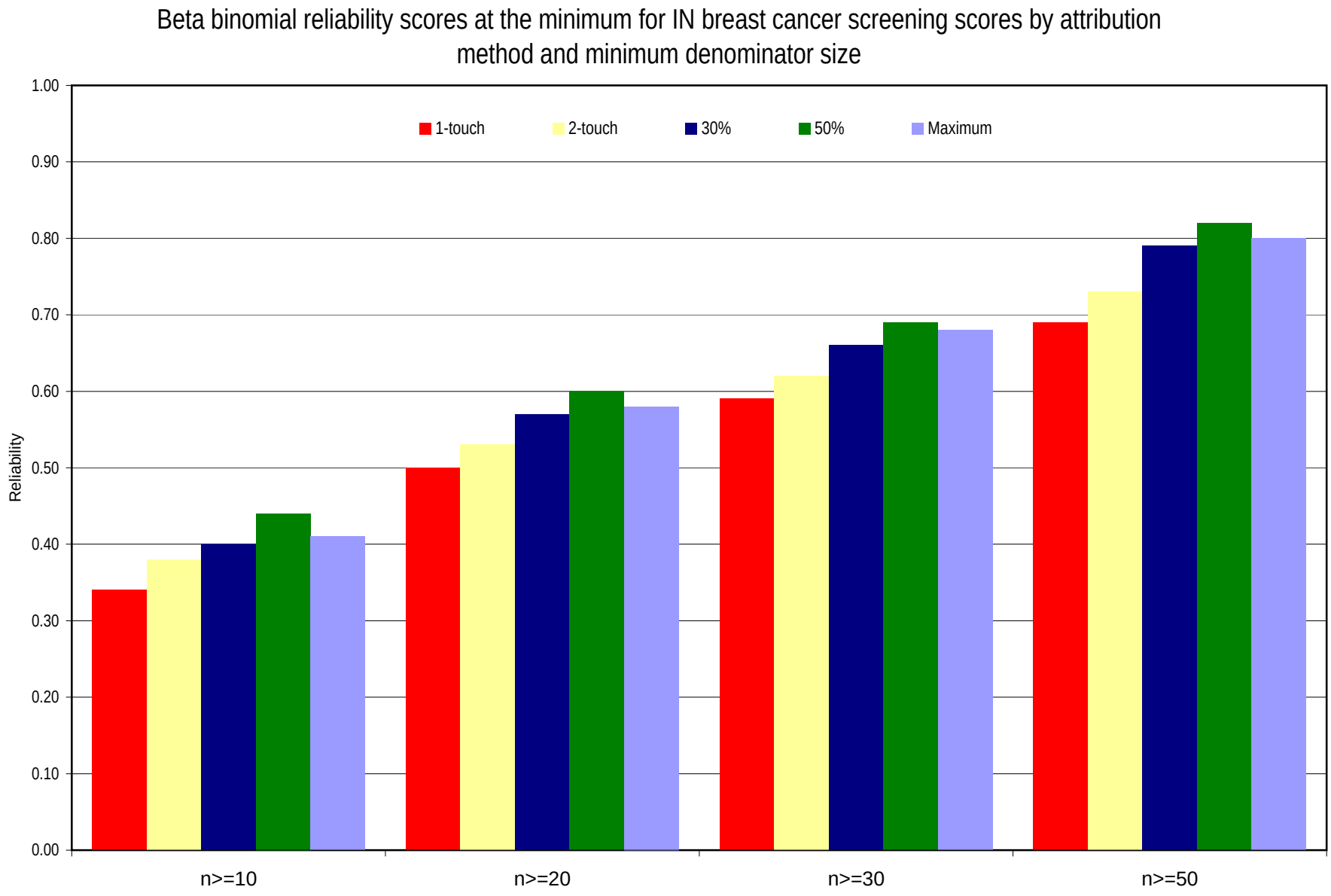

Nephropathy Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 2-Touch Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
75%
70%
0.47
0.74
0.32
70%
0.66
0.80
0.62
70%
0.75
0.85
0.74
70%
0.86
0.91
0.86
N>=1
N>=10
N>=20
N>=30
N>=50

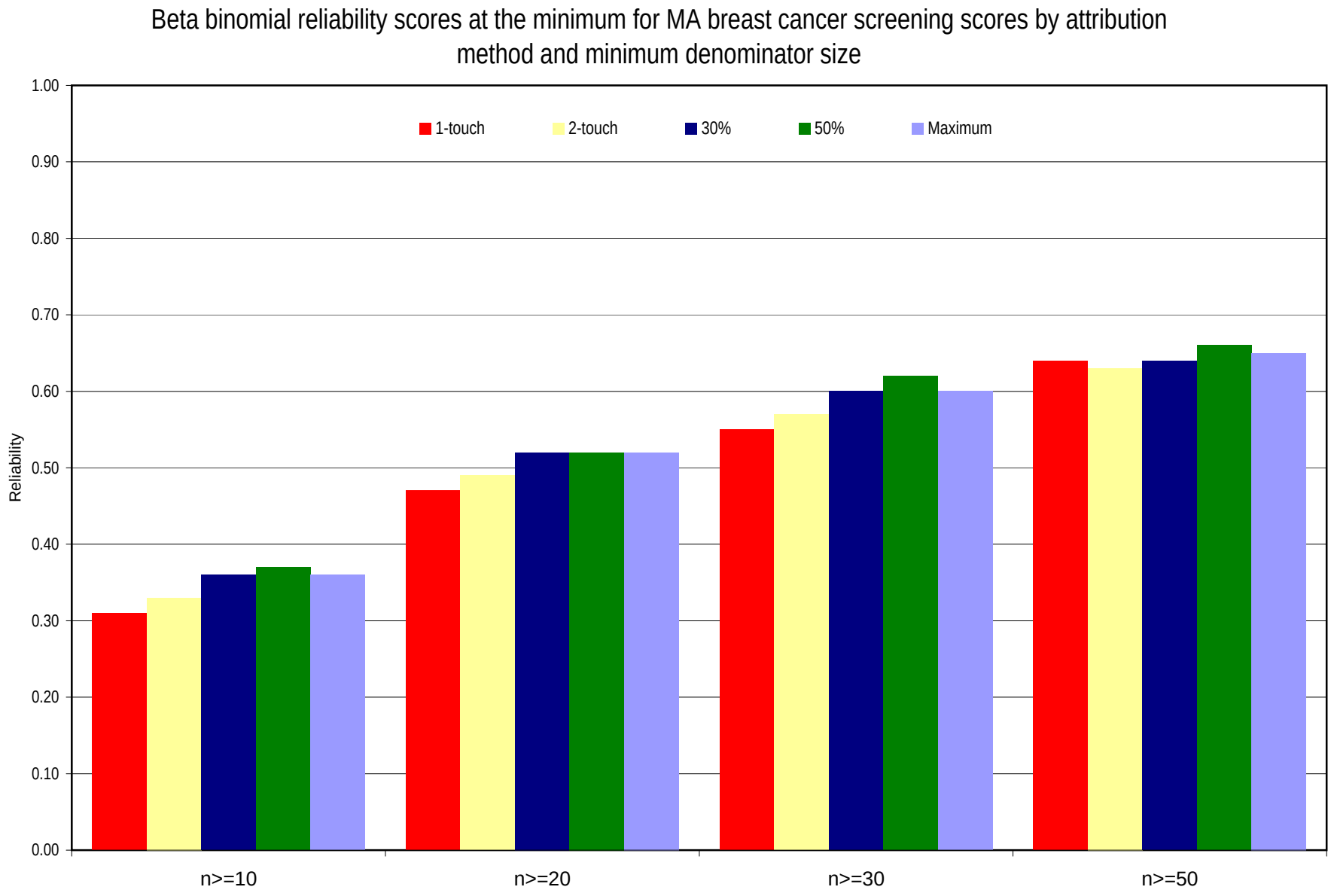

Nephropathy Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 30% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
72%
67%
0.53
0.76
0.40
66%
0.71
0.82
0.69
66%
0.78
0.85
0.79
65%
0.85
0.90
0.86
N>=1
N>=10
N>=20
N>=30
N>=50

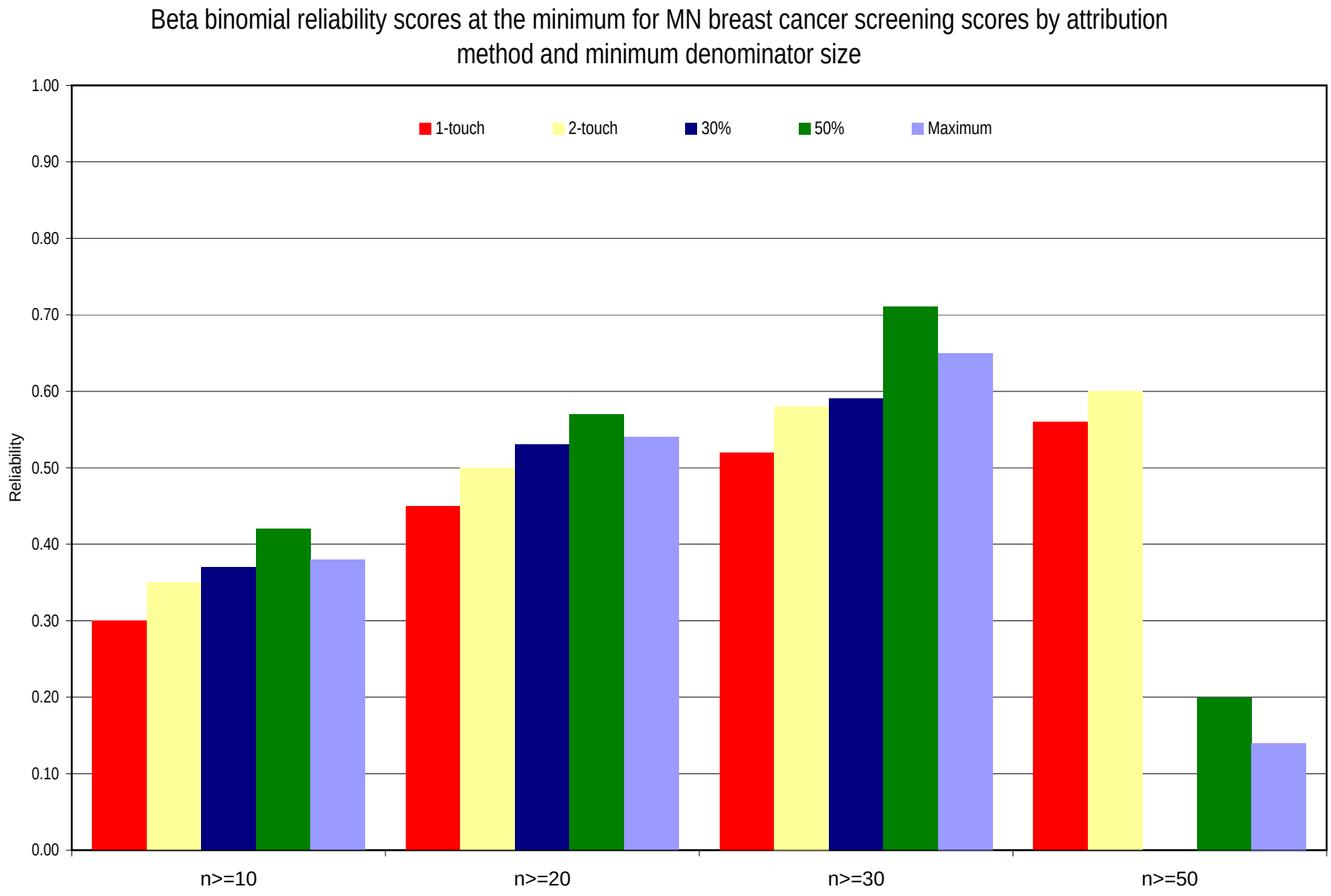

Nephropathy Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 50% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
72%
66%
0.57
0.77
0.49
65%
0.74
0.83
0.75
64%
0.79
0.86
0.81
62%
0.84
0.89
0.85
N>=1
N>=10
N>=20
N>=30
N>=50

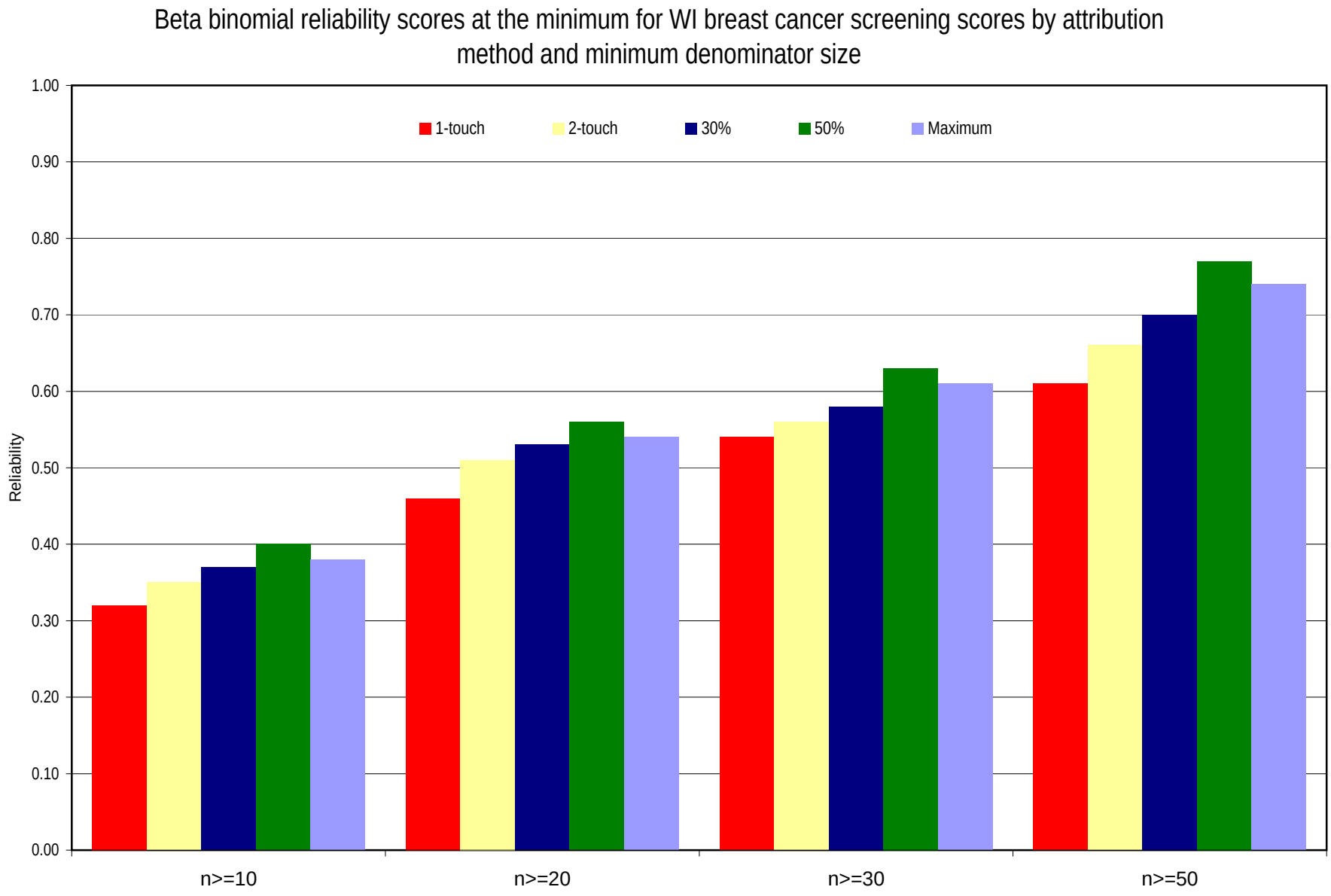

Nephropathy Testing Rates, 2006, AZ, Reliability at Various Minimum Denominators, 50% Rule
Rate
Beta-bin @ min
Beta-bin @ avg
Cronbach's Alpha
Percent/Reliability
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
72%
67%
0.55
0.76
0.45
66%
0.72
0.82
0.71
66%
0.79
0.86
0.80
64%
0.85
0.89
0.86
N>=1
N>=10
N>=20
N>=30
N>=50

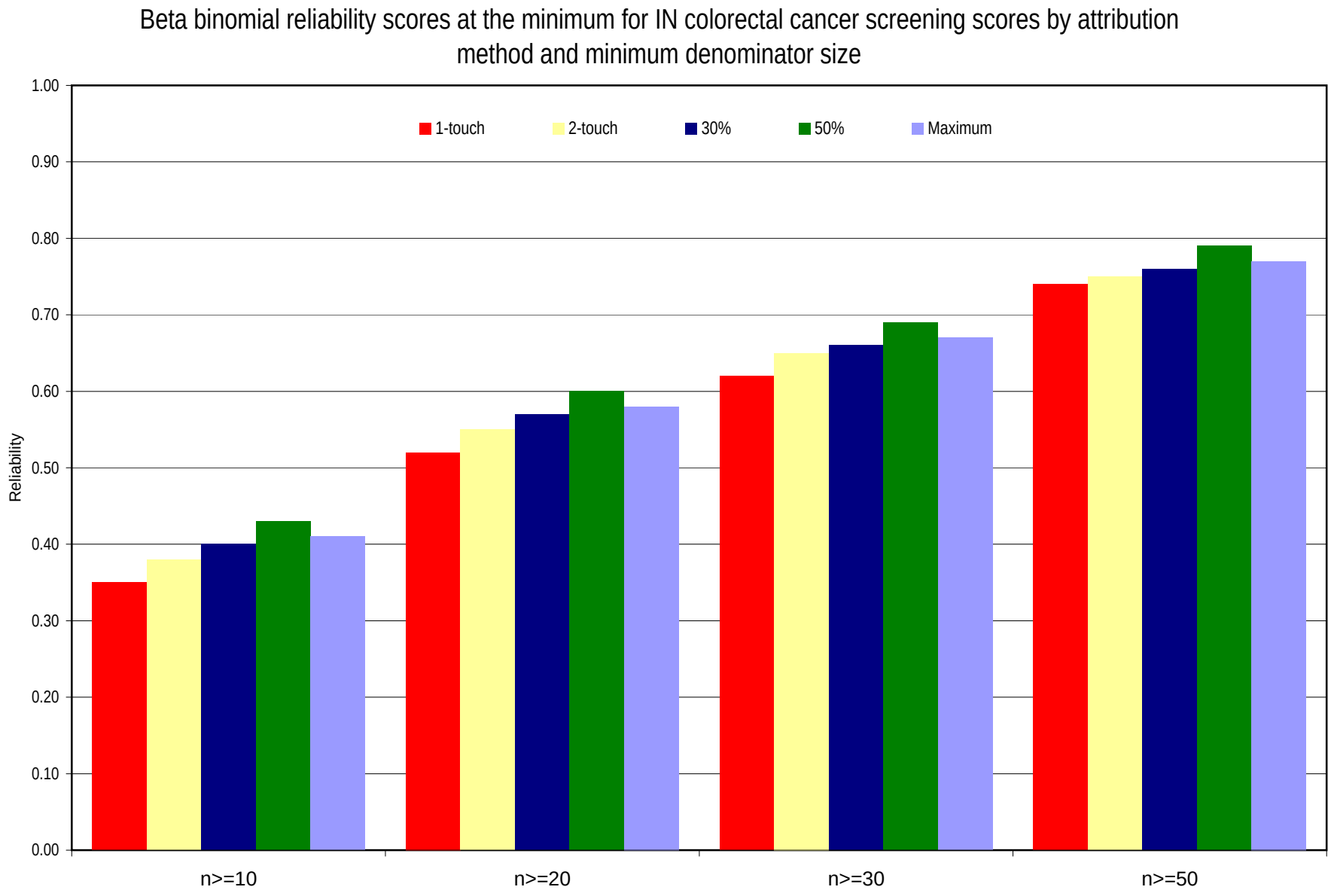

Beta binomial reliability scores at the minimum for AZ breast cancer screening scores by attribution method and minimum denominator size
1-touch
2-touch
30%
50%
Maximum
Reliability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
n>=10
n>=20
n>=30
n>=50


