AHRQ

AHRQ QUALITY INDICATORS

# Prevention Quality Indicators

SOFTWARE DOCUMENTATION

**AHRQ Quality Indicators**

# Prevention Quality Indicators: Software Documentation, Version 2.1 – SAS

Department of Health and Human Services
Agency for Healthcare Research and Quality
www.qualityindicators.ahrq.gov

October 2001
AHRQ Pub. No. 02-R0202
Revision 2 (January 10, 2003)

## Citation

AHRQ Quality Indicators -- Prevention Quality Indicators: Software Documentation, Version 2.1 – SAS. Rockville, MD: Agency for Healthcare Research and Quality, 2001. AHRQ Pub. No. 02-R0202.

## Abstract

The value of information on health care quality has never been so widely recognized; yet many organizations lack the resources and/or expertise to build a quality information program from the ground up. Recognizing this, the AHRQ Quality Indicators were developed specifically to meet the short-term needs for information on health care quality using standardized, user-friendly methods and existing sources of data.

This module of the AHRQ Quality Indicators was designed to capitalize on the availability of administrative data on inpatient stays to produce information about 16 Prevention Quality Indicators (PQIs): *potentially avoidable hospitalizations for ambulatory care sensitive condition* (ACSC) indicators, which involve admissions that evidence suggest could have been avoided, at least in part, through better access to high-quality outpatient care.

This document is the software documentation for the Prevention Quality Indicators software Version 2.1, which is provided on the AHRQ Web site. The software was developed in SAS, for use on a personal computer. By making this tool available, we hope to assist others in producing information on health care quality more cost effectively.

Details on the development of the Prevention Quality Indicators can be found in "Guide to Prevention Quality Indicators: Hospital Admission for Ambulatory Care Sensitive Conditions."

# Acknowledgments


***Core Project Team***
Mark McClellan, M.D., Ph.D., *principal investigator*
Kathryn M. McDonald, M.M., *EPC coordinator*
Sheryl M. Davies, M.A.
Jeffrey Geppert, J.D.
Patrick Romano, M.D., M.P.H.
Kaveh G. Shojania, M.D.

***Other Contributors***
Amber Barnato, M.D.
Paul Collins, B.A.
Bradford Duncan M.D.
Michael Gould, M.D., M.S.
Paul Heidenreich, M.D.
Corinna Haberland, M.D.
Paul Matz, M.D.
Courtney Maclean, B.A.
Susana Martins, M.D.
Kristine McCoy, M.P.H.
Suzanne Olson, M.A.
L. LaShawndra Pace, B.A.
Mark Schleinitz, M.D.
Herb Szeto, M.D.
Carol Vorhaus, M.B.A.
Peter Weiss, M.D.
Meghan Wheat, B.A.

***Consultants***
Douglas Staiger, Ph.D.

The following staff from **Social & Scientific Systems**, **Inc.**, developed this software product, documentation, and guide:

***Programmers***
Leif Karell
Kathy McMillan
Fred Rohde

***Technical Writer***
Patricia Caldwell

***Graphics Designer***
Laura Spofford

Contributors from the **Agency for Healthcare Research and Quality**:
Anne Elixhauser, Ph.D.
H. Joanna Jiang, Ph.D.
Denise Remus, Ph.D., R.N.
Margaret Coopey, R.N., M.G.A, M.P.S.

We also wish to acknowledge the contribution of the peer reviewers of the evidence report and the beta-testers of the software products, whose input was invaluable.

## Table of Contents

**Abstract** ..... **i**

**Acknowledgments** ..... **ii**

**Introduction** ..... **1**

**Components of the Prevention Quality Indicators module** ..... **1**

**Quick Reference** ..... **3**
Figure 1. Processing Steps ..... 4
Prevention Quality Indicator Module Details ..... 5

**Background** ..... **6**

**Data elements and coding conventions** ..... **7**
Input for the AHRQ Quality Indicators ..... 7
Coding of diagnoses, procedures, and DRGs ..... 8
Missing values ..... 9

**Processing steps** ..... **12**
1.Identify outcomes in inpatient records ..... 12
2.Identify populations at risk from Census population data ..... 12
3.Calculate observed (raw) Prevention Quality Indicator rates ..... 12
4.Risk adjust the Prevention Quality Indicator rates. ..... 12
5.Create MSX smoothed rates. ..... 13

**Program descriptions** ..... **14**
Program 1: CONTROL.SAS ..... 14
Program 2: PQFMTS.SAS ..... 15
Program 3: PQSASA1.SAS ..... 15
Program 4: PQSASA2.SAS ..... 16
Program 5: PQSASA3.SAS ..... 19

**Reviewing the printed output** ..... **21**
PQSASA1.SAS ..... 21
PQSASA2.SAS ..... 21
PQSASA3.SAS ..... 22

**Interpreting the results.** ..... **24**
Observed means. ..... 24
Risk-adjusted rates. ..... 24
MSX smoothed rates (optional). ..... 25
Comparing observed, risk-adjusted and smoothed rates ..... 26