Beta binomial reliability scores at the minimum for IN breast cancer screening scores by attribution method and minimum denominator size
1-touch
2-touch
30%
50%
Maximum
Reliability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
n>=10
n>=20
n>=30
n>=50

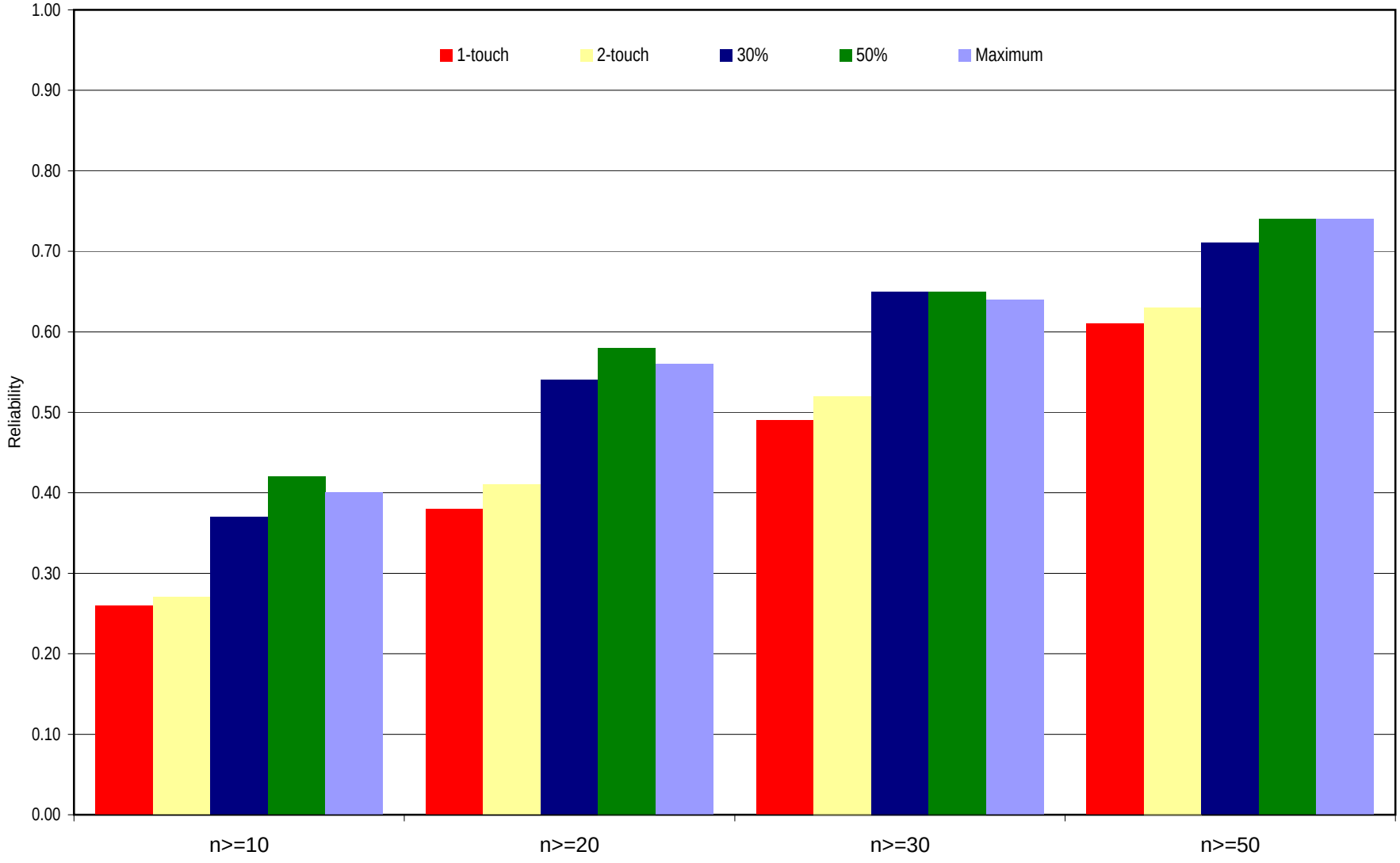

Beta binomial reliability scores at the minimum for MA breast cancer screening scores by attribution method and minimum denominator size
1-touch
2-touch
30%
50%
Maximum
Reliability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
n>=10
n>=20
n>=30
n>=50

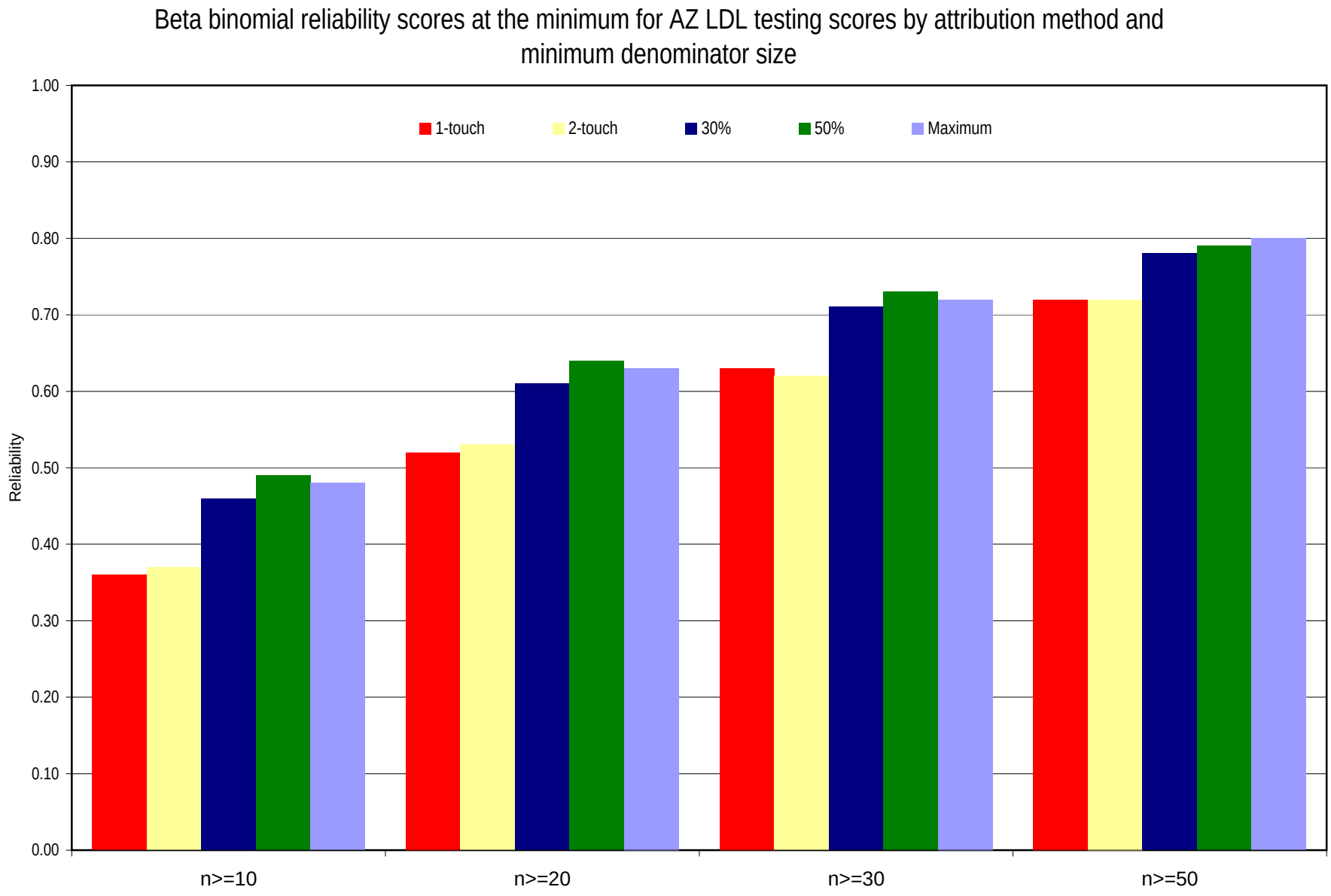

Beta binomial reliability scores at the minimum for MN breast cancer screening scores by attribution method and minimum denominator size
1-touch
2-touch
30%
50%
Maximum
Reliability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
n>=10
n>=20
n>=30
n>=50

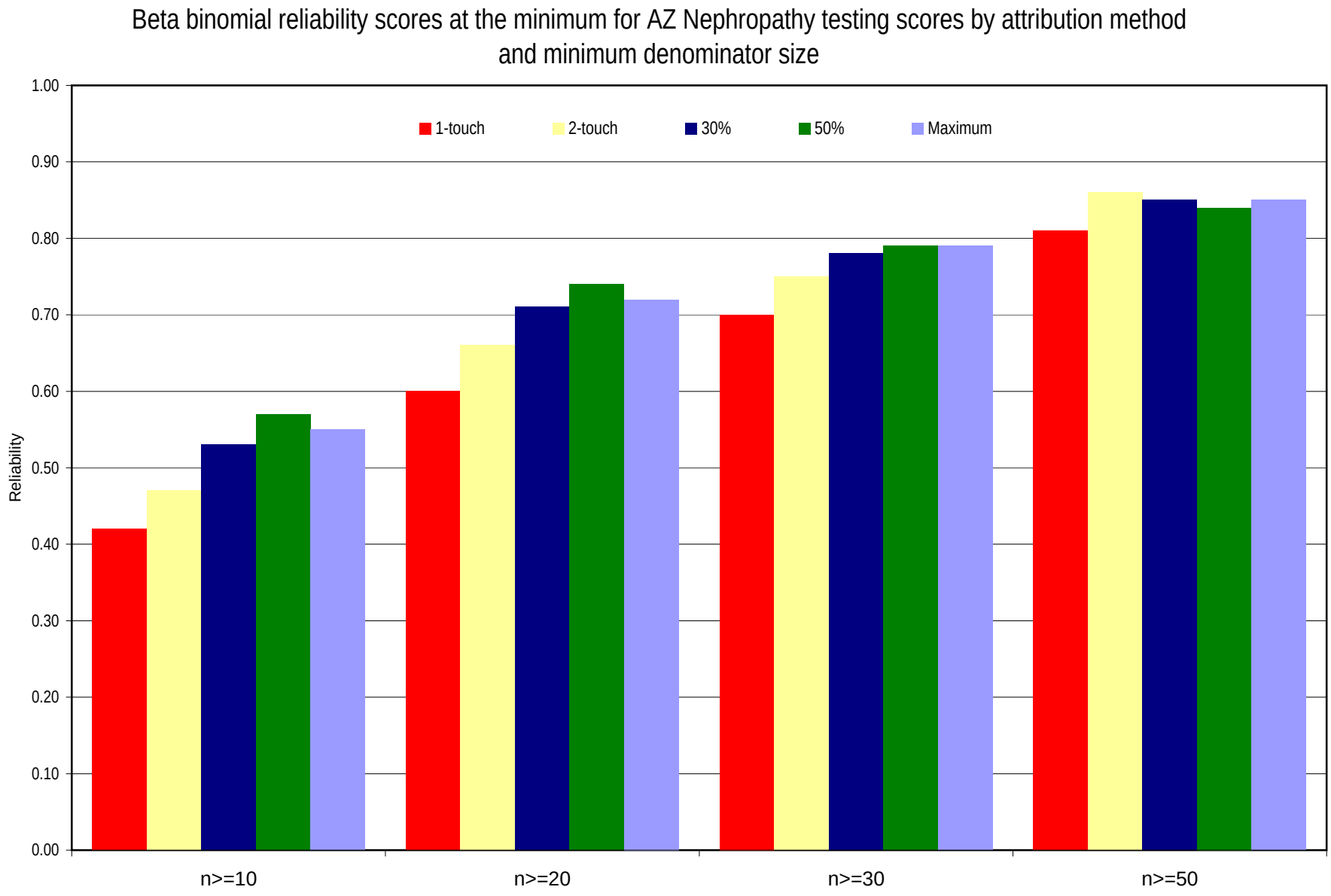

Beta binomial reliability scores at the minimum for WI breast cancer screening scores by attribution method and minimum denominator size
1-touch
2-touch
30%
50%
Maximum
Reliability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
n>=10
n>=20
n>=30
n>=50


Beta binomial reliability scores at the minimum for IN colorectal cancer screening scores by attribution method and minimum denominator size
1-touch
2-touch
30%
50%
Maximum
Reliability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
n>=10
n>=20
n>=30
n>=50

Beta binomial reliability scores at the minimum for AZ HbA1c testing scores by attribution method and minimum denominator size




Beta binomial reliability scores at the minimum for AZ LDL testing scores by attribution method and minimum denominator size
1-touch
2-touch
30%
50%
Maximum
Reliability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
n>=10
n>=20
n>=30
n>=50


Beta binomial reliability scores at the minimum for AZ Nephropathy testing scores by attribution method and minimum denominator size
1-touch
2-touch
30%
50%
Maximum
Reliability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
n>=10
n>=20
n>=30
n>=50

Reliability Scores for Selected Denominators (50% Rule)

† % eligible physicians w/ rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ % beneficiaries in denominator w/ a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

* Reliability calculated using the beta-binomial model and average denominator size

** Reliability calculated using the beta-binomial model and minimum denominator size

Measure	State	Minimum denominator	Average denominator	Mean physician rate	Standard deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physician‡	Beta-bin @ average*	Beta-bin @ minimum**	Cronbach's alpha
Breast Cancer Screening	AZ	10	24.4	66.7%	17.3%	19.6%	59.0%	0.68	0.47	0.31
		20	34.4	69.0%	15.6%	10.0%	42.2%	0.75	0.63	0.59
		30	45.0	70.3%	15.1%	5.0%	27.6%	0.80	0.73	0.72
		50	66.1	72.6%	13.9%	1.3%	10.8%	0.84	0.80	0.81
	CA	10								
		20								
		30								
		50								
	IN	10	26.8	61.6%	16.9%	23.4%	69.4%	0.67	0.44	0.21
		20	33.9	62.7%	15.7%	14.9%	55.8%	0.72	0.60	0.55
		30	42.2	63.8%	14.9%	8.0%	37.3%	0.76	0.69	0.68
		50	61.3	62.4%	15.9%	1.7%	11.3%	0.85	0.82	0.83
	MA	10	22.2	73.5%	14.9%	21.8%	58.2%	0.56	0.37	0.15
		20	30.8	74.4%	13.1%	10.7%	39.4%	0.63	0.52	0.48
		30	40.5	75.6%	12.2%	4.4%	21.4%	0.68	0.62	0.60
		50	62.8	75.8%	9.9%	0.7%	5.0%	0.71	0.66	0.64
	MN	10	17.4	74.7%	15.7%	15.6%	45.3%	0.56	0.42	0.26
		20	26.0	77.3%	13.3%	4.6%	20.1%	0.63	0.57	0.54
		30	35.5	77.4%	14.0%	1.0%	5.9%	0.75	0.71	0.73
		50	53.7	80.8%	7.5%	0.0%	0.4%	0.21	0.20	0.47
	WI	10	20.7	71.5%	15.8%	25.8%	58.3%	0.58	0.40	0.22
		20	29.1	73.1%	14.0%	11.2%	35.6%	0.65	0.56	0.53
		30	37.9	73.9%	12.6%	4.3%	17.9%	0.68	0.63	0.63
		50	57.4	74.2%	13.1%	0.3%	2.2%	0.79	0.77	0.80

Reliability Scores for Selected Denominators (50% Rule)

† % eligible physicians w/ rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ % beneficiaries in denominator w/ a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

* Reliability calculated using the beta-binomial model and average denominator size

** Reliability calculated using the beta-binomial model and minimum denominator size

Measure	State	Minimum denominator	Average denominator	Mean physician rate	Standard deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physician‡	Beta-bin @ average*	Beta-bin @ minimum**	Cronbach's alpha
Colorectal Cancer Screening	AZ	10								
		20								
		30								
		50								
	CA	10								
		20								
		30								
		50								
	IN	10	109.0	39.1%	15.2%	44.0%	75.9%	0.89	0.43	0.00
		20	123.0	39.8%	14.6%	38.4%	74.7%	0.90	0.60	0.47
		30	132.0	40.5%	14.3%	34.9%	73.3%	0.91	0.69	0.63
		50	148.0	41.2%	14.2%	30.0%	70.3%	0.92	0.79	0.78
	MA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MN	10								
		20								
		30								
		50								
	WI	10								
		20								
		30								
		50								