**Benchmark timings** ..... **28**

**User Support** ..... **28**

## List of Tables

**Table 1. Prevention Quality Indicator (PQI) Variables** ..... 5

**Table 2. Prefixes for the Prevention Quality Indicator (PQI) Variables** ..... 5

**Table 3. Prevention Quality Module Contents** ..... 5

**Table 4. Data Elements and Coding Conventions** ..... 10

**Table 5. PQSASA2 aggregation possibilities** ..... 18

# AHRQ Quality Indicators, Software Documentation, Version 2.1: Prevention Quality Indicators

## Introduction

This documentation describes the software for implementing the Prevention Quality Indicator (PQI) module of the AHRQ Quality Indicators Version 2.1 (Revision 2) and highlights features of the analytic approach of particular interest to new users. Detailed information about the Prevention Quality Indicators (PQI) is contained in the companion document "*Guide to the Prevention Quality Indicators,"* which includes information about the development of the indicators and their definitions.

The software consists of a series of computer programs that:

- Assign and calculate PQIs from hospital discharge abstracts or UB (uniform billing) claims data.
- Print PQI results in SAS format or as an ASCII text file (comma delimited) that can be imported into word processing documents, spreadsheets, or graphics software, at the user's option.
- Provides the option to generate rates by Metropolitan Statistical Areas (MSAs) or by county for urban counties (rural areas are defined by county in either case).
- Calculate rates using either the hospital location or the county of patient residence.
- Create risk-adjusted rates that adjust for casemix differences.
- Create smoothed rates that reduce fluctuations over time due to random variation.

This documentation includes:

- A discussion of the data elements necessary to produce the AHRQ PQIs and the uniform coding conventions recommended for those elements.
- Descriptions of the PQI processing steps in nontechnical language.
- Detailed descriptions of the functions of each PQI SAS program.

***Critical user information is highlighted with this symbol:***

## Components of the Prevention Quality Indicators module

As shown in Table 3 on page 5, the Prevention Quality Indicators module consists of five SAS programs and 13 ASCII text files that contain auxiliary data. These programs and text files are described in the subsequent sections of this document. The programs were developed for use in Statistical Analysis System (SAS), version 6.12 or higher, on an IBM compatible computer.

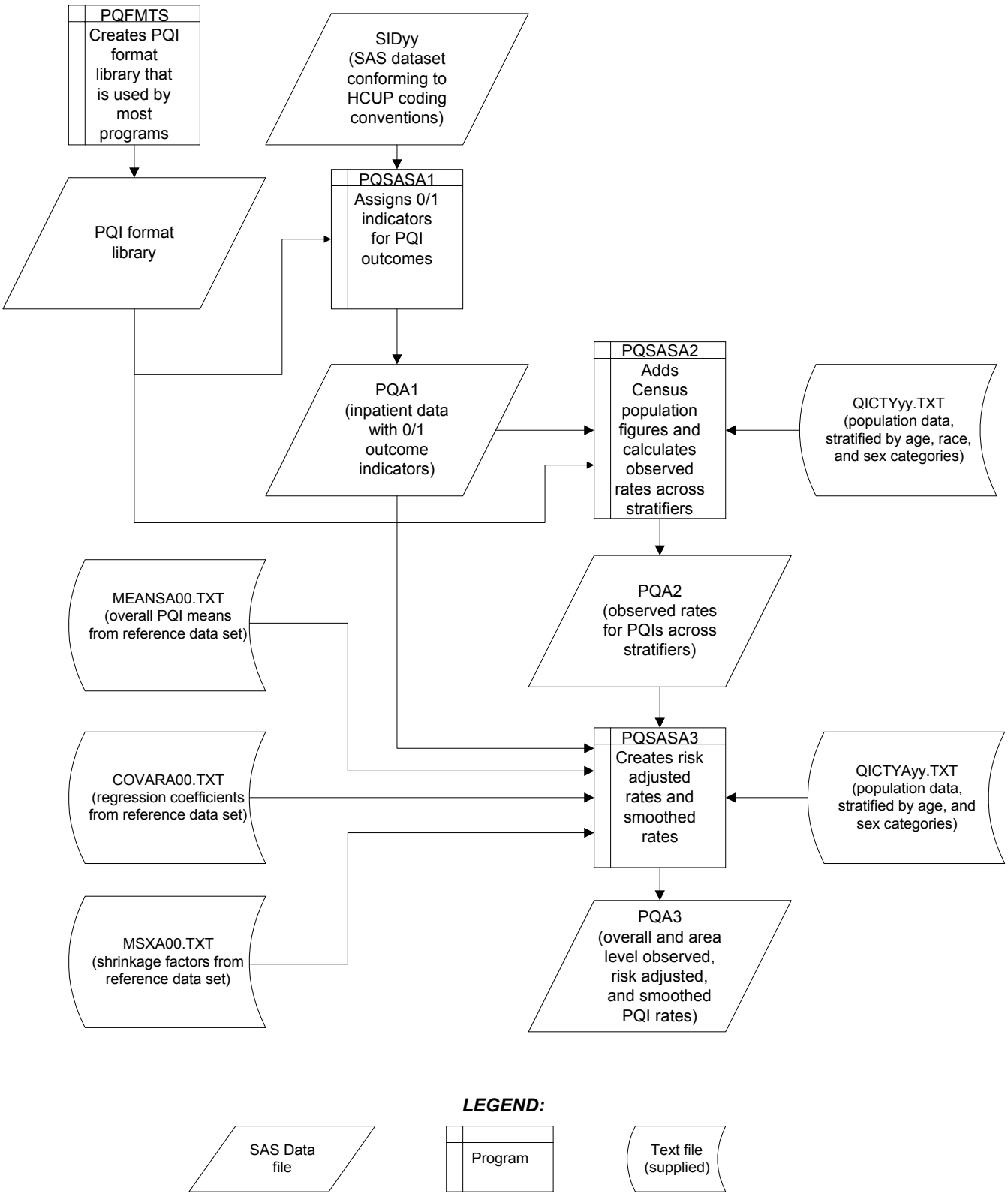
A note on the naming conventions for the Prevention Quality Indicators (PQI) module programs. The programs have names of the form PQSASAi. The first two characters "PQ" of the program name indicates a Prevention Quality Indicators program, to distinguish it from other modules that have been or will be released from AHRQ. The next three characters of the program name are "SAS" and are present to