Reliability Scores for Selected Denominators (50% Rule)

† % eligible physicians w/ rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ % beneficiaries in denominator w/ a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

* Reliability calculated using the beta-binomial model and average denominator size

** Reliability calculated using the beta-binomial model and minimum denominator size

Measure	State	Minimum denominator	Average denominator	Mean physician rate	Standard deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physician‡	Beta-bin @ average*	Beta-bin @ minimum**	Cronbach's alpha
HbA1c Testing	AZ	10	25.2	81.0%	14.2%	20.7%	63.5%	0.65	0.42	0.30
		20	35.5	82.3%	12.8%	10.7%	46.3%	0.71	0.58	0.58
		30	48.1	82.9%	11.3%	5.1%	30.0%	0.75	0.65	0.65
		50	73.1	83.3%	10.3%	1.5%	13.5%	0.80	0.74	0.75
	CA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	IN	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MN	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	WI	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			

Reliability Scores for Selected Denominators (50% Rule)

† % eligible physicians w/ rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ % beneficiaries in denominator w/ a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

* Reliability calculated using the beta-binomial model and average denominator size

** Reliability calculated using the beta-binomial model and minimum denominator size

Measure	State	Minimum denominator	Average denominator	Mean physician rate	Standard deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physician‡	Beta-bin @ average*	Beta-bin @ minimum**	Cronbach's alpha
Retinal Eye Exam	AZ	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	CA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	IN	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MN	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	WI	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			

Reliability Scores for Selected Denominators (50% Rule)

† % eligible physicians w/ rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ % beneficiaries in denominator w/ a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

* Reliability calculated using the beta-binomial model and average denominator size

** Reliability calculated using the beta-binomial model and minimum denominator size

Measure	State	Minimum denominator	Average denominator	Mean physician rate	Standard deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physician‡	Beta-bin @ average*	Beta-bin @ minimum**	Cronbach's alpha
LDL Testing	AZ	10	25.2	76.8%	16.7%	20.7%	63.5%	0.71	0.49	0.41
		20	35.5	78.8%	14.4%	10.7%	46.3%	0.76	0.64	0.63
		30	48.1	79.4%	13.7%	5.1%	30.0%	0.81	0.73	0.74
		50	73.1	79.9%	12.3%	1.5%	13.5%	0.85	0.79	0.80
	CA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	IN	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MN	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	WI	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			

Reliability Scores for Selected Denominators (50% Rule)

† % eligible physicians w/ rate is the number of physicians with at least 1 beneficiary in their denominator divided by the number of eligible physicians

‡ % beneficiaries in denominator w/ a physician is the number of beneficiaries who were attributed to at least 1 physician divided by the number of beneficiaries in the denominator

* Reliability calculated using the beta-binomial model and average denominator size

** Reliability calculated using the beta-binomial model and minimum denominator size

Measure	State	Minimum denominator	Average denominator	Mean physician rate	Standard deviation	% eligible physicians w/ a rate†	% beneficiaries in denominator w/ a physician‡	Beta-bin @ average*	Beta-bin @ minimum**	Cronbach's alpha
Nephropathy Testing	AZ	10	25.2	65.7%	20.0%	20.7%	63.5%	0.77	0.57	0.49
		20	35.5	65.1%	19.8%	10.7%	46.3%	0.83	0.74	0.75
		30	48.1	63.8%	18.8%	5.1%	30.0%	0.86	0.79	0.81
		50	73.1	61.8%	17.1%	1.5%	13.5%	0.89	0.84	0.85
	CA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	IN	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MA	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	MN	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			
	WI	10				0.0%	0.0%			
		20				0.0%	0.0%			
		30				0.0%	0.0%			
		50				0.0%	0.0%			

Appendix 10

Final Cycle 3 Report
Report Appendix

Appendix A: BQI Project Measures

Appendix B: BQI Cycle 3 Measures, by Pilot

Centers for Medicare &
Medicaid Services

Better Quality Information to Improve Care for Medicare Beneficiaries

HHSM-500-2006-MD002C
Modification No. MD0008

Delmarva Foundation
9240 Centreville Road
Easton, MD 21601

July 31, 2008

Better Quality Information to Improve Care for Medicare Beneficiaries

(BQI) Project

Final Cycle 3 Report

This Cycle 3 Data Aggregation report has been prepared by Delmarva for the purpose of providing a comprehensive picture of the activities, accomplishments, and challenges associated with the third round of data aggregation. This report describes the experiences of four pilot sites: Center for Health Information & Research (CHIR), Indiana Health Information Exchange (IHIE), Minnesota Community Measurement (MNCM), and the Wisconsin Collaborative for Healthcare Quality (WCHQ), who are participating in Cycle 3. The two BQI sites not participating in Cycle 3 aggregation are the California Cooperative Healthcare Reporting Initiative (CCHRI) and Massachusetts Health Quality Partners (MHQP). This Cycle 3 report contains the following sections:

Executive Summary 3
Introduction and Common Themes 6
Measure Selection Challenges and Lessons Learned 8
Data Aggregation and Analysis Challenges and Lessons Learned 10
MNCM Experience Beginning Provider Data Submission 12
MNCM Public Reporting of Composite Measures: Effect on Provider Participation 13
Preliminary, Un-validated Cycle 3 Measure Results 14
Comparison of Cycle 3 to Previous Reporting Cycles 17
Experiences with Producing Individual Physician Level Results 19
Other Issues and Lessons Learned 22
Conclusions 23

Appendices

Appendix A: BQI Pilot Measures ..... A 1-2
Appendix B: BQI Cycle 3 Measure Specifications ..... B 1-6

Executive Summary

The Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project is a special study funded by the Centers for Medicare & Medicaid Services (CMS) under the Quality Improvement Organization (QIO) 8th Statement of Work (SOW). The BQI project will analyze, evaluate and report the advantages and challenges of six separate regional pilot sites ranking the quality of outpatient ambulatory office medical care provided in their geographic regions. By completion of the BQI project in October 2008, each pilot site will have aggregated Medicare FFS administrative data with their other, regionally-derived data through at least three of four cycles. Cycle 1 covered data from Calendar Year (CY) 2005, Cycle 2 CY 2006, and Cycle 4 will report results based on CY 2007 data. Cycle 3 is intended to distill additional lessons associated with the some of the six pilots performing a semi-annual data aggregation.

The three pilot sites reporting semi-annual data aggregation in this report are CHIR, IHIE, and WCHQ. In addition, MNCM is in the process of performing semi-annual measure reporting on the most recent 12 months, July 2007 – June 2008. These data will be reported in the final BQI report. In this report, MNCM reports a rich experience with a new, parallel data aggregation method which consists of collecting data directly from providers, including lessons learned and contrasts with measures calculated from administrative data.

An executive summary of lessons, issues, and conclusions from Cycle 3 follows. MNCM results are reported in specified sections. Results in bullets below are consistent across the three other sites (CHIR, IHIE, WCHQ), unless otherwise specified. Delmarva takes complete responsibility for the conclusions stated below, as well as their supporting analysis and synthesis. We believe, however, that the conclusions accurately reflect the broad consensus of the four pilot sites participating in Cycle 3, as well as the other sites participating in the project.

Context of Cycle 3:

- Cycle 1 covered calendar year (CY) 2005 dates of service. All six BQI sites reported five quality measures.
- Cycle 2 covered CY 2006 dates of service. All six BQI sites reported an expanded set of twelve quality measures, five of which were consistent across all six pilots. (To view all the measures calculated across all cycles, please refer to Appendix A)
- For Cycle 3 CHIR, IHIE and WCHQ reported measures from a rolling twelve month data collection period, July 2006 through June 2007.
- In addition, MNCM reports its experience with implementing seven measures directly submitted by providers.

Measure Selection Challenges and Lessons Learned

- Cycle 3 did not result in measure selection issues for most of the sites.
- Measures that require enhanced administrative data, such as lab and pharmacy data, present additional challenges with data completeness and aggregation. (IHIE)

- MNCM selected five diabetes measures and four vascular measures for initial direct data submission by providers. Their providers preferred outcome measures over process measures, judging outcomes to be more relevant to their quality of care (see Appendix A for specific measures).

Data Aggregation & Analysis Challenges & Lessons Learned

- The work of the additional semi-annual aggregation stressed resources at all the pilots.
- WCHQ only reported seven of twelve measures, because their provider-supplied reporting is staggered semi-annually to accommodate the providers' workload.
- IHIE again found data completeness issues with lab and pharmacy data sources.

MNCM experience beginning Provider Data Submission

- Provider Direct Data Submission eliminates many issues of attribution and validation.
- The tradeoff of using this method is the need for auditing to assure accuracy of submitted measures. Education, training and monitoring of auditing staff and subcontractors is necessary.

MNCM Public reporting of composite measures, effect on Provider participation

- MNCM used "all-or-nothing" composite measures combining five diabetes measures into one diabetes score, and four vascular measures into one vascular score. These composite disease-specific measures were well received by the public, media, and payers.
- Providers continued to use individual measures for internal self-improvement.
- Media attention, public reporting, and the use of composite measures for pay-for-performance by Blue Cross/Blue Shield resulted in increased provider participation.

Comparison of Cycle 3 Results to Cycles 1 and 2

- Measures calculated on a rolling-twelve-month cycle were not meaningfully different from the same measures calculated on a calendar-year annual basis.

Issues Related to Individual Physician Level Calculation and Reporting

- Most of the pilot sites prefer clinic site or group level for public reporting.
- CHIR in Arizona is progressing with its plan to engage more physicians in validating their measures and using the information for quality improvement.

Other Issues and Lessons Learned

- Precise and consistent definition of every measure's numerator and denominator is essential to compare results. Slight changes of definition can have marked effects.
- Continuing communication and face-to-face meetings have been valuable in building a learning community among the pilots and deriving more lessons learned.

Take-home Conclusions Regarding Cycle 3

- Measures calculated from a semi-annual rolling twelve months do not appear to offer significant additional information over annual calendar-year measures.

- Aggregating measures more frequently causes significant workload issues.
- For most customers and most uses, annual reporting of quality measures current within 6-18 months is sufficient.
- Providers need some measures quarterly and more current for their own process improvement. These measures can be obtained from internal sources when needed.

Introduction and Common Themes

The Better Quality Information (BQI) to Improve Care for Medicare Beneficiaries Project is a special study funded by the Centers for Medicare & Medicaid Services (CMS) under the Quality Improvement Organization (QIO) 8th Statement of Work (SOW). The BQI project will analyze, evaluate and report the advantages and challenges of six separate regional pilot sites ranking the quality of outpatient ambulatory office medical care provided in their geographic regions.

A national system to measure the quality and value of health care would give patients and consumers the information they need to make informed decisions, as well as give care providers the measurements they need to manage and improve their performance. In addition to nationally recognized standard measures of quality, such a reporting system would require accurate and reliable collection of useful information. Experience in collecting and using medical quality information in the outpatient office setting is scarce.

During the BQI project, the six pilot sites will each report on twelve quality measures, five of which are consistent across all six sites. The sites use differing techniques for collecting data, various regional data bases, and independent self-designed ranking methodologies which pre-existed the BQI project. In spite of these differences, it is the goal of the BQI project to distill valuable lessons learned from the experiences of these sites, to leverage this experience by combining information from multiple public and private sources, and to transfer this knowledge to the next generation of medical quality measurement and reporting.

By completion of the BQI project in October 2008, each pilot site will have aggregated Medicare FFS administrative data with their other, regionally-derived data through at least three cycles. The first cycle reported five measures from each site, based on calendar year (CY) 2005 data (see Appendix A for specific measures). Cycle 1 was completed in November 2007. Cycle 2 of the BQI project was completed in Spring 2008 and each site reported results for twelve measures based on CY 2006 data (see Appendix A for specific measures).

In addition to the three cycles of data aggregation based on calendar years 2005- 2007, the BQI project asked the pilot sites to commit to performing a semi-annual cycle of aggregation and reporting. This semi-annual cycle of aggregation and reporting is referred to as “Cycle 3,” and is intended to investigate the issues surrounding reporting quality measures at more frequent intervals. The measurement period targeted for Cycle 3 was the twelve months spanning July 1, 2006 through June 30, 2007. The last BQI data aggregation cycle, which is based on calendar year 2007 data, is therefore referred to as “Cycle 4.”

Three of the six pilot sites- Center for Health Information & Research – Arizona State University (CHIR), Indiana Health Information Exchange (IHIE), and Wisconsin Collaborative for Healthcare Quality (WCHQ) committed to performing the Cycle 3 semi-annual data aggregation. The remaining three sites, California Cooperative Healthcare Reporting Initiative (CCHRI), Massachusetts Health Quality Partners (MHQP), and Minnesota Community Measurement (MNCM), initially declined to perform Cycle 3, largely because their methods of obtaining regional data were based on annual reports, making a six month evaluation impossible. MNCM, however, began during

2006 an additional method of local measure collection based on direct data submission from providers, and saw this as an opportunity to also attempt more frequent semi-annual reporting.

This Cycle 3 Data Aggregation Report has been prepared by Delmarva to provide a comprehensive picture of the activities, accomplishments, challenges and lessons learned from semi-annual aggregation. CHIR, IHIE, and WCHQ report their experience with aggregating data for a rolling twelve months July 2006- June 2007. The report also includes MNCM's experience with direct data submission, and provides their insights regarding the advantages of direct reporting by providers, the utility of composite measures for public reporting, and the effect of public reporting and media attention, as well as pay-for-performance initiatives, on the providers' desire to voluntarily participate in data collection and submission.

This Cycle 3 report contains the following sections:

- Measure Selection Challenges & Lessons Learned
- Data Aggregation & Analysis Challenges & Lessons Learned
- MNCM Experience Beginning Provider Data Submission
- MNCM Public Reporting of Composite Measures, Effect on Provider Participation
- Preliminary, Un-validated Cycle 3 Measure Results
- Comparison of Cycle 3 Results to Cycles 1 and 2
- Issues Related to Individual Physician Level Calculation and Reporting
- Other Issues and Lessons Learned
- Conclusions from the Cycle 3 Experience

Measure Selection Challenges and Lessons Learned

At the beginning of the BQI project, it was anticipated that the measures to be calculated would be taken from the list of AQA Starter Set of 26 Measures[1]. This list represents measures that aligned with agreed-upon parameters and addressed agreed-upon specific conditions and areas[2]. However, while these measures can be used to measure ambulatory care performance in general, the BQI project discovered that not all of them were applicable to the project. The Data Conquerors (a workgroup comprising the technical and analytical representatives from the pilot sites and Delmarva) reviewed each measure in detail for applicability to the project. In evaluating the measures, the BQI project had to consider the availability of data as well as the applicability of the measure to the Medicare population.

After it became clear that the AQA starter set of 26 measures would not provide sufficient measures for the project (each pilot site was required to select 12 for the entire project), the pilot sites and Delmarva worked together to identify additional measure sets that could be reviewed for applicability to the BQI project. The measure sets reviewed include the National Committee on Quality Assurance (NCQA) Health Plan Employer Data and Information Set (HEDIS®) measurement set (for which physician-level specifications were in draft) as well as other National Quality Forum (NQF) endorsed measures, such as those from the Physician Consortium for Performance Improvement (PCPI). In addition, several pilot sites (MNCM, WCHQ) opted to include their measure specifications, which were largely based on existing measure sets such as NCQA HEDIS®.

The BQI pilot sites were active participants throughout the measure selection process. Cycle 2 included a total of twenty-five measures across all six pilot sites. Each pilot continued and refreshed the five measures they reported in Cycle 1, and each site also calculated seven additional measures for Cycle 2. The selection of measures varies by pilot; however, as of Cycle 2 five measures are being calculated across all six pilot sites: Breast Cancer Screening, Colorectal Cancer Screening, Low-Density Lipoprotein (LDL) Testing for Diabetics, Hemoglobin A1c (HbA1c) Testing for Diabetics, and Nephropathy Testing for Diabetics. Each pilot's selection of their additional measures was based not only on data availability, but also relevance to and acceptance by the local community. For example, some pilot sites have access to laboratory results data either from the labs themselves (IHIE) or through their data collection methodology (WCHQ and MNCM), and are therefore able to include disease control measures in their list. The full list of measures that are being calculated can be found in Appendix A.

The two pilot sites calculating measures based on expanded administrative data (CHIR and IHIE) chose to refresh the results for the same twelve measures calculated during Cycle 2. For the two pilots sites calculating measures based on provider data submission, subsets of the Cycle 2 measures were chosen. Appendix A provides the details by pilot. All six pilot sites will calculate their selected twelve measures for the final cycle of data aggregation Cycle 4, covering calendar year 2007 data.

[1] http://www.aqaalliance.org/files/RevisedStarterSetApril2006.doc
[2] http://www.aqaalliance.org/files/AQAParametersforSelectingAmbulatoryCare.doc

For Cycle 3, WCHQ members stratified their community data into four payer categories. Identifying the payer has not been integral to the WCHQ data submission process prior to the BQI project, and many of their member organizations do not have a process for capturing payer information when performing data extraction. For Cycle 3, WCHQ developed a policy requiring members to report their data stratified by payer. This has allowed WCHQ to compare measure results for Medicare FFS patients to patients with other insurance in Cycle 3. To manage the administrative burden on the providers of performing this chart review in a payer-stratified manner, WCHQ only collected six diabetes measures and one uncomplicated hypertension measure for Cycle 3 (see Appendix A).

MNCM did not have any measure-related challenges or lessons learned for Cycle 3 reporting. CHIR had no new measure related issues for Cycle 3. They identified some data handling errors described in the data section below.

Data Aggregation and Analysis Challenges and Lessons Learned

The most significant data aggregation challenges for Cycle 3 were related to the workload of performing mid-year data analysis. In some cases, this required an extra cycle of provider reporting, and in other cases re-programming the date parameters for the data analysis. The four sites performing Cycle 3 data aggregation overcame this barrier in differing ways. CHIR and IHIE adjusted work process schedules. IHIE had to adjust the data run schedule for the rollout of their Quality Health First program to accomplish Cycle 3. WCHQ selected a smaller subset of their 12 quality measures for Cycle 3 reporting to limit the workload, and also adjusted some of their other work goals to accomplish the additional semi-annual data analysis. MNCM, as noted above, used its DDS process.