distinguish the program from SPSS versions of the software. The sixth character of the program name is always “A” to denote a program that is for the production of Area indicator rates that use county and Metropolitan Statistical Area (MSA) area populations as denominators. The last character (i) designates the number of the specific program.

## Quick Reference

The subsequent two pages are intended to serve as a quick reference to assist in reading this documentation and in reviewing the Prevention Quality Indicator (PQI) module outputs. Processing steps (Figure 1) are shown first followed by a listing of the module variables (Table 1), variable prefixes (Table 2), and module contents (Table 3). One suggestion is to print these two pages in duplex mode on one sheet of paper that can then be easily referred to as the need arises.

***Figure 1. Processing Steps***



## *Prevention Quality Indicator Module Details*

| Table 1. Prevention Quality Indicator (PQI) Variables | | | | | | |
|---|---|---|---|---|---|---|
| **Indicator number** | **Numerator (admissions for ACSC)** | **Denominator** | **Age categories** | | | |
| | | | **0 to 17** | **18 to 39** | **40 to 64** | **65 +** |
| 1 | Diabetes short term cx | Area population | | Included | Included | Included |
| 2 | Perforated appendix | **Appendicitis*** | Included | Included | Included | Included |
| 3 | Diabetes long term cx | Area population | | Included | Included | Included |
| 4 | Pediatric asthma | Area population | Included | | | |
| 5 | COPD | Area population | | Included | Included | Included |
| 6 | Pediatric gastroenteritis | Area population | Included | | | |
| 7 | Hypertension | Area population | | Included | Included | Included |
| 8 | CHF | Area population | | Included | Included | Included |
| 9 | Low birth weight | **Births*** | NA | NA | NA | NA |
| 10 | Dehydration | Area population | Included | Included | Included | Included |
| 11 | Bacterial pneumonia | Area population | Included | Included | Included | Included |
| 12 | Urinary infection | Area population | Included | Included | Included | Included |
| 13 | Angina w/o procedure | Area population | | Included | Included | Included |
| 14 | Diabetes uncontrolled | Area population | | Included | Included | Included |
| 15 | Adult asthma | Area population | | Included | Included | Included |
| 16 | Lower extremity amputation, DM | Area population | | Included | Included | Included |

ACSC – Ambulatory Care Sensitive Condition.
* The denominator includes all hospital admissions for this condition in the area.

| Table 2. Prefixes for the Prevention Quality Indicator (PQI) Variables | | |
|---|---|---|
| **Prefix** | **Contents** | **Example (for PQI # 1, Diabetes short–term complications)** |
| T | Inpatient numerator (top) | TAPQ01 |
| P | Population denominator (pop) | PAPQ01 |
| O | Observed Rate | OAPQ01 |
| R | Risk-adjusted Rate | RAPQ01 |
| S | Smoothed Rate | SAPQ01 |

| Table 3. Prevention Quality Module Contents | |
|---|---|
| **SAS programs** | **ASCII text files †** |
| CONTROL.SAS | QICTY97.TXT |
| PQFMTS.SAS | QICTY98.TXT |
| PQSASA1.SAS | QICTY99.TXT |
| PQSASA2.SAS | QICTY00.TXT |
| PQSASA3.SAS | QICTY01.TXT |
| | QICTY**A**97.TXT |
| | QICTY**A**98.TXT |
| | QICTY**A**99.TXT |
| | QICTY**A**00.TXT |
| | QICTY**A**01.TXT |
| | MEANSA00.TXT |
| | COVARA00.TXT |
| | MSXA00.TXT |

† The ASCII text files are provided with the Prevention Quality module, and are necessary for the programs to run. There is no need to convert these files to SAS format for use with the software.