The primary challenge for CHIR during Cycle 3 was completing a large data run in a short time period. While they received the Cycle 3 data in May 2008, they were still finalizing the Cycle 2 analyses and reports and trying to improve physician engagement with validation of their data. Both of these efforts challenged their available resources and they were unable to begin Cycle 3 analysis until June. With an early July 2008 deadline for Cycle 3, CHIR felt that adding this cycle of data aggregation was very burdensome.

CHIR identified a significant data calculation issue during the running of Cycle 3 data. The denominators of several measures were noted to steadily increase through Cycles 1, 2, and 3. For Cycle 2 this had been attributed to data completeness issues, but as the denominators continued to grow significantly in Cycle 3 CHIR detected a programming error. In some cases the next Cycle's data was being added to the data from the previous cycles instead of over-writing the previous data. As a result the program code has been corrected and Cycle 2 data was re-calculated. CHIR completed Cycle 3 for all 12 measures calculated in Cycle 2. Comparing the first three cycles, CHIR noticed the measure results for Medicare FFS patients became more consistent with the all-payer measure results with each subsequent cycle.

IHIE is receiving very current clinical measures from its local sources, usually one to three months after service, and finds that the age of the 2006-2007 Medicare FFS data decreases its relevance to providers. IHIE feels that, as provider-submitted data becomes more complete and current, using administrative data from earlier time periods will have less relevance to patients and payers as well as providers.

IHIE completed Cycle 3 for the same twelve measures calculated in Cycle 2, some of which required clinical information. IHIE continued to experience the same issues (previously reported in Cycle 2) regarding those measures needing lab and pharmacy data: their admirable region-wide clinical database does not capture all area pharmacies and non-participating labs. IHIE finds that consumers have many alternatives for getting prescriptions filled that are not captured in claims systems. Retinal eye examinations for diabetics can be performed by multiple specialists in multiple settings and are difficult to document reliably. IHIE is still pursuing several regional laboratories that are not yet submitting their outpatient results into the region-wide clinical database.

Another issue IHIE identified is difficulty in keeping track of Medicare data received from CMS through the BQI data use agreements (DUAs), distinct from other data they receive for the same Medicare beneficiaries directly from their providers. This latter data is not restricted from feedback to providers to the degree that their DUA with Delmarva for the BQI project restricts the data from CMS. IHIE feels separating patient-specific data by source is problematic because of the DUA variations.

Because WCHQ collects its measures from provider-submitted all-payer data, asking providers to stratify their patients by payer is an extra administrative burden for them. WCHQ mitigates this by having providers submit data in a staggered manner, five measures in January and seven measures in July. For Cycle 3, covering six months of data, WCHQ did not budget with its data subcontractor, Centers for Health Systems Research and Analysis, for a partial-year data run on the CMS data. Therefore, the Cycle 3 results from WCHQ are based on WCHQ data only, stratified by payer. For Cycle 3 WCHQ received six additional months of data (for a total of 12 months) on six of its diabetes measures and one hypertension measure, which are identified in Appendix A. WCHQ feels that their Cycle 3 results lend increasing confidence to the conclusion that the quality of care provided to Medicare FFS beneficiaries does not vary greatly from that provided to commercially- insured patients.

Now that MNCM has completed two cycles of DDS for two consolidated measures, their next step will be speeding up the data collection cycle and beginning semi-annual reporting. To fulfill its commitment to Cycle 3 of the BQI project, MNCM will begin reporting Optimal Diabetes Care (ODC) and Optimal Vascular Care (OVC) measures for a rolling twelve months twice annually. During the month of July 2008, MNCM will collect provider data submissions for its first semi-annual report, covering dates of service 7/1/2007- 6/30/2008. The Cycle 3 results reported by MNCM below are for calendar year 2007 dates of service.

Direct Data Submission at MNCM and WCHQ

While direct data submission and reporting is relatively new to MNCM, WCHQ has always collected its measurement data based on provider submission of data. At WCHQ, reporting is done at the physician group level. At MNCM, direct data submission has allowed the site to collect and report data at the clinic site level. In previous data aggregation cycles, and historically, MNCM has only reported at the medical group level.

The MNCM and WCHQ direct data submission process is very similar, with two specific differences. MNCM receives patient-level, identified data from which it pulls a sample of patients for chart auditing and data abstraction validation. WCHQ receives group-level numerators and denominators for the measures, but also receives de-identified patient-level data for measure calculation validation.

MNCM Experience Beginning Provider Data Submission

Since 2002, MNCM has been calculating and publicly reporting quality measures for providers in the Minnesota region. Initially these measures were calculated from administrative data largely obtained from health plans, and validated by providers. In 2006 MNCM decided to develop an additional reporting system, called Direct Data Submission (DDS), which used clinical data extracted and submitted by medical groups to MNCM. There were several reasons MNCM wished to move to provider data submission.

MNCM's experience up to 2006 had shown strong interest by the providers in their own individual measures, for use for internal improvement processes, but patients had not shown as strong an interest in the public websites. This was felt to be because the measures were too technical for public relevancy, timeliness and complexity of presentation, and also because the measures were calculated at the corporate group level which in the Minnesota market often meant large groups with multiple clinic locations all being scored together. DDS was anticipated to be more timely, to be able to address validation and attribution directly since the data was submitted by the providers, and to allow data collection at the clinic site level. Including chart review also allowed selection of measures that were richer in clinical relevance to both providers and patients.

For the initial DDS effort, five measures for "Optimal Diabetes Care" (ODC) were selected. While data on the five separate measures was reported to the providers, for public reporting the five measures were rolled-up into an "all-or-nothing" composite measure for ODC. This single, composite measure was simpler to understand and preferred by patients and payers. The five ODC components were:

- LDL <100
- HbA1c <7
- BP <130/80
- Documented non tobacco user
- On aspirin if age <=40

The first three of these five measures are outcome measures. This move to prefer outcome over process measures when possible was a response to MNCM's provider stakeholders, who felt that outcome measures were a better reflection of quality care than process measures.

In order to address accuracy of data submission, MNCM created an audit process for participating sites. They used the NCQA "8 and 30" sampling methodology to review charts at each clinic. Eight records are initially selected randomly for review. If these eight records perfectly agree with the submitted data, the audit is closed. If there is an error detected in the first eight records, then 22 more records are randomly selected, with a total of 90% compliance (for all 30 records) required for participation. Corrective remedy plans at the clinic site level are expected for deficiencies, that is, discrepancies between source data found in the clinic sites' medical records and the data submitted to MNCM.

MNCM's first "pilot" for DDS was performed in the Spring 2007, for all 2006 dates of service. The initial data collection was performed on an Access database distributed to providers on CD. In this first, voluntary group, 20 medical groups with 191 clinic sites submitted data and passed validation audits. Four medical groups dropped out or failed validation in entirety, and one medical group had two clinics drop out. Altogether, fifteen (15) clinic sites either dropped out or did not pass the audit. 182 of the clinics had a denominator of diabetic patients larger than 29 and were included in the analysis. These clinics represented a population of 83, 611 Minnesota patients with diabetes.

Overall success with the "all-or-nothing" composite score for the 5 ODC measures was only 14.1%, with a range from 0% to 50%. There was significant pubic and media attention to this initial report. Problematic measures included achieving and documenting non-smoking status in diabetic patients, and accurately capturing aspirin use, since not only do aspirin-containing products have many names, they are usually not purchased through pharmacy benefits.

MNCM Public Reporting of Composite Measures: Effect on Provider Participation

These composite measures for diabetes and coronary disease were well- accepted by patients, payers and even providers because they were easy to understand, timely, and accurate. Providers continued to receive the non-compounded individual measure results as well for their internal process improvement.

As a result of the public attention, medical groups in the MNCM market that did not initially participate with this reporting found themselves under considerable pressure from their patients and payers to report their data as well. Therefore, in the spring of 2008 a second cycle of DDS was performed using ODC and an "Optimal Vascular Care" (OVC) composite measure.

For this second cycle the number of medical groups voluntarily extracting and reporting jumped from about 26 to 58, representing over 300 regional clinic sites and 127,000 diabetic patients. Average rate of success with all five diabetes measures increased slightly to 17.5 % (range, 0-48%).

MNCM performed a survey of participating groups to evaluate the relationship of having an electronic medical record (EMR) and optimal quality measures. Only twelve out of sixty-five medical groups, or about 18.5%, reported having an EMR in 2007. Groups with EMR tended to have higher scores, but the difference may not have been statistically significant.

Table 1 – Measure Rates for Groups Using Paper Charts and Groups with an EMR

Measure	Overall Mean Rate	Rate for Groups using Paper Charts	Rates for Groups with an EMR
Optimal Diabetes Care	17.2%	12.7%	19.8%
Optimal Vascular Care	33.1%	32.3%	34.7%

Preliminary, Un-validated Cycle 3 Measure Results

Despite the limitations and challenges encountered, the BQI pilot sites have been able to generate measure results, at either the physician group or individual physician level, and have provided their Cycle 3 results to Delmarva.

Descriptive statistics for these results are presented in Tables 2a-2d. Unless otherwise noted, the results presented are aggregated, multi-payer results. All results are not fully validated nor have the pilots completed all statistical testing (for validity and reliability).

Table 2a: Center for Health Information & Research, Cycle 3 Descriptive Statistics

Center for Health Information & Research (Individual Physician Level) Timeframe: July 1, 2006 – June 30, 2007							
Measure	Average Rate (1)	Standard Deviation	Rate Range	Denominator Range	Average Denominator	Number of Physicians Reporting (2)	Number of Physicians (3)
Retinal Eye Exam	31.6%	25.4%	0.0% - 100.0%	1 - 381	19	1,180	1,294
HbA1c Testing	68.6%	23.2%	0.0% - 100.0%	1 - 401	27	1,260	1,294
LDL Testing	77.8%	21.8%	0.0% - 100.0%	1 - 401	27	1,260	1,294
Breast Cancer Screening	54.0%	23.2%	0.0% - 100.0%	1 - 192	23	1,278	1,294
Colorectal Cancer Screening	25.6%	13.9%	0.0% - 92.0%	8 - 837	102	1,294	1,294
Nephropathy testing	32.4%	31.2%	0.0% - 100.0%	1 - 262	11	1,044	1,294
Depression Medication Management	19.4%	32.6%	0.0% - 100.0%	1 - 29	1	333	1,294
Cardiovascular –LDL testing	68.6%	28.9%	0.0% - 100.0%	1 - 69	8	1,173	1,294
Beta blocker treatment after heart attack	52.4%	46.1%	0.0% - 100.0%	1 - 6	0	209	1,294
Heart Failure: ACE inhibitor/ARB therapy	49.5%	34.2%	0.0% - 100.0%	1 - 40	3	901	1,294
Warfarin for patients w/CHF & atrial fibrillation	62.5%	41.3%	0.0% - 100.0%	1 - 10	1	585	1,294
CAD patients receiving lipid lowering therapy	41.2%	26.5%	0.0% - 100.0%	1 - 139	10	1,145	1,294

(1) Average of individual physician rates.
(2) Number of physicians with at least 1 patient in the denominator.
(3) Number of physicians included in Cycle 3 data file.

Table 2b: Indiana Health Information Exchange, Cycle 3 Descriptive Statistics

Indiana Health Information Exchange (Physician Group Level) Timeframe: July 1, 2006 – June 30, 2007							
Measure	Average Rate (1)	Standard Deviation	Rate Range	Denominator Range	Average Denominator	Number of Groups Reporting (2)	Number of Groups (3)
Breast Cancer Screening	67.8%	11.8%	56.0% - 100.0%	2 - 7,033	2,171	12	12
Colorectal Cancer Screening	46.0%	12.4%	25.0% - 65.0%	4 - 22,193	6,127	12	12
Beta-Blocker treatment after heart attack	4.1%	3.6%	0.0% - 10.0%	1 - 60	21	8	12
Persistence of beta-blocker therapy- Post MI	35.9%	26.5%	17.0% - 100.0%	1 - 57	29	8	12
Cardiovascular LDL-testing	68.1%	25.7%	0.0% - 100.0%	3 - 4,962	1,353	11	12
Cardiovascular LDL-control	19.0%	12.5%	0.0% - 39.0%	3 - 4,962	1,353	11	12
HbA1c testing	69.6%	15.2%	36.0% - 92.0%	2 - 6,536	1,994	11	12
LDL-testing	50.5%	21.8%	26.0% - 100.0%	2 - 6,536	1,994	11	12
Diabetes LDL cholesterol control	14.0%	7.1%	0.0% - 30.0%	2 - 6,536	1,994	11	12
Nephropathy testing	57.6%	15.4%	16.0% - 73.0%	2 - 6,536	1,994	11	12
Retinal Eye Exam	89.9%	6.3%	76.0% - 100.0%	2 - 6,536	1,994	11	12
HbA1c control	13.1%	9.5%	0.0% - 30.0%	2 - 6,536	1,994	11	12

(1) Average of physician group rates.
(2) Number of groups with at least 1 patient in the denominator.
(3) Number of participating IHIE groups.

Table 2c: Minnesota Community Measurement, Cycle 3 Descriptive Statistics

Minnesota Community Measurement (Physician Group Level) Timeframe: Calendar Year 2007									
Measure	Average Rate (1)	Standard Deviation	Rate Range		Denominator Range		Average Denominator	Number of Groups Reporting (2)	Number of Groups (3)
Optimal Diabetes Care	14.9%	8.0%	3.3%	33.3%	30	14,729	1,477	56	n/a
HbA1C control	55.3%	9.7%	32.8%	81.4%	11	9,240	840	56	n/a
Diabetes LDL control	52.9%	9.7%	26.7%	70.9%	8	9,483	863	56	n/a
Blood pressure control	47.3%	12.3%	18.8%	75.0%	7	8,104	799	56	n/a
Daily Aspirin Use	69.1%	11.9%	35.0%	97.9%	14	12,935	1,160	56	n/a
Documented Tobacco free	82.2%	8.1%	41.7%	90.7%	19	12,521	1,234	56	n/a

(1) Average of physician group rates.
(2) Number of groups submitting data.
(3) Not applicable for direct data submission.

Table 2d: Wisconsin Collaborative for Healthcare Quality, Cycle 3 Descriptive Statistics

Wisconsin Collaborative for Healthcare Quality (Physician Group Level) (1) Timeframe: July 1, 2006 – June 30, 2007									
Measure	Average Rate (2)	Standard Deviation	Rate Range		Denominator Range		Average Denominator	Number of Groups Reporting (3)	Number of Groups (4)
HbA1c testing	68.1%	6.5%	49.4%	76.8%	367	27,854	6,500	20	20
HbA1c control	47.9%	7.8%	31.1%	59.9%	367	27,854	6,500	20	20
LDL testing	82.9%	5.5%	66.7%	92.7%	367	27,854	6,500	20	20
Diabetes LDL cholesterol control	51.3%	5.1%	37.6%	61.7%	367	27,854	6,500	20	20
Blood pressure control	42.1%	5.0%	33.8%	51.9%	367	16,283	4,224	20	20
Nephropathy Testing	73.0%	9.3%	53.9%	86.7%	367	27,854	6,500	20	20
Hypertension-Blood Pressure Control	65.5%	3.7%	60.5%	72.5%	375	28,788	7,416	18	20

(1) All-payer results using WCHQ measurement methodology.
(2) Average of physician group rates.
(3) Number of groups submitting data.
(4) Number of WCHQ membership groups.

Comparison of Cycle 3 to Previous Reporting Cycles

For the most part, the pilot sites reported that Cycle 3 was similar to previous aggregation cycles.

IHIE continued to use its existing Quality Health First (QHF) measures and conducted its aggregation in the same manner. One specific difference was that Cycle 3 consisted of data from July 1 2006 to June 30, 2007. This data set is different from the QHF program monthly runs so this required an additional data specific run which utilized server space, programming time and efforts in addition to the work efforts for aggregation and programming of QHF monthly reports.

At WCHQ, data collection from its providers and subsequent measure calculation was not fundamentally different than other reporting cycles. Cycle 3 allowed CHIR to discover a glitch in its software, which has already been discussed, but like the others, their Cycle 3 experience was not different than previous aggregation cycles. While MNCM used a new data collection method to report measures for Cycle 3, this effort runs parallel with its other data aggregation efforts.

For the three pilots who used a twelve-month rolling measurement period, a review of the measure results shows no meaningful differences between the rates for Cycle 3 and those previously reported. Tables 3a-3c present the Cycle 2 and Cycle 3 results for each of these three pilots, for the measures that were consistent across both cycles.

Table 3a: Center for Health Information & Research, Cycle 2 and 3 Measure Results

Center for Health Information & Research (Individual Physician Level) - Cycles 2 and 3		
Measure	Cycle 2 CY 2006	Cycle 3 July 1, 2006 – June 30, 2007
Retinal Eye Exam	31.70%	31.60%
HbA1c Testing	66.60%	68.60%
LDL Testing	78.90%	77.80%
Breast Cancer Screening	56.30%	54.00%
Colorectal Cancer Screening	27.90%	25.60%
Nephropathy testing	33.40%	32.40%
Depression Medication Management	16.70%	19.40%
Cardiovascular –LDL testing	65.50%	68.60%
Beta blocker treatment after heart attack	41.80%	52.40%
Heart Failure: ACE inhibitor/ARB therapy	47.60%	49.50%
Warfarin for patients w/CHF & atrial fibrillation	57.00%	62.50%
CAD patients receiving lipid lowering therapy	46.20%	41.20%

Table 3b: Indiana Health Information Exchange, Cycle 2 and 3 Measure Results

Indiana Health Information Exchange (Physician Group Level) - Cycles 2 and 3		
Measure	Cycle 2 CY 2006	Cycle 3 July 1, 2006 – June 30, 2007
Breast Cancer Screening	65.30%	67.80%
Colorectal Cancer Screening	41.10%	46.00%
Beta-Blocker treatment after heart attack	13.00%	4.10%
Persistence of beta-blocker therapy- Post MI	39.50%	35.90%
Cardiovascular LDL-testing	68.10%	68.10%
Cardiovascular LDL-control	21.30%	19.00%
HbA1c testing	63.80%	69.60%
LDL-testing	42.40%	50.50%
Diabetes LDL cholesterol control	10.60%	14.00%
Nephropathy testing	52.50%	57.60%
Retinal Eye Exam	85.80%	89.90%
HbA1c control	11.50%	13.10%

Table 3c: Wisconsin Collaborative for Healthcare Quality, Cycle 2 and 3 Measure Results

Wisconsin Collaborative for Healthcare Quality (Physician Group Level) - Cycles 2 and 3		
Measure	Cycle 2 July 1, 2005 – June 30, 2006	Cycle 3 July 1, 2006 – June 30, 2007
HbA1c testing	68.80%	68.10%
HbA1c control	48.30%	47.90%
LDL testing	83.10%	82.90%
Diabetes LDL cholesterol control	51.30%	51.30%
Blood pressure control	39.20%	42.10%
Nephropathy Testing	73.70%	73.00%
Hypertension-Blood Pressure Control	62.20%	65.50%

Experiences with Producing Individual Physician Level Results

When the BQI project started, it was primarily focused on reporting performance measures at the individual physician level. Over the course of the project, interest in physician group-level reporting increased. Most of the pilot sites have expressed a strong preference for group-level reporting at the present time.

There are several reasons for this preference for group-level measures. First, two of the pilot sites, WCHQ and MNCM, collect measures at the group level and therefore cannot drill down the measures to individuals. Second, physician office care is often done in a group environment with individual physicians intentionally handing off specific types of care to other members of the group. This indicates that the group level is a more natural and appropriate level to report on than the individual physician level. Third, there are concerns that small denominator sizes at the individual physician level produce performance measures that are not reliable. Last, physicians who are beginning measure reporting are more comfortable with the group level and the pilot sites value and need the active involvement of the physicians in their efforts, in order to be successful.

Pilot Sites that committed to a non-patient identifiable provider reporting file at the individual physician level

As agreed in the Technical Proposal dated February 13, 2008, (Table 3-6, p.3-18) two of the pilot sites committed to creating a reporting file of non-patient identifiable provider measures at the level of the individual physician. With Cycle 2, both of these two pilot sites, CCHRI and CHIR, have calculated their measures at the physician level and have created reporting files at the individual physician level which have been submitted with this Cycle 2 report.

CCHRI has been calculating and giving feedback to providers at the individual physician level for several years. To date, their public reporting has been at the group level only. CCHRI was the only pilot site to express that the providers in their market were requesting a move to public reporting of measures at the individual physician level. The BQI project has afforded CCHRI the ability to test individual physician-level quality measure reporting. CCHRI is in the initial stages of this measure reporting effort. They have begun sharing individual-physician quality measure reports to the Medical Directors of a sample of physicians in California, as a way to pilot test the report process. The second phase of the feedback / reporting process will be a staggered mailing of reports to approximately 15,000 physicians. In both instances, responses and inquiries from the physicians are welcomed. The results of this pilot effort will be included in the final BQI report.

CHIR has completed measure calculation at the individual physician level, but has just begun validation and has not yet considered public reporting of the information. Public reporting in general has been an issue since the project's inception. CHIR has never engaged in public reporting before, and the climate in Arizona has not historically been conducive to public reporting. There has not been substantial change in this sentiment over the life of the BQI project. As a corrective plan to address the issue of public reporting, CHIR has employed multiple strategies to garner buy-in at the provider level and the institution level. CHIR has been working with the state medical associations for more than a year to establish buy-in and to help form their Expert Physician Panel. This

engagement process took months, and continues to be impacted by work schedules as well as the overall perception of the project and its relationship to pay-for-performance.

Despite the organization's neutral stance, they continue to be met with what appears to be a philosophical opposition to public reporting; there has been voiced concern that the Federal government will use this mechanism to reduce reimbursement for services rendered. CHIR has worked with its Expert Physician Panel to devise a process by which physicians may feel confident in result reporting. CHIR felt the idea of physicians leading each other through the process with CHIR's help would improve the chance of buy-in, as the panel would be able to demystify the process for its peers and their medical associations would be supportive of the effort. When CHIR assembled the panel, it was clear that there was a wide range of understanding about all of the different initiatives they had heard of--there was also an evident range of interest. Therefore, CHIR selected one physician who met their criteria for ability to turn around the results as well as being "on board" with the idea. CHIR is engaging this physician as a "physician champion." Having worked through the process with him, CHIR had him present to the Center stakeholders as someone who had "lived" through the process. CHIR has secured two more physicians on the panel who are going to participate in the process in July. All three will attend a meeting in July to discuss their experiences with other interested parties.

This validation process seems, for those who understand the overall purpose of the project, to be effective when working with a physician or physician group one at a time. Unfortunately, the scope of BQI (both time required to secure participation and conduct validation) will not allow for results to be reviewed with each physician or even each physician group in order for all participants to feel that their results have been adequately validated.

Pilot sites that committed to aggregating and reporting patient-identifiable feedback to providers at the individual physician level

In the Technical Proposal of February 13, 2008, (Table 3-6, p.3-18) two of the four remaining pilot sites that planned to create reporting files only at the physician group level still committed to aggregating and giving feedback to the providers at the individual physician level. Of these two sites, IHIE has successfully begun this process and has reported on its preliminary experience. WCHQ has identified logistical limitations to individual-level calculations, has analyzed the issues, and has planned and executed an alternative to individual-level reporting in their community.

At IHIE, the community feels that reporting must occur at the physician group level for several reasons:

1. The providers practice as a team, team effort is encouraged; therefore ratings should be for the team.
2. Physician extenders such as interns, residents, Nurse Practitioners, Physician Assistants and Nurse Midwives are not counted by CMS as providers, but assignment of the care provided by these individuals to any one member of the supervising team would be arbitrary and possibly distort the measurement of quality.

3. One patient may be attributed to multiple providers within a group, who all managed the patient at various times, and that this would result in this one patient's outcome having greater weight in the measures than a patient who was counted only once for the group.

WCHQ has identified some logistical limitations to fulfilling the commitment to report feedback to physicians at the individual physician level. The current process WCHQ uses to calculate measures is to receive data directly from providers as a numerator and denominator for each measure at the self-defined provider group level. Rates are then calculated and posted for public review on their website at www.wchq.org. As a result, WCHQ cannot calculate measure results at the level of the individual physician, but continues to be committed to exploring the issues surrounding individual physician-level reporting

A second logistical issue is that each of the provider organizations participating with WCHQ has developed their own internal methodologies for attribution of patients to physicians within their groups. These attribution methodologies have not been reviewed or standardized by WCHQ. Trying to obtain this data from the organizations for the BQI project would therefore result in inconsistent data that would not be comparable site to site.

Based on the information WCHQ obtained in a literature review, their leadership is not fully convinced that measurement and reporting at the individual physician level have sufficient scientific rigor at this time. The WCHQ Board of Directors (BOD) is particularly concerned about some of the methodological and technological issues that come about when discussing the power of measures to assess individual performance due to sample sizes, sensitivity of measures and specificity of measures in small patient populations.

In addition, the WCHQ BOD feels strongly that the work of their organization has consistently promoted the idea of a care team, and that assessment of individual performance within this care team undermines the system of accountability to the team.

Because of the above concerns WCHQ does not intend to calculate measures and give provider feedback at the individual physician level at this time. As an alternative plan to address their commitment, WCHQ planned and sponsored a conference with the AHRQ to address the issues at the core of this decision. WCHQ feels this work will move the agenda forward toward building consensus and establishing some guidelines for future individual physician measurement. At this conference, there was agreement that there was value in physician level *measurement*, however, there was still dissention about the public release of such information. In addition, the participants agreed that a team approach to care giving is important to understand. However, more areas of further research existed than areas of agreement. Some of these areas of further research identified include:

- Regarding attribution: Is individual physician an adequate model, or a concept with individuals in separate practices? How can this be addressed?
- Is there a way to simplify measurement and reporting?
- How do we create a bridge between claims and clinical data?

- How do we develop standards that can be applied to the diverse physician practices across the US?
- How do we address the regional variations in coding?
- Have we done adequate research on the size of a practice and its impact on measure results?

Pilot sites that committed to reporting and giving feedback at the physician group level only

The remaining two pilot sites, MHQP and MNCM, have fulfilled their commitment in the Technical Proposal of February 13 (Table 3-6, p.3-18) to aggregate, report feedback to providers, and create provider reporting files at the group level only.

MHQP has not previously done physician-level reporting for any of its annual HEDIS® measure reports. Though the measures are calculated down to the individual physician level, they are rolled up to a provider-defined group level for analysis and reporting. The MHQP Physician Council has concerns about the validity and reliability of the data at the individual physician level.

Based on its experience, however, MHQP believes that improvements in quality are more related to the organizational structure within which a physician practices than to the individual physician. Their reasoning regarding the delivery of care by teams and groups is similar to that reported by other pilot sites.

MNCM also receives its data directly from provider groups as a group-level numerator and denominator, similar to the system at WCHQ. For MNCM, the only means for reporting at the individual physician level would be through the expansion of their direct data submission process, and it would also require significant resources at the medical group level.

Beyond that, for the measures MNCM currently reports, their philosophy is that the unit of analysis and quality improvement is the clinic-site level, and they have recently considered moving towards a "team level." MNCM's chronic care measures, especially, are measures of a care system at a given medical group.

Other Issues and Lessons Learned

Three of the four pilot sites reporting for Cycle 3 mentioned, as an important lesson, the precision necessary in defining both the numerators and denominators of each quality measure, and making them consistent across sites if comparisons are to be valid. This was an issue agreed to by all six pilot sites at our recent Cycle 3 face-to-face meeting. Small variations in definitions can make significant differences in either the numerator, denominator, or both.

IHIE and WCHQ separately commented on the benefits they perceived from clear communication and an increasing sense of community and exchange among the stakeholders of the BQI project. Both found face-to-face meetings a beneficial addition to web-based and conference call interactions. Delmarva has found that actively

soliciting input from all sites on shared issues combined with interactive face-to-face meetings has effectively created frank and productive discussions and clarified issues and lessons learned.

Conclusions

Cycle 3 data reporting was placed into the BQI project among the other, calendar-year-based cycles to study the issues and effects of performing data aggregation on a more-frequent-than-annual cycle. Delmarva believes the results presented in this Cycle 3 report support the following conclusions regarding semi-annual or mid-year reporting:

1. Measures calculated from a semi-annual rolling twelve months do not appear to offer significant additional information over annual calendar-year measures. The reasons for some observed variations in both directions will require further study, but at this point seem related to technicalities of definitions of specific measures, data collection cycle issues, or both.
2. Aggregating measures more frequently causes significant workload issues. Whether the site in question collected data from administrative databases, provider sources, or both, the mid-year aggregation was essentially as much work as the annual calculation. As electronic medical records become more sophisticated and complete, this problem may ameliorate.
3. The final point remains, however, that for most customers and most uses, annual reporting of quality measures current within six to eighteen months is sufficient. Providers need some measures quarterly and more current for use in their own process improvement. These measures can in most cases be obtained from internal sources when they are needed.

Appendix A: BQI Project Measures

1 = Measure calculated during Cycle 1 aggregation
2 = Measure added during Cycle 2 aggregation
3 = Measure calculated during Cycle 3 aggregation
4 = Measure added during Cycle 4 aggregation[1]

Measure	Specifications	NQF endorsed	AQA Starter Set	AZ	CA	IN	MA	MN[2]	WI[3]
PREVENTION									
Breast Cancer Screening	**HEDIS®– BCS**	**Y**	**Y**	**1,3**	**1**	**1,3**	**1**	**1**	**1**
Colorectal cancer screening	**HEDIS®– COL**	**Y**	**Y**	**1,3**	**1**	**1,3**	**1**	**2[4]**	**1**
Pneumonia vaccination in 65+ yrs pop	CMS / NCQA	Y	N						2[5]
DIABETES									
LDL testing	**HEDIS®– CDC[6]**	**Y**	**Y**	**1,3**	**1**	**1,3**	**1**	**1**	**1**
Diabetes LDL cholesterol control	HEDIS®– CDC[1]	Y	Y			2,3	4	2	2
HbA1c testing	**HEDIS®– CDC[1]**	**Y**	**Y**	**1,3**	**1**	**1,3**	**1**	**1**	**1**
HbA1c control	HEDIS®– CDC[1]	Y	Y			2	4	2	1
Nephropathy testing	**HEDIS®– CDC[1]**	**Y**	**N**	**2,3**	**2**	**1,3**	**2**	**1**	**2**
Retinal eye exam	HEDIS®– CDC[1]	Y	Y	1,3		2,3	1	1	
Blood pressure control	HEDIS®– CDC[1]	Y	N					2	2
Optimal diabetes composite	MNCM	N	N					2[7]	
HbA1C control	MNCM	N	N					3[8]	
Diabetes LDL control	MNCM	N	N					3[8]	
Blood pressure control	MNCM	N	N					3[8]	
Daily Aspirin Use	MNCM	N	N					3[8]	
Documented Tobacco free	MNCM	N	N					3[8]	
Diabetes – smoking status	MNCM	N	N					2	
Diabetes – daily aspirin use	MNCM	N	N					2	
Optimal Vascular Control	MNCM	N	N					3[8]	
CARDIOVASCULAR CONDITIONS									
Cardiovascular – LDL testing	HEDIS–CMC[9]	Y	N	2,3	1	2,3	2		2
Cardiovascular – LDL control	HEDIS–CMC[2]	Y	N			2,3			2
CAD patients receiving lipid-lowering therapy	PCPI	Y	Y	2,3	2		2		

[1] All measures marked calculated during Cycles 1 or 2 are calculated during Cycle 4 unless specifically noted as Cycle 3 or Cycle 4 additions by the use of "3" or "4."
[2] MNCM is calculating HbA1C control, LDL control, and BP control based on the direct data submission process.
[3] WCHQ will use HEDIS® specifications for Medicare FFS data analysis, and WCHQ specifications for aggregated data.
[4] MN will collect this measure via Hybrid data collection method.
[5] Administrative and EMR data only, no survey.
[6] CDC = Comprehensive Diabetes Care. This measure is one of those included in this category.
[7] MNCM Optimal Diabetes Composite: HbA1C control <7.0, LDL – control <100, BP control (collect <130/80 and <140/90), Daily Aspirin Use, Documented Tobacco free
[8] Cycle 3 data collection via direct data submission by providers
[9] CMC = Cholesterol Management for Patients with Cardiovascular Conditions. This measure is one of those included in this category.

Measure	Specifications	NQF endorsed	AQA Starter Set	AZ	CA	IN	MA	MN[2]	WI[3]
HEART ATTACK / HEART FAILURE									
Beta blocker treatment after heart attack[10]	HEDIS®– BBH	Y	Y		2	2,3	2		
Persistence of beta blocker therapy – Post MI	HEDIS®– PBH	Y	Y	2,3	2	2,3	2		
HF: LVF testing	PCPI	Y	Y		2				
Heart failure: ACE inhibitor/ARB therapy	PCPI	Y	Y	2,3					
Warfarin for patients w/ CHF & atrial fibrillation	PCPI	Y	N	2,3	2				
MEDICATION MANAGEMENT									
Antidepression Medication Management – Acute Phase	HEDIS®– AMM	Y	Y	2,3					
Annual monitoring for patients on persistent medications: Ace inhibitors, Digoxin, Diuretics, Anti-convulsants, Statins	HEDIS®– MPM	Y	N		2[11]		2[12]		
OTHER									
Hypertension – blood pressure control	HEDIS®– CBP	Y	N						2
Use of Spirometry Testing in Assessment/Diagnosis of COPD	HEDIS®– SPR	N	N				2		

[10] 2008 HEDIS® specifications for public comment indicate potential retirement of this measure (http://www.ncqa.org/Programs/PublicComment/HEDIS2008/index.htm)

[11] Annual monitoring for patients on persistent medications: CCHRI:5- Components: Diuretics, Ace inhibitors, Digoxin, Anti-Convulsants, Statins

Cycle 1 based on 2005 data, HEDIS® 2006; Cycle 2 based on 2006 data, HEDIS® 2007

[12] Annual monitoring for patients on persistent medications: MHQP: ARBs & Ace inhibitors, calculated separately but reported together.

Appendix B:
BQI Cycle 3 Measures, by Pilot

	Nephropathy Testing	Retinal Eye Exam	Diabetes LDL testing	Diabetes LDL control	Diabetes HbA1c Testing	Diabetes HbA1c control	Diabetes Blood Pressure Control	Optimal diabetes composite	Diabetes – smoking status	Diabetes – daily aspirin use
Center for Health Information & Research (CHIR)										
Age Range	<70 OR >15	<70 OR >15	<70 OR >15	x	<70 OR >15	x	x	x	x	x
Measurement period	1 year (2 tests)	1 year (2 tests)	1 year (2 tests)	x	1 year (2 tests)	x	x	x	x	x
Denominator period	24 months	24 months	24 months	x	24 months	x	x	x	x	x
Specifications	Symmetry EBM Connect (http://www.symmetry-health.com/products/product_SEBM.php), possible algorithm modification									
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care, Medicaid FFS; Medicaid Managed Care									
Specialties / Medical Professionals	Attribution to Primary Care Physicians (Doctors of Medicine, Doctors of Osteopathy) in Internal Medicine, Family Practice, Pediatrics, Geriatrics, and OB/GYN									
Indiana Health Information Exchange (IHIE)										
Age Range	18-75	18-75	18-75	18-75	18-75	18-75	x	x	x	x
Measurement period	1 year	1 year	1 year	1 year	1 year	1 year	x	x	x	x
Denominator period	2 years	2 years	2 years	2 years	2 years	2 years	x	x	x	x
Specifications	Administrative. HEDIS® 2007, Medicare Care Management Performance Demonstration Quality Measurement & Reporting Specifications (HF, CAD measures), PCPI (Osteoporosis Management)									
Population covered	Commercial PPO (3 plans); Medicare FFS									
Specialties / Medical Professionals	Attribution to MDs (PCPs, and excluding such MDs such as radiology, etc), DOs, Nurse Practitioners and Physician Assistants. Specialties as per measure specifications, including primary care (family practice, internal medicine), OB/GYN, endocrinology, cardiology, gastroenterology.									
Minnesota Community Measurement (MNCM)										
Age Range	x	x	x	18-75	x	18-75	18-75	18-75	18-75	18-75
Measurement period	x	x	x	1 year	x	1 year	1 year	1 year	1 year	1 year
Denominator period	x	x	x	2 years	x	2 years	2 years	2 years	2 years	2 years
Specifications	MNCM									
Population covered	All populations - provider direct data submission									
Specialties / Medical Professionals	All as appropriate.									