## Background

The PQI module contains 16 indicators that measure hospital admissions for ambulatory care sensitive conditions (ACSC) across geographic areas. ACSCs represent conditions for which hospitalization could be avoided if the patient receives timely and adequate outpatient care. Many factors influence the quality of outpatient care, including access to care and adequately prescribed treatments, once care is obtained. In addition, patient compliance with those treatments and other patient factors may play a role. In general, areas with lower social-economic status tend to have higher admission rates for ACSCs than areas with higher social-economic status. As with utilization indicators, there are no "right rates" of admission for these conditions. Very low rates could signal inappropriate underutilization of healthcare resources while very high rates could indicate potential overuse of inpatient care.

Therefore, hospital admission for ACSCs is not a measure of hospital quality but a potential indicator of outpatient and community health care need. For example, if an area has a relatively high hospital admission rate for diabetes complications, the local healthcare providers should work with the community to identify reasons and strategies to address the problem.

The 16 ACSCs in the PQI module are listed in Table 1 on page 5. See also Appendix A of the "*Guide to the Prevention Quality Indicators"* for the definition of each indicator.

Except for perforated appendix and low birth weight, each indicator is measured as the number of hospital admissions for a particular ACSC per 100,000 residential population in an area. By default, the programs define areas by county; however, users have the option of specifying areas by Metropolitan Statistical Area (MSA) for urban counties. Rural areas are defined by county in either case.

The data required for measuring these indicators come from hospital discharge abstracts or billing claims (administrative data) which are readily available in many states. The residential population data are from the Bureau of Census.

The software generates observed, risk-adjusted, and smoothed rates for each indicator at the area level. Observed rates are the raw rates. Risk-adjusted rates are derived from applying the average casemix of a baseline file that reflects a large proportion of the U.S. hospitalized or residential population. Smoothed rates are estimates with removal of fluctuations over time due to random variation.

For information about how these indicators were identified, see the "*Guide to the Prevention Quality Indicators."*

# Data elements and coding conventions

## *Input for the AHRQ Quality Indicators*

The PQI software was written to process data from discharge data abstracts that contain information about hospital stays. The specific data elements that are needed for the PQI software are listed in Table 4. The PQI module was tested on data from the Healthcare Cost and Utilization Project (HCUP). HCUP is an ongoing Federal-State-private collaboration to build uniform databases from administrative hospital-based data collected by State data organizations and hospital associations.

- *The input data files for the Prevention Quality Indicators software must be in SAS.*

- *We recommend that users recode data elements in their input files to be consistent with the coding expected by the software. This will minimize the number of internal changes that will be necessary in the Prevention Quality Indicators software. For example, recoding the SEX data element in the input file to be consistent with the coding described in Table 4 (i.e., 1 for male and 2 for female) is easier than modifying all uses of the SEX data element in the PQI programs.*

- *Not every value for every data element is referenced by the PQI software. For example, admission source (ASOURCE) values are used only to identify transfers. However, we include the complete set of HCUP values to assist users of the uniform HCUP data files.*

- *To minimize internal changes to the software, all required elements should be present in the input data file. If a required element is not available, a dummy element should be provided. Failure to provide a dummy element will result in errors during the execution of the PQI programs.*

The data elements listed in Table 4 are those required for the input data files of both SPSS and SAS. When a variable is not required for the SPSS program the notation "Not used by the QI program" will be present in the Comments column. The SPSS program will still run if these variables are not present, however, the SAS program will not run correctly. Standardizing the variables and data values in the input data file will be helpful for users who have access to both statistical software packages.

The data element HOSPSTCO provides flexibility for users to calculate the PQIs by hospital location or by patient residence. If the user wants to calculate the PQIs using the population associated with the hospital location as the denominator, the values for this variable should the individual hospital FIPS[1] state/county codes. If the user wants to calculate PQIs based on the population of the MSA region or county associated with the patient residence, the values for this variable should be the FIPS state/county code associated with each individual patient's place of residence.

- *If the hospital FIPS code is used in HOSPSTCO, users should be aware that rates may be biased for hospitals which serve as regional referral centers. These hospitals are likely to treat patients from outside the MSA, county or even the state in which the facility is located.*

- *Using the patient FIPS state/county code for analysis may more accurately reflect the true population at risk. Evaluation of geographic variations in admissions for ambulatory care sensitive conditions by patient MSA or county of residence can result in better information to guide community or provider response.*

- *It is possible that some records in the input data file may be missing the patient FIPS code. Users should be aware that any records with missing values (in the HOSPSTCO data field) are excluded from the calculations of observed, risk-adjusted and smoothed PQI rates. They will be included in the output from the third program (PQSAS1.SAS).*

[1] Federal Information Processing Standard, as defined by the U.S. Department of Commerce, National Institute of Standards and Technology (formerly National Bureau of Standards).