Appendix B:
BQI Cycle 3 Measures, by Pilot

	Cardiovascular LDL testing	Cardiovascular LDL control	CAD patients receiving lipid-lowering therapy	Beta blocker treatment after heart attack	Persistence of beta blocker treatment post-MI	HF: LVF testing	HF: ACEI/ARB therapy	Warfarin for patient w/CHF and atrial fibrillation	Antidepression medication management
Center for Health Information & Research (CHIR)									
Age Range	18-75	x	>19	x	>35	x	>35	>19	> 18 2/3 yrs
Measurement period									
	1 year	x	1 year	x	1 year	x	1 year	1 year	20 months
Denominator period	24 months	x	1 year	x	1 year	x	1 year	1 year	1 year
Specifications									
	Symmetry EBM Connect (http://www.symmetry-health.com/products/product_SEBM.php), possible algorithm modification								
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care, Medicaid FFS; Medicaid Managed Care								
Specialties / Medical Professionals									
	Attribution to Primary Care Physicians (Doctors of Medicine, Doctors of Osteopathy) in Internal Medicine, Family Practice, Pediatrics, Geriatrics, and								
Indiana Health Information Exchange (IHIE)									
Age Range	18-75	18-75	x	>=35	>=35	x	x	x	x
Measurement period	1 year	1 year	x	1 year	1 year	x	x	x	x
Denominator period	2 years	2 years	x	18 months	18 months	x	x	x	x
Specifications									
	Administrative. HEDIS® 2007, Medicare Care Management Performance Demonstration Quality Measurement & Reporting Specifications (HF, CAD measures), PCPI (Osteoporosis Management)								
Population covered	Commercial PPO (3 plans); Medicare FFS								
Specialties / Medical Professionals									
	Attribution to MDs (PCPs, and excluding such MDs such as radiology, etc), DOs, Nurse Practitioners and Physician Assistants. Specialties as per measure specifications, including primary care (family practice, internal medicine), OB/GYN, endocrinology, cardiology, gastroenterology.								
Minnesota Community Measurement (MNCM)									
Age Range	x	x	x	x	x	x	x	x	x
Measurement period	x	x	x	x	x	x	x	x	x
Denominator period	x	x	x	x	x	x	x	x	x
Specifications	MNCM								
Population covered	All populations - provider direct data submission								
Specialties / Medical Professionals	All as appropriate.								

Appendix B:
BQI Cycle 3 Measures, by Pilot

	Breast cancer screening	Colorectal cancer screening	Annual monitoring for patient on persistent meds	Hypertension – blood pressure control	Use of spirometry testing	Osteoporosis management	Pneumococcal vaccination
Center for Health Information & Research (CHIR)							
Age Range	52-69	51-80	x	x	x	x	x
Measurement period							
	2 years	10 yrs if poss	x	x	x	x	x
Denominator period	2 years	10 yrs if poss	x	x	x	x	x
Specifications	Symmetry EBM Connect (http://www.symmetry-health.com/products/product_SEBM.php), possible algorithm modification						
Population covered	Commercial (HMO, POS, PPO); Medicare FFS; Medicare Managed Care, Medicaid FFS; Medicaid Managed						
Specialties / Medical Professionals	Attribution to Primary Care Physicians (Doctors of Medicine, Doctors of Osteopathy) in Internal Medicine, Family Practice, Pediatrics, Geriatrics, and OB/GYN						
Indiana Health Information Exchange (IHIE)							
Age Range	52-69	50-80	x	x	x	x	x
Measurement period	2 years	120 months	x	x	x	x	x
Denominator period	2 years	10 yrs if poss	x	x	x	x	x
Specifications	Administrative. HEDIS® 2007, Medicare Care Management Performance Demonstration Quality Measurement & Reporting Specifications (HF, CAD measures), PCPI (Osteoporosis Management)						
Population covered	Commercial PPO (3 plans); Medicare FFS						
Specialties / Medical Professionals	Attribution to MDs (PCPs, and excluding such MDs such as radiology, etc), DOs, Nurse Practitioners and Physician Assistants. Specialties as per measure specifications, including primary care (family practice, internal medicine), OB/GYN, endocrinology, cardiology, gastroenterology.						
Minnesota Community Measurement (MNCM)							
Age Range		x	x	x	x	x	x
Measurement period		x	x	x	x	x	x
Denominator period		x	x	x	x	x	x
Specifications	MNCM						
Population covered	All populations - provider direct data submission						
Specialties / Medical Professionals	All as appropriate.						

Appendix B:
BQI Cycle 3 Measures, by Pilot

	Nephropathy Testing	Retinal Eye Exam	Diabetes LDL testing	Diabetes LDL control	Diabetes HbA1c Testing	Diabetes HbA1c control	Diabetes Blood Pressure Control	Optimal diabetes composite	Diabetes – smoking status	Diabetes – daily aspirin use
Wisconsin Collaborative for Healthcare Quality (WCHQ)										
Age Range	18-85	x	18-85	18-85	18-85	18-85	18-85	x	x	x
Measurement period	1 year	x	1 year	1 year	1 year	1 year	1 year	x	x	x
Denominator period	2 years	x	2 years	2 years	2 years	2 years	2 years	x	x	x
Specifications										
	WCHQ. Medical groups/providers determine denominators as per WCHQ methodology, and then provide numerator information.									
Population covered	All payers									
Specialties / Medical Professionals	PAs, NPs, DOs, MDs.									

Appendix B:
BQI Cycle 3 Measures, by Pilot

	Cardiovascular LDL testing	Cardiovascular LDL control	CAD patients receiving lipid-lowering therapy	Beta blocker treatment after heart attack	Persistence of beta blocker treatment post-MI	HF: LVF testing	HF: ACEI/ARB therapy	Warfarin for patient w/CHF and atrial fibrillation	Antidepression medication management
Wisconsin Collaborative for Healthcare Quality (WCHQ)									
Age Range	x	x	x	x	x	x	x	x	x
Measurement period	x	x	x	x	x	x	x	x	x
Denominator period	x	x	x	x	x	x	x	x	x
Specifications									
	WCHQ. Medical groups/providers determine denominators as per WCHQ methodology, and then provide numerator information.								
Population covered	All payers								
Specialties / Medical Professionals	PAs, NPs, DOs, MDs.								

Appendix B:
BQI Cycle 3 Measures, by Pilot

	Breast cancer screening	Colorectal cancer screening	Annual monitoring for patient on persistent meds	Hypertension – blood pressure control	Use of spirometry testing	Osteoporosis management	Pneumococcal vaccination
Wisconsin Collaborative for Healthcare Quality (WCHQ)							
Age Range		x	x	18-85	x	x	x
Measurement period		x	x	1 year	x	x	x
Denominator period		x	x	2 years	x	x	x
Specifications	WCHQ. Medical groups/providers determine denominators as per WCHQ methodology, and then provide numerator information.						
Population covered	All payers						
Specialties / Medical Professionals	PAs, NPs, DOs, MDs.						

Appendix 11

Data Feedback to Providers: Experience and Lessons Learned

Better Quality Information to Improve Care for Medicare Beneficiaries

(BQI) Project

Data Feedback to Providers: Experience and Lessons Learned

July 22, 2008

Introduction 2
Summary of Experience to Date 2
Center for Health Information and Research (CHIR) 3
California Cooperative for Healthcare Reporting Initiative (CCHRI) 5
Indiana Health Information Exchange (IHIE) 6
Massachusetts Health Quality Partners (MHQP) 7
Minnesota Community Measurement (MNCM) 9
Wisconsin Collaborative for Healthcare Quality (WCHQ) 10
Conclusion 10
Appendix A – IHIE QHF Validation ProcessExamples of Letters Sent to Providers 11
Appendix B – MHQP Examples of Letters Sent to Providers 13
Appendix C – MNCM Policy and Procedure: Appeals of Clinical Data and Results 16

Introduction

Data feedback to providers is a critical component of a successful physician level performance measurement and reporting initiative. The sharing of results with providers before it is reported allows physicians the opportunity not only to understand the process, but to buy into it as well. Because the ultimate goal of the Better Quality Information to Improve Care for Medicare Beneficiaries (BQI) project is to improve the quality of care for Medicare beneficiaries, data feedback to providers is a principle driver for achieving this objective. The six BQI pilot sites through their past and current experience with performance measurement and/or public reporting understand the need for this critical feedback loop. Working with a diverse group of pilot sites, with varying degrees of experience, allows the BQI project to explore different methods of giving feedback to providers as well as the challenges and lessons learned with doing so. This document explores the BQI pilot sites' experiences around provider data feedback and planned activities under the project. The BQI pilot sites included within this summary are:

- Center for Health Information and Research – Arizona State University (CHIR)
- California Cooperative for Healthcare Reporting Initiative (CCHRI)
- Indiana Health Information Exchange (IHIE)
- Massachusetts Health Quality Partners (MHQP)
- Minnesota Community Measurement (MNCM)
- Wisconsin Collaborative for Healthcare Quality (WCHQ)

Summary of Experience to Date

The BQI pilot sites have varying levels of experience with provider feedback. Methods for data feedback to providers range from 1) individual physician interaction with validating the results to 2) sharing the findings through forums with the physicians and other stakeholders before release, or 3) a compromise somewhere in between the spectrum of possible methods. It is also important to emphasize that the communities in which the pilot sites operate also influence the selected method(s). In the context of the BQI project, the pilot sites had to work within the confines of the QIO regulations for sharing information with providers. This also had an impact on the method(s) used during the project. The implementation of and experience with data feedback to providers for each BQI pilot site is described below.

Three of the pilots -- MHQP, MNCM, and WCHQ -- have been engaged in data feedback activities, at some level, for at least four years. IHIE began planning for the data feedback process in 2006 as a part of their Quality Health First SM program. CHIR and CCHRI engaged in the planning and implementation of data feedback to providers as a part of the BQI project. As the organizations' efforts around performance measurement reporting and provider feedback have evolved, so has their ability to share lessons learned and best practices in the areas of development, implementation, evaluation, physician engagement, and provider outreach.

Center for Health Information and Research – ASU (CHIR)

Public reporting in general has been an issue for CHIR since the project's inception. CHIR has never engaged in public reporting before and the climate in Arizona has not historically been conducive to public reporting; there has not been substantial change in this sentiment over the life of the project.

CHIR has never publicly reported non-aggregated de-identified results of any fashion largely due to the nature of their relationships with those who voluntarily provide them with data. Because of the nature of this relationship, there was never a provider data feedback loop to ensure the accuracy of the data provided. In contrast to many of the other BQI pilot sites who have a history of measuring quality of care and reporting results, CHIR does not. The advantage CHIR has brought to the BQI project has been their skill in working with large data and the opportunity to move the Arizona community forward towards transparency in healthcare.

Multiple strategies have been employed to garner buy-in at the provider and institution level to move forward with public reporting, which in turn would include a mechanism for data feedback to providers before reporting of results. Despite CHIR's neutral stance, they continue to be met with what appears to be a philosophical opposition to the idea – there has been some concern voiced that the Federal Government will use this reporting mechanism to reduce reimbursement for services rendered.

To begin the process of data feedback to providers, CHIR initiated a process to test (i.e., validate) the results achieved from calculating the measures on administrative data against what the provider may glean from his/her clinical record data. It is through this validation process that CHIR will be able to demonstrate to their provider community the accuracy of the rates being measured by administrative data. More than a year ago, CHIR began working with the state medical associations to establish buy-in for the project and to help seat their Expert Physician Panel, beginning the organization's process for data feedback to providers. This engagement process took months and continues to be impacted by the work schedules of the participating physicians, overall perception of the project, as well as, its perceived relationship to pay-for-performance. Still, CHIR has managed to consistently engage several physicians enough that they are willing to work through piloting CHIR's physician validation process. It is the physicians' belief that they need to have real results that are tangible to individual physicians in order to address the mentality that these types of endeavors happen "to" them and not "with" them.

Physicians participating in the pilot validation process are first asked to enter into a data sharing agreement with CHIR. Once this agreement is signed, the physician is provided a list of their attributed patients obtained through the CMS data to validate. The purpose of this is to validate the patient-to-provider attribution process and to determine whether there is sufficient agreement between the attribution methodology used by CHIR and the provider's record. Secondly, the physician is provided with indicators of

whether each patient met the measurement criteria and they are asked to review their medical records for that patient and indicate agreement/disagreement with each indicator. For example, the physician report would read Patient X received breast cancer screening and the physician would be asked to agree or disagree based on the information they have in the patient's medical record. The physicians that are currently participating in the validation process use electronic health records (EHRs) to validate the data. It is CHIR's intent to go through the validation with the piloting physicians and have the piloting physicians help them further engage their peers as physician leaders.

CHIR believes they have devised a process by which physicians may feel increasingly confident in the reporting of results. The physicians that buy into the process may, however, divide into those who feel comfortable about it in theory (probably the largest number) and those who feel comfortable with the practical application of the validation process given these caveats: 1) when reporting is done by a trusted entity, 2) when physician is allowed to review/preview/amend (data feedback to providers), and 3) when the physician is incentivized for participating in the validation process.

Significant progress has been made to get physicians to buy-in to the concept of data feedback to providers for the purposes of reporting results. One of the most significant successes that has helped CHIR move towards the goal of increased participation in data feedback to providers was the establishment of the Expert Panel of Physicians. Recognizing that there has not been any concerted, sustained effort in CHIR's region to participate in an initiative similar to BQI, constructing a group of physician leaders was both a first and significant step. However, a great amount of progress is still required for CHIR to widely review results with each physician or even each physician group. It will take buy-in from the physician leaders and their subsequent sharing of the benefits of the data feedback process to their colleagues to garner wide support and participation in this process. CHIR will continue to work towards this goal, but realizes that this will not occur before the end of the BQI Project.

One of the most important lessons CHIR learned through this process is that it takes time to get all the players to the table and to gain their trust and buy-in. Asking for physician time, even when incentivized, is extremely difficult given the many demands on their schedules – many on the CHIR Expert Panel routinely work on call and overnights in emergency care and all are actively engaged in direct care. The physicians participating in the validation process are compensated financially by CHIR for the administrative burden from participation in the pilot. The most effective method for helping physicians to understand the project and the process has been individualized contact. This has proven difficult for CHIR to manage as they juggle other project tasks and responsibilities. To support this work, CHIR has hired a Physician Liaison, who is a well-respected leader in the community, to assist on both provider and consumer engagement issues.

Ultimately, it is CHIR's intent to comply with posting of results to the degree with which CHIR can, given both their organization and their stakeholders' concerns regarding validity of results.

CHIR plans to post de-identified results on a University-sponsored website and explain how a beneficiary may use this type of reporting to gauge physician performance once the validation efforts with their physician leaders is complete. The website will also provide some context about the BQI project, quality measurement, etc. to accompany the results. A link to the CMS BQI website will also be included. CHIR anticipates that the website will be available on or around September 2008. Since the website will be sponsored by the university, CHIR does not have internal control over changes to the website. Therefore, they must work with other University personnel to do so which may cause some delay.

California Cooperative for Healthcare Reporting Initiative

CCHRI recently began the process of providing data feedback report to physicians for the Cycle 2 results. The report includes information on 11 quality measures – 3 Medicare Care Management Performance Demonstration and Quality Measures and 8 HEDIS 2006 measures. As a part of the BQI Project, CCHRI was required to calculate and report on 12 measures. However, they are only reporting on 11 measures because they found the Beta Blocker after a Heart Attack measure unreliable. CCHRI will continue to collect data on this measure and monitor it to see if it is able to be reported in the future.

CCHRI decided to establish the minimum patient threshold as the sample size for which 90 percent of physicians had at least 70 percent reliability. Therefore, they computed the observed 10th percentile of estimated reliability for each sample size n, and identified the value of n for which the 10th percentile of reliability was 0.70. This value of n was employed as the minimum patient threshold.

Physicians receive three types of score reporting for each measure:

1) Percentile rank compared to specialty peers, by measure:
    a. Displays how each physicians' measure scores compares to all other within their same specialty on a 1-100 scale.
    b. Scores calculated for physicians meeting the minimum patient threshold are shaded blue for easy identification. Physicians' not meeting the minimum patient threshold for measures are still shown, but in gray and are notified that they may not accurately reflect performance.

2) Performance score reported separately, by measure:
    a. Displays each physician's individual performance measure scores.
    b. Scores calculated for physicians meeting the minimum patient threshold are shaded blue for easy identification. Physicians' not meeting the minimum patient threshold for measures are still shown, but in gray and are notified that they may not accurately reflect performance.

3) Performance score, stratified by payer (i.e., Medicare vs. Commercial Patients):
    a. The measures include Medicare fee-for-service and commercial PPO patients.
    b. The scores for each patient group are reported separately.

c. Scores are reported at the individual physician level.
d. There may be too few patients in one or both of these patient subgroups to have reliable scores.

The initial phase of report distribution will be to the Medical Directors of a sample of physician groups in California. The Medical Directors will be used to pilot the process before the reports are sent out on a larger scale to individual physicians in California. The Medical Director will receive his/her individual report outlining their results during the specified period of service. The reports will contain the physician's numerator, rate, and rank compared to all other physicians within their specialty. The reports will be distributed as a PDF file via e-mail in late June. CCHRI is requesting comments/feedback from the Medical Directors by early July. A response to all comments/feedback will be addressed within 2 business days of the request.

The second phase of report distribution will be a staggered mailing to approximately 15,000 PPO physicians. These physicians will receive results at the individual physician level. This mailing will occur between July and August 2008. Physicians will be able to send comments/feedback to a designated email address where all inquiries will be addressed. CCHRI is currently planning to make sure all inquiries from physicians are answered in a timely fashion.

CCHRI is looking forward to hearing back from the Medical Directors and physicians on the data feedback reports. This is CCHRI's first experience with data feedback to providers at this level. Information on the success of this initiative as well as the lessons learned will be shared in future reports.

Indiana Health Information Exchange

IHIE has been sharing data with physicians through its Quality Health First SM Program (QHF) for approximately one year. This program utilizes the BQI Cycle 1 and Cycle 2 measures, in addition to some other measures that are exclusive to the QHF Program. Therefore, the experience in sharing data with physicians through the QHF Program is equivalent to the sharing of data experiences for the BQI Project.

IHIE has approximately 900 physicians enrolled in the QHF Program and most of those physicians are enrolled through large group contracts. The groups vary in resources and management. Each group has a quality manager responsible for coordinating the groups' quality efforts. Some groups have more electronic capabilities and prefer electronic data feedback where other have paper-based systems and need printed reports.

Each provider is very knowledgeable regarding the quality measures utilized in Cycles 1 and 2 since the measures are nationally recognized. The education of the provider groups has been focused on data aggregation activities since IHIE's QHF Program does not utilize chart review. The QHF measure specifications are very similar to the HEDIS Measure Specifications but are reported on a more current

system versus retrospectively. This has changed the nature of the data collection which has required additional physician education.

The physicians have been very excited about a community-wide quality focus but have also shared concerns regarding true quality measurements and misinterpreted measure rates. Some of the issues and concerns:

1) Coding Issues – patients who are pulled into denominators based on incorrect diagnosis coding from billing systems.

2) Non-Compliant Patients – physicians feel that if they order a test, drug or procedure that they are "doing the right thing" but cannot control patients who do not follow through with the orders. Health Savings Accounts, Payers, and Employers Benefit Packages may place restrictions on what providers can order and determine patients' out of pocket costs, which may affect patient compliance.

3) Sicker Patients – patient's age, co-morbidities, and health status is not taken into consideration affects the appearance of the physicians' scores.

4) Integrity of Data – providers feel the data is not readily available to measure the quality of care. There are so many obstacles in collecting correct data with the different options patients have on filling prescriptions and obtaining free or low cost procedures and screenings.

5) Level of Effort – provider time and staff time to actively participate, review, and correct quality reports have been the majority of concern in our provider community. This has led to more intrusion into the already hectic schedule and has imposed additional stress on an already over tasked staff.

Please see Appendix A for more detailed information on IHIE's QHF Validation Process.

Massachusetts Health Quality Partners

MHQP initiated its data feedback to providers plan in Cycle 2. Under this plan, MHQP provides the multi-payer results from the BQI project, at the physician group level, to the providers prior to their posting on the CMS website.

MHQP agreed to the public posting of its results for the three common measures for all sites: Breast Cancer Screening, Diabetes HbA1c testing, and Diabetes LDL-C testing for Cycle 2, based on performance year 2006.

Consonant with its policy of informing provider prior to public posting of any measure or survey results, MHQP implemented the process used when preparing to issue their annual public Clinical Quality reports on selected HEDIS measures. The process began by MHQP performing final analysis of the results to ensure

their level of comfort with the numerators, denominators, and rates. The results were then compared to the results of the same measures for the same time period reported based on the numbers and rates received annually from the commercial insurance plans on the managed care population in Massachusetts. While the numbers in the numerators and denominators were almost always lower for the BQI measures, the rates for these three measures were very close to MHQP's HEDIS results: for breast cancer the BQI results were 0.7% below MHQP HEDIS, for HBA1c, they were 2.3% lower, and for diabetic LDL-C they were 3% lower.

When this analysis was complete, MHQP added the data caveats to the Important Notes document that would be posted on the CMS website with the results. These were:

- These results have not yet been fully validated; we continue to work on validating the accuracy of the methods used to calculate the measures.
- Merging commercial and Medicare data results in loss of detail that might be important, as for many measures, groups performing well for commercial members do not always perform equally well for Medicare members and vice versa.

MHQP has started the process of informing the physician group contacts that the results would soon be posted, and made each group's results available to them before the posting occurred. The steps involved in this process were as follows:

1) Create individual results data files for each group
2) Post data files on interactive, password protected website
3) Send via email a cover letter explaining the public posting to each group, and notify them that they could access their group's results on the MHQP website by using their group's password, which was provided via telephone. In addition, MHQP provided a summary of the data sources and an overview of how the results differ from the HEDIS results from MHQP that each group is accustomed to seeing annually. (See Appendix B)
4) Include a copy of the Important Notes with the email.

MHQP sent the email to the physician groups on May 30, 2008. They only received two responses back from the physician groups. One group contact wanted to clarify the populations included in the BQI measure results, i.e. does the BQI population include commercial managed care and Medicare managed care. The contact also wanted to know how PPO patients had been attributed to a PCP. The other response was from a group contact who informed MHQP that they should send results directly to two of the groups whose results had been sent to him for distribution. The contact also asked MHQP to change the addresses associated with results from a few of his constituent groups, which they subsequently changed. These revisions were sent to Delmarva and the revised file was posted soon after on the CMS website.

On June 3, 2008, MHQP sent out a second email to the group contacts to inform them that as of that date MHQP had made the results of all 12 BQI Cycle 2 results available to them on the website, including those that would not be reported publicly.

Based on the minimal feedback from the physician groups, MHQP feel this method of providing the data to the physicians prior to posting of the results was effective.

Minnesota Community Measurement

MNCM began collecting data in 2002 for 2001 dates of service. For the first two years of reporting, feedback was provided to the participating medical groups but there was no public reporting. Feedback given to the medical groups included graphs and tables of their own measures, and blinded feedback of all the participating medical groups.

Beginning in 2004 for 2003 dates of service, rates were posted on the MNCM website. Medical groups continued to receive their group's feedback.

Before Cycle 2 results of the BQI Project were posted on the CMS website, MNCM sent email notification to each of the providers. The providers, i.e. physician groups, were given their specific data and offered a chance to provide feedback. None of the providers expressed any concerns with the data that was sent to them.

MNCM has developed a process for providers to appeal or dispute their rates. Medical groups are given an opportunity before the rates are published to ask questions and express concerns. There is a formal process for medical groups to file an appeal with the Quality Audit committee of the MNCM board.

The related policies, Data Dispute Appeals and Appeals of Performance Data and Results, are included in Appendix C.

With their experience in providing feedback and reporting measures, MNCM has learned a few lessons. It is important to always give providers a preview period – you want to make sure they understand what is being reported about them and to let them know they have an avenue to express any questions or concerns. In the long-run, this just adds to buy-in and good will. MNCM's goal has always been to report the most accurate data and to inspire groups to make improvements – if you alienate them along the way, then you aren't achieving your purpose. Another lesson is to not expect that you will be able to publicly report new measures. MNCM has had the experience where they have planned to report, but then, due to measure development changes/concerns, they were not able to do so. The lesson there was always plan for a trial period and if all goes well, then publicly report.

MNCM also experienced some challenges with data feedback and public reporting. One of the challenges they experienced was how to best display the data. MNCM learned that two different displays are needed – one for providers and one for consumers. The MNCM website does not serve the consumers at this point.

Wisconsin Collaborative for Healthcare Quality

Given that WCHQ is measuring and reporting at the practice group level, they have developed physician group feedback report for BQI to be submitted to the Quality Improvement Officers for each organization. A letter was drafted by WCHQ and reviewed and approved by Delmarva prior to sending the results from Cycles 1 and 2 to the physician groups. The letter contained an explanation of the project and the corresponding group's data appearing at the end of the letter. WCHQ plans to implement the same process for future cycles.

The provision of reports to WCHQ's member groups prior to publication is a requirement based on their principles for participation. Only the data specific to a practice group will be sent to that practice prior to public dissemination on the CMS website maintaining the data use agreements set forth in the BQI contract. Also, after each cycle's publication on the CMS website, WCHQ will notify members and update the link to the data on the WCHQ website. To date, WCHQ has had very minimal feedback from their physicians groups on this process.

Because the data that is being published on the CMS website is the aggregate data for all patients regardless of their insurance status and has been previously published on the WCHQ website, the organization does not plan to create a separate set of measure views for the BQI Project. WCHQ felt that interested parties could look to the WCHQ measures and find the same data results as on the CMS website reducing confusion and questions about the data's source.

Conclusion

The goal of data feedback to providers is to gain physician confidence in the measure rates that are being reported as well as allowing them to feel that they are a part of the process. As demonstrated in this report, it also allows them the opportunity to preview the data and provide feedback before the data is publicly reported. The BQI pilot sites have made progress around provider feedback methods within their respective communities. Based on the information provided within this summary, the experience of the sites with data feedback to providers varies widely. As the pilot sites that are just beginning to give this type of information

back to the providers, in their respective communities, continue to implement and improve on their processes we will have more lessons learned to share regarding the most effective ways to provide this type of information.

Appendix A: IHIE QHF Validation Process

Technical Validation Process

Data Flow, Storage, Calculation and application of Measure Rules

Claims and Enrollment Data are received via CD Rom and Encrypted FTP files from payers, data is logged and stored on the Indiana Health Information Exchange (IHIE) server, converted to HL7 format, and sent to Regenstrief Institute (RI). RI stores the data in entity assigned Silos in the Indiana Network for Patient Care (INPC) data base, runs denominator queries aggregating all data in INPC Silos and submits to IHIE via Encrypted FTP. The data will then be run for output in report format where the calculation and rules for the measures will be applied. Each step of the above process will have built-in checks and balances to ensure the data integrity does not vary across each step of the process, see Technical Data Validation Process (documentation of this process is currently underway).

Internal Validation Process

Accuracy of Measure Rates and Clinical Relevance

Reports are validated internally by the Quality Health First SM Physician Liaisons and programming staff. The report validation includes raw procedure and raw test data. The validation consists of checking measure rules programmed against data appearing in reports, alerts and reminders to ensure the data is accurately reflected by the measure rules. The data is also compared by looking at diagnosis and procedure coding to ensure accurate ranges for cohorts and exclusions. The attribution reports are validated by looking at the ages of patients to ensure appropriate attribution and looking for duplication of patients on reports. The measure metrics reports are validated by looking at the percentiles to validate the calculations and to validate the alerts (failures) against the alerts appearing on the alert and reminder report.

External Validation Process

Accuracy of Measure Rates and Clinical Relevance

Reports will initially be delivered to and reviewed with physicians and physician groups by an IHIE Physician Liaison to validate the following. The reports will include drilldown capability at a patient-by-patient level. After initial review, physicians will have the opportunity to submit corrections to IHIE via the Data Reconciliation Process. This validation process will include:

- Physician review on patient/physician attribution
- Physician review on numerators and denominators, including:
    - Patients in the denominator for which physician has documentation of exclusions (i.e. mastectomy)
    - Patients not included in the denominator for which physician has documentation of denominator criteria

- Patients in denominator not included in the numerator for which physician has documentation of measures met (i.e. mammogram report from another physician or facility)
- Patients in the denominator not included in the numerator for which the testing/screening was done in the physician office.
- Patient data attested to accuracy by provider but not verifiable will be only excluded from measure and will not be used for inclusion in the numerator.

Clinical data to substantiate corrections (Data Reconciliation Process) to the reports will be submitted by the physician office directly to IHIE via fax by submission of a standard QHF form, directly from an EMR when applicable, or via the web-based application data through the QHF portal when available. The clinical data received will be used to correct the denominators and/or numerators. Because these data are coming directly from physician offices, we will initiate an inter-rate reliability process to ensure accuracy (See IRR Policy and IRR Tools).

Inter-Rater Reliability Policy

Any clinical data sent by a physician office directly to Indiana Health Information Exchange for Quality Health First (point of care lab data and documentation to correct numerator compliance and exclusionary documentation) will be subject to inter-rater reliability testing.

_ All documentation must be clear, legible and include patient's full name, date of birth, physicians name and date service was performed or the date the physician signed off on documentation recorded in the medical record.

_ Scores will be determined by accuracy of each item, date of service, test or service performed, and result if applicable.

_ Inter-rater reliability testing not meeting 100% may be subject to a full audit of clinical data received from a physician's office at any given reporting timeframe. This may result in a delay of future Quality Health First quality reports and reported to Payers who administer the incentive payments.

Forms:
Point of Care data submission forms
Data Reconciliation forms: Adult, Pediatric and Demographic
IRR Tool

Appendix B: Examples of Letters Sent To Providers

Letter from Massachusetts Health Quality Partners to their Providers Regarding the Public Posting of the BQI Results and Review of Provider Specific Data

May 30, 2008

Dear Colleague,

As you may know, MHQP is one of six organizations in the country that was selected to be a "Better Quality Information" (BQI) pilot by the Centers for Medicare and Medicaid Services (CMS). One of the major goals of the BQI project is to provide recommendations on the most effective methods to aggregate Medicare claims data with data from other payers in order to produce the most accurate and comprehensive measures of the quality of services being provided by physicians to Medicare beneficiaries. MHQP brings to this process the methodological rigor we bring to all of our measurement work.

Public Posting of Results
In addition to testing methodology, this project involves posting performance results for measures on the CMS BQI public website. MHQP has agreed to have CMS use the Massachusetts medical group level results for 2006 for breast cancer screening and diabetes HbA1c and LDL-C testing. In accordance with MHQP's commitment to always make data available to physicians prior to public release, we are making your group's results available on our MHQP private website, so that you can have a preview of what will be shown.

When it is posted by CMS, the data will be available in a downloadable table from http://www.cms.hhs.gov/BQI on or before June 15. The posting will include an **Important Notes** document which appears with the results on the CMS site (and is attached here). The Important Notes describes the measure results and associated numerators and denominators that can be found in each file. It is important to read the information in this document because it explains the limitations to the measures data contained in each file, including the differences across the 6 BQI pilot communities.

How to access your group's data before the CMS posting:
You can access these results by going to MHQP's private website at http://private.mhqp.org. Please contact Anne Gagné at 617-402-5026 for the pass code you can use to access and download your reports on line. When you log in to the site, click on "Reports" to access the BQI results that CMS will be posting. As noted above, CMS will be reporting on three measures for MHQP's population: Breast Cancer Screening, HbA1c and LDL-C testing for diabetics. We report by group when there is a denominator of a minimum of 30 and at least 2 physicians. Therefore, your particular group may not have results for all three measures posted.

Data Sources:
For this project MHQP has merged Medicare fee for service (FFS) claims with managed care and PPO claims from Blue Cross Blue Shield of Massachusetts, Fallon Community Health Plan, Harvard Pilgrim Health Care, and Health New England. Data from 2005 were used to produce an initial set of 5 measures last Fall. Claims data from calendar year 2006 were used to measure performance on 12 services, just completed last week. Performance on these same 12 measures will be calculated again this summer, using calendar year 2007 data. All of the measures have been endorsed by the National Quality Forum, and all but one are NCQA HEDIS measures. (See attached list of measures)

How these results vary from the HEDIS results you regularly receive from MHQP:

HEDIS Results:
Population included: All managed care members, and Administrative Services Only (ASO) participants from the state's five large health plans: Blue Cross Blue Shield of Massachusetts, Fallon Community Health Plan, Harvard Pilgrim Health Care, Health New England, and Tufts Health Plan.

Data source: Numerators and denominators provided by these health plans, grouped and calculated by MHQP, and audited by an independent consultant.

BQI Project Results on CMS Website:
Population included: Medicare FFS beneficiaries and all fully insured adult managed care and PPO members of the health plans named above minus Tufts, and NOT including members insured through self-funded employers with Administrative Services Only (ASO) contracts with the plans.

Data source: Actual claims from commercial plans and Medicare. All claims, plus files on member enrollment and providers were sent to a data aggregator who could accommodate merging and storing all of this administrative data.

Results OVERALL:
In the BQI reports, *numbers* in both patient denominators and numerators are often lower than the HEDIS reports. However, with some exceptions the measure *rates* are comparable between the HEDIS and the BQI data for the commercially insured population.

The Medicare results are in most cases significantly lower than the commercial results and our analysis suggests they vary in other ways as well. The merging of the commercial and Medicare data does not provide consumers, in our opinion, a useful understanding of a group's performance. Despite this fact CMS will be reporting only *merged* data. Most groups will therefore have lower scores reported than their commercial-only scores, but higher scores than they would have had if the Medicare results were reported separately.

If you have any questions about the BQI project or other work that MHQP is doing, please do not hesitate to contact Polly Marvin at 617-402-5022.

Letter from Massachusetts Health Quality Partners to their Providers Regarding the Ability to Review all of the BQI Results

June 4, 2008

Dear Colleague:

Last week we sent a letter informing you that CMS would be publicly posting up to three measures from the BQI project work, on their public website in the near future.

Please note that as of this afternoon, all 12 BQI measures are available on our website - including those which will not be publicly reported. So if you go on to the private site (see below) you will see all of your group's measure results for the BQI project.

Again, please note that we report by group when there is a denominator of a minimum of 30 and at least 2 physicians. Therefore, your particular group may not have results for all measures posted.

How to access your group's data before the CMS posting:
You can access these results on the MHQP's private website http://private.mhqp.org. Please contact Anne Gagné at 617-402-5026 for the pass code you can use to access and download your reports on line. When you log in to the site, click on "Reports" to access the BQI results for all 12 measures.

Reminder: CMS will be publicly posting results on three measures for MHQP's population: Breast Cancer Screening, HbA1c and LDL-C testing for diabetics. Also, on your group's reports there is a column noting which rates will be publicly reported.

Please feel free to contact us with any questions or comments.

Appendix C: MNCM Policy and Procedure – Appeals of Clinical Data and Results

Policy Name: Appeals of Performance Data and Results	**Effective Date:** July, 2007
Policy Owner: Executive Director, MNCM	**Revision Date:** October, 2007; December 2007
Category of Policy: Quality Audit Committee of the Board	

DEFINITIONS:

Appeal of performance result – a formal request by an entity, on which MNCM reports performance results, for reconsideration of these performance results with the goal of finding a mutually acceptable solution.

POLICY STATEMENT:

MNCM has a formal process by which an entity can request reconsideration of a performance result either currently posted or to be posted on the MNCM Web site (www.mnhealthcare.org). Before each posting of results on the MNCM Web site, MNCM staff will provide every reportable entity two weeks to review their result(s). After this two week period, MNCM will post results on its web site.

The review process timeline begins with an e-mail message sent to each reportable entity along with their data and the following information:

- The review period timeline and the date the data will be posted on the MNCM website
- For concerns about the performance data – the timeframe to get the concern to MNCM, the MNCM staff person name and phone number and e-mail address
- MNCM will respond to the concern ASAP and get review decision back to medical group
- If the medical group has an ongoing concern about the performance results after the response from MNCM; the timeline for and how to submit an appeal (including proper areas/topics for appeal)

MNCM will post on its web site and publish in the final, written report the results on each entity unless a formal appeal has been submitted*. As appropriate, performance results under appeal by a reportable entity will be noted on the MNCM web site as "results under review". Publication of the final, written report will be held until appeals have been resolved and results have been finalized.