*The software will generate rates for each county included in the HOSPSTCO data field. Users are encouraged to limit cases in their input file to those patients discharged from the geographic area of interest. For example, if you are using data from the state of Wisconsin and are interested in generating a rate for Wisconsin, you should remove the cases where the patient's county of residence (FIPS code) is from another state. Otherwise the total or overall rate will include the population of the counties outside Wisconsin.*

This version of the software provides the user with the option of producing output by MSA or by county. When MSA is selected, urban areas are defined by MSAs. When county is selected urban areas will be defined by county. Rural areas are always defined by county.

*In the software programs, the MSALEVL parameter indicates whether MSA codes should be created from the FIPS state/county codes in the input dataset. If the user wishes to analyze data at the MSA level, this parameter should be set to 1. If the user wishes to analyze data at the county level, this parameter should be set to 0.*

## *Coding of diagnoses, procedures, and DRGs*

*Diagnoses and procedures must be coded using the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM)*

*Significant modifications to ICD-9-CM codes occurred in the early 1990s. PQI definitions only reflect codes valid after October 1, 1994, and therefore may not accurately analyze data collected before 1994.*

- Diagnoses and procedure codes should be in character strings.
- Diagnosis and procedure codes should not contain any decimal points.
- Diagnosis and procedure codes should be left justified. Diagnosis codes with fewer than five digits should be padded with spaces (not zeros) on the right. Procedure codes with fewer than four digits should be padded with spaces (not zeros) on the right.

*The PQI programs assume accurate and specific coding of ICD codes. If an ICD code is specified using 5 digits, the software recognizes only 5-digit versions of that code and not versions truncated at 3 or 4 digits. Similarly, if an ICD code is specified using 4 digits, the software recognizes on 4-digit versions of that code and not versions truncated at 3 digits. For example, 4281 is the ICD-9-CM diagnosis code for left heart failure. If your data are coded less specifically (i.e., 428 is used to designate "heart failure"), these less specific codes are not recognized by the software and will result in undercounting cases.*

- Diagnosis-related groups (DRGs) are those derived from the Centers for Medicare & Medicaid Services (previously Health Care Financing Administration) Medicare grouper. The software expects that you will be using the DRG effective on the discharge date. The software now refers to data elements DRG and MDC. Your data should be coded accordingly. The software may be modified at the user's option to use other types of DRGs. However, the impact of using other types of DRGs should be evaluated carefully before making such a change.

*Version 2.1 of the software accounts for ICD-9-CM and DRG coding effective through September 30, 2002*.

## *Missing values*

- *The PQI programs do not distinguish among different types of missing values.*

- *Cases that are missing data in fields used for indicator definitions or for risk adjustment (such as age and sex) will be excluded from analysis.*

Data files of hospital discharge abstract data may have numeric data elements coded using special SAS missing "dot" values. For example, besides the standard SAS value of "." for missing data, there might also be present values of ".A" for invalid data, ".B" for data unavailable from a particular source, ".C" for inconsistent data. However, the PQI programs do **not** distinguish among the different types of missing codes. Therefore, all types of missing values in the input data to the PQI module can be represented by a single value - missing data (.) for numeric variables and blank (' ') for alphanumeric (or character) variables.

## Table 4. Data Elements and Coding Conventions

| Variable name | Description | Format | Value description | Comments |
|---|---|---|---|---|
| KEY | Sequence number. Unique case identifier. | Numeric | User defined unique numeric identifier for each discharge record | Used to facilitate data exploration and possible trouble shooting. |
| AGE | Age in years at admission | Numeric | Age in years | |
| RACE | Race of patient. | Numeric<br>1<br>2<br>3<br>4<br>5<br>6 | White<br>Black<br>Hispanic<br>Asian or Pacific Island<br>Native American<br>other | The values of 1,2,3 are used directly in the QI software. All other ethnicity codes are mapped to an 'other' category. |
| SEX | Sex of patient | Numeric<br>1<br>2 | Male<br>Female | |
| HOSPSTCO | Hospital location (FIPS† State/county code) | Numeric<br>ssccc | Modified Federal Information Processing Standards State/county code | Available at: http://www.census.gov/population/estimates/county/ |
| HOSPID | Data Source hospital number | Numeric | Hospital identification number | Used to facilitate data exploration and possible trouble shooting. |
| ASOURCE | Admission Source | Numeric<br>1<br>2<br>3<br>4<br>5 | ER<br>Another hospital<br>Another fac. incl. LTC<br>court/law enforcement<br>routine/birth/other | The values 2 and 3 are referenced by the QI code (to identify transfers from another hospital) |
| DISP | Disposition of patient | Numeric<br>1<br>2<br>3<br>4<br>5<br>6<br>7<br>20 | Routine<br>Short-term hospital<br>Skilled nursing facility<br>Intermediate care<br>Another type of facility<br>Home health care<br>Against medical advice<br>Died | Not used by the QI program. |
| DRG | Diagnosis Related Group | Numeric | DRG from federal DRG Grouper | |
| MDC | Major Diagnostic Category | Numeric | MDC from federal DRG grouper | |
| DX1 – DX30 | Diagnoses ICD-9-CM codes. DX1 is the principal diagnosis, DX2-DX30 are secondary diagnoses. | String, 5 character | Diagnoses code | Users with more or fewer secondary diagnoses should modify the parameter &NDX in CONTROL.SAS to reflect the number of diagnoses. |
| NDX | Count of the number of diagnoses (Dx) on the discharge record. | Numeric | | If the 1st discharge record has 5 Dx and the 2nd has 10, NDX would be 5 on the 1st and 10 on the 2nd. Not used by the SAS program. |
| PR1 – PR30 | Procedure ICD-9-CM codes. PR1 is the principal diagnosis, PR2-PR30 are secondary procedures. | String, 4 character | Procedure code | Users with more or fewer secondary procedures should modify the parameter &NPR in CONTROL.SAS to reflect the number of diagnoses. |
| NPR | Count of the number of procedures (Pr) on the discharge record. | Numeric | | If the 1st discharge record has 1 Pr and the 2nd has 2, NPR would be 1 for the 1st record and 2 for the 2nd. Not used by the SAS program. |

## User options

The PQI software reflects the development and implementation of the software in data available to AHRQ. Our goal was to develop the tools, illustrate their use, and then encourage others to adopt and use the tools for their own applications. As a result, we expect and encourage users of the software to consider whether and how to modify the PQI software to better serve their local conditions or interests.

*Modifications to the definitions of outcomes of interest (numerators) or populations at risk (denominators) are possible but not desirable. Maintaining consistent definitions is important. Once definitions are altered, the ability to make comparisons of PQIs based on original definitions is lost. We encourage users to identify ways to improve the PQI methodology and to share their suggestions with us for future updates.*

AHRQ will provide national estimates using the PQIs through its web-based query system, HCUPnet (http://hcup.ahrq.gov/HCUPnet.asp). If users change definitions of the PQIs, it will not be possible to compare users' results to the national estimates in HCUPnet.

## Processing steps

Each Prevention Quality Indicator (PQI) expressed as a rate, is simply defined as:

Outcome of interest / Population at risk

Conceptually, five steps are necessary to produce the PQI rates. The following describes the steps and how the software performs them.

### *1. Identify outcomes in inpatient records*

Inpatient records are marked to indicate whether they contain the outcome of interest (numerator or "top") for each of the AHRQ PQI measures.

This is done by setting a series of flag variables, each of which corresponds to the numerator for a particular PQI. For example, if the inpatient record meets the conditions for inclusion in the outcome for PQI #1, then the outcome indicator for PQI #1 is set to 1.

This step requires one pass through the discharge-level data and outputs a discharge-level data file containing the flag indicator variables for the outcomes for each PQI. This output file also contains the original input variables, with the exception of the diagnosis and procedure variables which have been removed to reduce the size of the output file.

### *2. Identify populations at risk from Census population data*

The populations at risk (the denominators for calculating the PQI rates) are derived from Census population figures by areas defined by county in all areas of the US, or by MSA in urban areas and by county in rural areas.

### *3. Calculate observed (raw) Prevention Quality Indicator rates*

Using the output data from step 1 and Census population data from step 2, the PQI rates are calculated for user-specified combinations of stratifiers. These stratifiers include area, age groups, race/ethnicity categories, and sex.

*The programs calculate observed PQI rates regardless of the number of cases available. However, PQI rates based on only a few cases should be interpreted with caution.*

In work at AHRQ, we do not report rates based on fewer than 30 cases in the numerator or the denominator. This exclusion rule serves two purposes:

- It eliminates unstable estimates based on too few cases.
- It helps protect the identities of hospitals and patients.

This step outputs the observed rates to user-specified levels by area, age, sex, and race/ethnicity categories, alone or in combination. For example, rates can be output at the area level, at the age group level, or at the area level by age groups.

### *4. Risk adjust the Prevention Quality Indicator rates.*

Overall file means and regression coefficients from a baseline database (reflecting a large proportion of the U.S. population) are applied to the observed rates to risk-adjust the rates observed

in the user's data. These baseline file means and regression coefficients are provided as part of the PQI module. The risk-adjusted rates will then reflect the age and sex distribution of areas in the baseline file rather than the distribution for the areas covered by the users' data. This will allow risk-adjusted rates produced by various users to be compared directly to one another. The overall means and regression coefficients were derived from AHRQ's State Inpatient Databases (SID) for 29 states. The codes to generate these baseline means and coefficients are not part of the PQI module.

### *5. Create MSX smoothed rates.*

Shrinkage factors are applied to the risk-adjusted rates for each PQI in a process called multivariate signal extraction (MSX). These shrinkage factors were calculated from the SID database of 29 states. For each PQI, the shrinkage estimate reflects a 'reliability adjustment' unique to each indicator. The less reliable the PQI over time and across areas, the more the estimate 'shrinks' the PQI toward the overall area mean. The resulting rate will appear "smoother" than the raw rate, meaning the year-to-year fluctuations in performance are likely to be reduced. More information on interpreting smoothed rates is contained in the interpretation section of this document. The shrinkage factors are provided as part of the PQI software and do not need to be calculated by users.