As a policy, MNCM will not make performance result changes unless an entity can prove errors or if a pattern of data issues/errors can be identified by MNCM based upon multiple comments and/or concerns. Furthermore, if the performance results under investigation and appeal (and subsequent request for recalculation and public reporting of rates) are within confidence intervals of original rate, MNCM has the option to not extend resources to investigate appeal or recalculate rates.

It is also a policy that MNCM will not postpone public reporting or suppression of web posted and published results solely based upon an entity not agreeing to a MNCM measurement specification, data collection process or method and/or that their internally collected rates do not match MNCM results. These concerns will be shared with the Reporting Advisory and/or Data Planning Committee, as appropriate.

If after discussions with MNCM staff, an entity has remaining concerns about the accuracy of data posted on the MNCM web site, published in a report (or to be published) these concerns must be formally presented to MNCM in the form of a written appeal and have evidence that supports the concern.

All appeals will be reviewed by the MNCM Executive Director who will make recommendations to the MNCM Quality Audit Committee. The MNCM Quality Audit Committee has the authority to make decisions about the dispensation of the appeal.

* MN Community Measurement will give the option to any new entity being reported on for the first time on established measures to suppress their data from being publicly reported for the first year.

- The decision to publicly report or to suppress the data will be made after the entity has seen its data.
- This option does not apply to clinic (sites) that are part of medical groups on which MNCM has previously reported**.**

ASSOCIATED PROCEDURES:

1. Medical groups will be sent their performance results along with information on the review process timeline.
2. If a medical group has concerns about their performance result, they may communicate their concerns to MNCM staff within the timeline specified.
3. MNCM staff will respond to the concern and provide additional information as requested and as possible.
4. Medical Group will investigate performance results and submit evidence to MNCM if they are requesting a change in their performance results.
5. MNCM staff will review the evidence and Executive Director will make a decision based upon the medical group request and evidence submitted.
6. If Executive Director's decision is not in medical group's favor, medical group may submit a formal and written appeal to the MNCM Executive Director.
7. Appeals will be reviewed by the MNCM Executive Director.
8. The MNCM staff will investigate the appealed situation, document findings and form recommendations.
9. All appeals will be brought to the MNCM Quality Audit Committee along with the documented investigation, the findings and recommendation of the MNCM Executive Director.
10. The MNCM Quality Audit Committee will determine if additional information is needed, and how best to gather and receive this information. However, any appellant who wishes to make a presentation to the Committee shall be granted the opportunity to do so.
11. The MNCM Quality Audit Committee will make decisions about the dispensation of each

12. MNCM staff will annually collate information on the appeals and produce a summary of total number of appeals, what was appealed, how they were adjudicated and timing of appeals submitted.

LINKS & REFERENCES:

- MNCM Policy and Procedure: Verifying Accuracy of the Data
- MNCM Policy and Procedure: Working with Physicians on Measurement and Making Measurement Methodology Available

Appendix 12

Glossary of Terms

Appendix 12
BQI Glossary of Terms

Administrative data

Data collected for administration of medical encounters. These data can include claims for services received in all settings. They also may include demographic and other billing-related information. The term "expanded administrative data" is used to refer to health care billing data from claims, PLUS health care information such as laboratory test results, pharmacy benefits, and other electronically-available data. Administrative data is distinguished from clinical medical record information, though as more medical records become electronic the distinction will become increasingly blurred.

Agency for Healthcare Research and Quality (AHRQ)

The nation's lead federal agency for research on health care quality, costs, outcomes and patient safety. AHRQ is the health services research arm of the U.S. Department of Health and Human Services (HHS), complementing the biomedical research mission of its sister agency, the National Institutes of Health. The agency is home to research centers that specialize in major areas of health care research, including: clinical practice and technology assessment, health care organization and delivery systems, and primary care. AHRQ is a major source of funding and technical assistance for health services research and research training at leading U.S. universities and other institutions. As a science partner, the agency works with the public and private sectors to build the knowledge base for what works—and does not work—in health and health care and to translate this knowledge into every-day practice and policy-making.
http://www.ahrq.gov/,
http://www.rwjf.org/qualityequality/product.jsp?id=30712

Aligning Forces for Quality: The Regional Market Project (AF4Q)

A national program of the Robert Wood Johnson Foundation (RWJF) designed to help communities across the country improve the quality of health care for patients with chronic conditions such as diabetes, asthma, depression and heart disease. The premise of AF4Q is that no single person, group or profession can improve the quality of care without the support of others. AF4Q seeks to drive quality improvement by aligning key forces, including health care providers (physicians/physician groups, nurses, and clinics), health care purchasers (employers and insurers) and health care consumers (patients). http://www.rwjf.org/qualityequality/product.jsp?id=30712

Ambulatory Care

Medical care provided on an outpatient basis—therefore, not requiring a person to be admitted to the hospital. It is provided in physicians' offices, clinics, emergency departments, outpatient surgery centers and hospital settings that do not involve a patient staying overnight. http://www.rwjf.org/qualityequality/product.jsp?id=30712

Ambulatory Quality Alliance (AQA)

A broad-based coalition of physicians, consumers, purchasers, health insurance plans and others who seek to improve health care quality and patient safety through a collaborative process in which key stakeholders agree on a strategy for measuring performance at the physician or group level; collecting and aggregating data in the least burdensome way; and reporting meaningful information to consumers, physicians and other stakeholders to inform choices and improve outcomes.
http://www.aqaalliance.org/,
http://www.rwjf.org/qualityequality/product.jsp?id=30712

Attribution

The process by which the responsibility for a patient's health care is assigned to a health care provider (patient to provider attribution),

OR

The process by which an individual physician is associated with a physician group (physician to group attribution).

Beneficiary

The name for a person who has health care insurance through the Medicare or Medicaid program. http://www.medicare.gov/Glossary/Search.asp?SelectAlphabet=ALL&Language=English

Bridges to Excellence (BTE)

A not-for-profit organization developed by employers, physicians, health care services, researchers, and other industry experts with a mission to create significant leaps in the quality of care by recognizing and rewarding health care providers who demonstrate that they have implemented comprehensive solutions in the management of patients and deliver safe, timely, effective, efficient, equitable and patient-centered care.
http://www.bridgestoexcellence.org/Content/ContentDisplay.aspx?ContentID=20

Chartered Value Exchanges (CVEs)

A specific example of regional medical coalitions. The CVE concept was developed by the U.S. Department of Health and Human Services (HHS) under the Value-driven Healthcare Initiative to have a nationwide system and network of collaboratives at the regional/local public-private level. A CVE has been defined as a multi-stakeholder collaborative taking clear action in their community to convene community purchasers, health plans, providers, and consumers to advance the four cornerstones of the Value-Driven Healthcare. CVEs are chartered by the AHRQ.
http://www.hhs.gov/valuedriven/communities/index.html, http://www.hhs.gov/valuedriven/communities/valueexchanges/exchanges.html

Claims Data

That part of electronically-available administrative medical data that specifically relates to billing for a provided medical service.

Composite Measure

A measure of performance or quality that is constructed by combining in some manner several other sub-measures. The most common example used by BQI pilot sites was an "all-or-nothing" disease-specific measure. In this type of composite measure, several quality measures dealing with, for example, diabetes care are considered together, and for the composite measure to be counted successful ALL the separate sub-measures must be successful. This allows reporting a single, over-all "diabetes care score" that represents success on all the related diabetes measures. Composite measures may be created by a variety of schemes, and may reflect other aspects of care than one particular disease management.

Consumer

An individual who uses, is affected by, or is entitled or compelled to use a health-related service. http://www.rwjf.org/qualityequality/product.jsp?id=30712

Correlation

It measures the strength and direction of a linear relationship between two quantitative variables.

Data Aggregation

For purposes of this report, combining Medicare FFS data with other data.

Denominator Event

The event or state that defines a case as eligible for inclusion in the denominator. http://www.qualitymeasures.ahrq.gov/resources/glossary.aspx

Fee-for-Service (FFS)

An arrangement under which patients or a third party pay physicians, hospitals, or other health care providers for each encounter or service rendered. http://www.rwjf.org/qualityequality/product.jsp?id=30712

Health Plan Employer Data and Information Set (HEDIS)

A set of health care quality measures designed to help purchasers and consumers determine how well health plans follow accepted care standards for prevention and treatment. Formerly known as the Health Plan Employer Data Information Set, health plans can receive accreditation on HEDIS measures from certain organizations, such as the National Committee on Quality Assurance. http://www.ncqa.org/tabid/187/Default.aspx, http://www.rwjf.org/qualityequality/product.jsp?id=30712

HIPPA

An act that contains federal regulations intended to increase privacy and security of patient information during electronic transmission or communication of "protected health information" (PHI) among providers or between providers and payers or other entities. It includes a framework to allow the appropriate sharing of a patient's PHI among their legitimate medical providers for purposes of improving care and preventing harm. http://www.psnet.ahrq.gov/glossary.aspx

Managed Care

Health insurance plans intended to reduce unnecessary health care costs through a variety of mechanisms, including: economic incentives for physicians and patients to select less costly forms of care, programs for reviewing the medical necessity of specific services, increased beneficiary cost sharing, controls on inpatient admissions and lengths of stay, the establishment of cost-sharing incentives for outpatient surgery, selective contracting with health care providers, and the intensive management of high-cost health care cases. The programs may be provided in a variety of settings, such as health maintenance organizations (HMO), independent practice associations (IPA), and preferred provider organizations (PPO), etc. http://www.qualitymeasures.ahrq.gov/resources/glossary.aspx

Medicaid

A joint Federal and State program that helps with medical costs for some people with low incomes and limited resources. Medicaid programs vary from state to state, but most health care costs are covered if you qualify for both Medicare and Medicaid. http://www.medicare.gov/Glossary/Search.asp?SelectAlphabet=ALL&Language=English

Medicare

Medicare is health insurance for people age 65 or older, under age 65 with certain disabilities, and any age with End-Stage Renal Disease (ESRD) (permanent kidney failure requiring dialysis or a kidney transplant).Medicare has the following parts: Part A (Hospital Insurance), Part B (Medical Insurance), Part C (Medicare Advantage Plans, like an HMO or PPO), and Part D (Medicare prescription drug coverage). http://www.medicare.gov/Publications/Pubs/pdf/11082.pdf

Medicare Part D

Prescription drug coverage offered under Medicare. http://www.medicare.gov/Publications/Pubs/pdf/11082.pdf

National Committee on Quality Assurance (NCQA)

A private, nonprofit organization dedicated to improving health care quality through measurement, transparency and accountability. NCQA has been a central figure in driving improvement throughout the health care system, helping to elevate the issue of health care quality to the top of the national agenda. The organization has helped build consensus around important health care quality issues by working with large employers, policy-makers, doctors, patients and health plans to decide what's important, how to measure it and how to promote improvement.
http://www.ncqa.org/
http://www.rwjf.org/qualityequality/product.jsp?id=30712

National Provider Identifier (NPI)

A 10-digit, intelligence free numeric identifier (10 digit number). Intelligence free means that the numbers do not carry information about health care providers, such as the state in which they practice or their provider type or specialization. The NPI will replace health care provider identifiers in use today in HIPAA standard transactions. Those numbers include Medicare legacy IDs (UPIN, OSCAR, PIN, and National Supplier Clearinghouse or NSC). The provider's NPI will not change and will remain with the provider regardless of job or location changes. NPIs may identify individual physicians, medical groups, hospitals, or any other legal entity engaged in providing medical care.
http://www.cms.hhs.gov/NationalProvIdentStand/downloads/npi_fs_geninfo_010906.pdf

National Quality Forum (NQF)

A nonprofit membership organization created to develop and implement a national strategy for health care quality measurement and reporting. Established as a public-private partnership, the NQF has broad participation from all parts of the health care system, including national, state, regional and local groups representing consumers, public and private purchasers, employers, health care professionals, provider organizations, health plans, accrediting bodies, labor unions, supporting industries and organizations involved in health care research or quality improvement. Together, the organizational members of the NQF work to promote a common approach to measuring health care quality and fostering system-wide capacity for quality improvement. Quality improvement measures endorsed by the NQF are considered the gold standard.
http://www.qualityforum.org/,
http://www.rwjf.org/qualityequality/product.jsp?id=30712

Numerator Event

The event that tells whether the patient received the service being measured.

Pay-for-Performance

A method for paying hospitals and physicians based on their demonstrated achievements in meeting specific health care quality objectives. The idea is to reward providers for the quality—not the quantity—of care they deliver. http://www.rwjf.org/qualityequality/product.jsp?id=30712

Payer

The entity that assumes the risk of paying for medical treatments. Examples include uninsured patients, self-insured employers, health plans or HMOs. http://www.rwjf.org/qualityequality/product.jsp?id=30712

Performance Measure

Sets of established standards against which health care performance is measured.
http://www.rwjf.org/qualityequality/product.jsp?id=30712

Physician Consortium for Performance Improvement (PCPI)

A consortium convened by the American Medical Association (AMA) to enhance quality of care and patient safety by taking the lead in the development, testing, and maintenance of evidence-based clinical performance measures and measurement resources for physicians. http://www.ama-assn.org/ama/pub/category/2946.html

Physician Quality Reporting Initiative (PQRI)

An initiative authorized through the Medicare, Medicaid, and SCHIP Extension Act of 2007. It is a financial incentive for health care professionals to improve the quality of care that they provide. http://www.cms.hhs.gov/pqri/ http://www.rwjf.org/qualityequality/product.jsp?id=30712

Physician Voluntary Reporting Program (PVRP)

An initiative built on Medicare's comprehensive efforts to substantially improve the health and function of its beneficiaries by preventing chronic disease complications, avoiding preventable hospitalizations, and improving the quality of care delivered. Under the voluntary reporting program, physicians who chose to participate helped to capture data about the quality of care provided to Medicare beneficiaries, in order to identify the most effective ways to use the quality measures in routine practice and to support physicians in their efforts to improve quality of care. http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=1701

Process Measure

Performance measures may be described in three general categories: Structural measures, Process measures, and Outcome measures. Structural measures relate to the existence of a capability, and are yes-or-no. For example, Is there a plan for evacuation in emergencies? Yes-or-No? Outcome measures relate to an actual clinical result, for example, was there a post-operative infection? How often? These measures require clinical information to calculate. Process measures relate to how many times a recommended process is used in providing care. For example, how often was the recommended antibiotic given prior to surgery? Process measures are intended as a surrogate or substitute for an outcome measure. Presumably, if the recommended process is followed the outcome will be improved. Process measure data are easier to obtain than outcome measures, and for this reason are commonly used as new areas of measurement are attempted.

Public Reporting

A type of reporting that makes information about physician and physician group performance available for consumers to use to compare the performance of local physicians/physician groups. The expectation is that a comparative public report of local physicians' performance in treating people with chronic illnesses will motivate and improve performance. http://www.rwjf.org/qualityequality/product.jsp?id=30712

Quality Improvement Organization

Organizations that contract with CMS to review care provided to Medicare beneficiaries. http://www.usdoj.gov/atr/public/health_care/204694/appendixc.htm

Reliability	For the purposed of this report, consistency that determines the extent to which providers' scores can be distinguished from each other.
Skilled Nursing Facility (SNF)	A nursing facility with the staff and equipment to give skilled nursing care and/or skilled rehabilitation services and other related health services. http://www.medicare.gov/Glossary/Search.asp?SelectAlphabet=ALL&Language=English
Validity	The degree to which the measure is associated with what it purports to measure. http://www.qualitymeasures.ahrq.gov/resources/glossary.aspx
Value-Driven Healthcare Initiative	An initiative whose purpose is to have a collaborative system that integrates interoperable health information technology, evidence-based quality standards, and payments systems that will help achieve high-quality, cost-effective care for patients. http://www.hhs.gov/valuedriven/news/faq.html
Tax ID (TIN)	The number used for every doctor and every corporation for Internal Revenue purposes, included in billing data.
Unique Physician Identification Number (UPIN)	Unique identifier assigned to each physician by CMS.
Group UPIN 6	A number that is the six-digit UPIN for a medical group.
Group UPIN 4	A number that identifies the specific, geographic, clinic-site location within each group UPIN6.

Appendix 13

References and Resources

Appendix 13
References and Resources

Selecting Appropriate Measures of Physician Performance for Medicare Beneficiaries

Jordan HS, Pine M, Elixhauser A, et al. Cost-effective enhancement of claims data to improve comparisons of patient safety. *J Patient Saf*. 2007;3(2):82-90.

Analysis of Addressing Methodological Questions

Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG.. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281(22): 2098-2105.

Krein SL, Hofer TP, Kerr EA, Hayward RA. Whom should we profile? Examining diabetes care practice variation among primary care providers, provider groups, and health care facilities. *Health Serv Res*. 2002;37(5): 1159-1180.

Landon BE, Normand S-LT. Performance measurement in the small office practice: challenges and potential solutions. *Ann Intern Med*. 2008;148: 353-357.

Teleki SS, Shaw R, Damberg CL, McGlynn EA. Working Paper- Providing performance feedback to individual physicians: current practice and emerging lessons. RAND HEALTH. 2006 Working Report 381-CHCF.

Reporting of Results to Physicians and Consumers

Marshall MN, Romano PS, Davies HTO. How do we maximize the impact of the public reporting of the quality of care? *Int J Qual Health Care*. 2004;16(1): i57-i63.

Milstein A, Lee TH. Comparing physicians on efficiency, *N Engl J Med*. 2007;357(26):2649-2652.

Teleki SS, Shaw R, Damberg CL, McGlynn EA. Working Paper- Providing performance feedback to individual physicians: current practice and emerging lessons. RAND HEALTH. 2006 Working Report 381-CHCF.

Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005;293(10): 1239-1244.

Individual Physician-Level Measurement and Reporting

Krein SL, Hofer TP, Kerr EA, Hayward RA. Whom should we profile? Examining diabetes care practice variation among primary care providers, provider groups, and health care facilities. *Health Serv Res*. 2002;37(5): 1159-1180.

Milstein A, Lee TH. Comparing physicians on efficiency, *N Engl J Med*. 2007;357(26):2649-2652.