These five steps reflect the PQI module production in a nutshell. The next section of this document describes the specifics of each software component of the PQI module software.

# Program descriptions

This section describes the five SAS programs that assign, calculate, and print the Prevention Quality Indicators.

For each program there is a description, a list of input and output files, and an explanation of changes to the program that may be required. The flow of data through the PQI module programs is shown in the flowchart in Figure 1 of page 4.

If you desire to create and examine observed PQI rates, then the PQFMTS, CONTROL, PQSASA1, and PQSASA2 programs will need to be used. If you also wish to create risk-adjusted and smoothed PQI rates, then you will also need to run the PQSASA3 program.

## *Program 1: CONTROL.SAS*

The CONTROL.SAS code contains all of the SAS statements that you need to modify in order to run the remaining PQI module programs (with one exception noted below). This is where you specify, for example, the path names or locations of all your input and output data sets and whether you want rates calculated at the county level or the MSA/county level. There is also a TITLE line that the user can fill in. The items that need to be changed in order to run the PQSASA1, PQSASA2, and PQSASA3 programs are described below under each program.

- *The CONTROL.SAS code is automatically inserted into the remaining programs when they are executed. You do **not** need to run CONTROL.SAS as a separate program.*

- *All of the changes needed for the CONTROL.SAS code are listed in the following descriptions for programs 2 – 5. Each program description contains a section labeled "Changes:" that enumerates the modifications for CONTROL.SAS that are associated with that particular program.*

- *Lines of code to be modified are identified by the phrase "<===USER MUST modify" which means the user **must** make the change or by "<===USER MAY modify" which means the modification is optional.*

In all of the remaining programs included with the PQI module, there is one line of code near the beginning of each program that needs to be modified. This line of code starts with the phrase "FILENAME CONTROL". You ***must*** modify this line to point to the path name where you stored your modified copy of the CONTROL.SAS file.

CONTROL.SAS has a collection of parameters that allow you to read in data from and write out data to different locations or pathnames. For example, "LIBNAME1 IN1" points to the location of your input data set for program PQSASA1 and "LIBNAME OUT1" points to the location of the output data set created by the PQSASA1 program. The parameters allow these locations to be different from one another. However, if you wish to read in and write out all of the data to the same location, it is easiest to just make a global change in CONTROL.SAS changing "C:\pathname" to the single directory location that you wish to use.

There is a global PRINT parameter in CONTROL.SAS that as a default prints all of the output summary records at the end of the PQSASA2 and PQSASA3 programs. These printouts may be quite large if you have a large database with a large number of areas (eg, counties, MSAs) and if you choose the results to be finely stratified. If you wish to turn these prints off, set PRINT to have a value of zero in CONTROL.SAS.

- The global MSALEVL parameter in CONTROL.SAS by default instructs all programs to define geographic areas by county for all areas in the input datasets. To define urban areas by Metropolitan Statistical Area (MSA) rather than by county, change the MSALEVL parameter from 0 to 1.

## *Program 2: PQFMTS.SAS*

The PQFMTS program defines a format library, which contains the diagnosis and procedure screens necessary for assigning the outcomes of interest. This format library is used by most of the succeeding programs.

The first few formats in this program are for the stratifiers (age, sex, race/ethnicity). These are followed by formats that are used for the identification of outcomes of interest in the PQSASA1 program, and by formats that map FIPS county codes to metropolitan statistical areas (MSAs) in the PQSASA1 and PQSASA2 programs. At the end of the program are formats used to assign age categories for the risk adjustment process implemented in PQSASA3.

The MSA definitions are from June 30, 1999 according to the federal Office of Management and Budget (OMB) circular 99-04. The mapping is from county to MSA, except in New England where counties are assigned to NECMAs (New England County Metropolitan Area). OMB defines NECMAs as a county-based alternative to the city- and town-based New England MSAs and CMSAs (Consolidated MSAs). For example, Hampden and Hampshire counties in Western Massachusetts are assigned to the Springfield MA NECMA, even though Holland town in Hampden County is part of the Boston MSA.

The PQI module version 2.1 contains formats that incorporate ICD-9-CM and DRG coding through September 30, 2001.

Input: None.

Output: Permanent SAS format library (FMTLIB).

Changes:

1. You must modify the "FILENAME CONTROL" statement at the beginning of the program to change the path so that it points to your location of the CONTROL.SAS code.

2. In CONTROL.SAS, specify the location (path) of the format library in the "LIBNAME LIBRARY" statement.

## *Program 3: PQSASA1.SAS*

The PQSASA1 program processes hospital discharge abstract data and flags inpatient records if they contain outcomes of interest. Outcome indicator names have prefix of "T" (Top). Stratifier variables are constructed at the beginning of the program.

This program assumes that the input data file (consisting of inpatient discharge abstract data) conforms to specific variable names, attributes, and coding conventions. See Table 4 on page 10 for variable names and attributes for the input data file.

Partial definitions of the Prevention Quality Indicators are given in Table 1 on page 5. This table is presented to assist those individuals who desire to examine the software source code statements. Complete definitions of the indicators are given in Appendix A of the "*Guide to the Prevention Quality Indicators".*

Input:

1. User supplied SAS inpatient data set consisting of administrative hospital discharge abstract data. This data set is a discharge level file with an array of diagnosis and procedure codes, among other data elements.

2. SAS format library (FMTLIB) created from the PQFMT program.

Output:

1. SAS dataset containing inpatient records with input variables, stratifiers, and flag indicators (TAPQxx) for the outcomes of interest that will later form the numerators for the PQI rates.

2. PROC MEANS (with N, NMISS, MIN, MAX, MEAN, and SUM) of all of the numeric variables in the output data file and a PROC CONTENTS of the output data file are output to hardcopy.

3. A PROC PRINT of the first 24 observations in the output data file is output to hardcopy.

Changes:

1. The "FILENAME CONTROL" statement at the beginning of the program must be changed to reflect the location of the CONTROL.SAS code.

2. In CONTROL.SAS, specify the locations (paths) of the input and output files in the "LIBNAME IN1" and "LIBNAME OUT1" statements.

3. In CONTROL.SAS, specify the names of the input and output files in the "INFILE1" and "OUTFILE1" statements.

4. To define urban areas by MSA rather than by county, change the MSALEVL parameter in CONTROL.SAS from 0 to 1. Rural areas will be defined by county no matter how MSALEVL is specified.

5. *The PQI module assumes that your starting SAS data set contains 30 diagnoses, and 30 procedures. If these numbers are different in your input inpatient data, then modify the NDX and NPR parameters in CONTROL.SAS.*

6. *If your variables are named differently from those used in the PQI software (see table 4 on page 10), you MUST either rename those variables prior to running this program or modify PQSASA1 to include the renaming of your variables to match those expected by the software. Similarly, if your variables are not the same type, i.e., character or numeric, as those expected by the software (see Table 4), you MUST modify your variables prior to running this program.*

## *Program 4: PQSASA2.SAS*

The PQSASA2 program calculates the observed or raw rates for the area-level Prevention Quality Indicators, using the data derived in the previous step (PQSASA1). These observed rates are stratified by combinations of area, sex, age, and race categories. The program first totals the indicator flags created by the PQSASA1 program, and then for each of the desired stratifiers divides these totals by the pertinent residential population. The population denominators are stored in variables with names that have a prefix of "P" (Pop). The Observed rates are stored in variables that have a prefix of "O".

Input:

1. The SAS dataset that was created in the previous step (PQSASA1). This is a discharge-level file that contains inpatient records with input variables, stratifiers, and the 16 new flag indicators for the PQI outcomes of interest.

2. A text file with Census area residential populations, stratified by area, age, sex, and ethnicity categories. Five such files are currently provided along with the PQI module software. The files are QICTY97, QICTY98, QICTY99, QICTY00, and QICTY01. The user should select the file for the year that best matches the user's discharge datafile.

*Users do **not** need to convert the ASCII text file to a SAS dataset for use with the software.*

3. SAS format library (FMTLIB) created from the PQFMT program.

Output:

1. SAS dataset with summary records that contain observed rates (OAPQxx variables where xx refers to the indicator number), the counts of outcomes that formed the numerators of the rates (TAPQxx variables), and the residential population totals that formed the denominators of the observed rates (PAPQxx variables). The output file has records summarized to the various combinations of stratifiers specified in the TYPELEVL parameter that is described in item #5 in the "Changes" section below.

2. An optional ASCII (comma delimited) text file that the user can then import into a spreadsheet. This text file contains the same information as the SAS output dataset. The text file will contain the OAPQxx observed rates, the TAPQxx counts of outcomes that formed the numerators of the rates, and the PAPQxx residential population totals. The different records/rows in the text file will correspond to the different TYPELEVL stratification combinations.

3. A PROC MEANS (with N, NMISS, MIN, MAX, MEAN, and SUM) of the area-level summary records that shows statistics for the OAPQxx observed rates, the TAPQxx counts of outcomes that formed the numerators of the rates, and the PAPQxx residential populations totals. These means will only be generated if the user included a value of 8 for the TYPELEVL parameter discussed in item 5 in the "Changes" section below.

4. A PROC CONTENTS of the output SAS summary dataset is output to hardcopy.

5. A PROC PRINT of the output summary dataset is provided at the end of the hardcopy printout. This printout may be quite large depending on the number and the types of stratifications that the user requests with the TYPELEVL parameter discussed in item 5 in the "Changes" section below. If the user does not wish to generate this printout, then the global "PRINT" parameter in CONTROL.SAS code should be set to zero.

Changes:

1. The "FILENAME CONTROL" statement at the beginning of the program must be changed to reflect the location of the CONTROL.SAS code.

2. In CONTROL.SAS, specify the locations (paths) of the input and output SAS files in the "LIBNAME IN2" and "LIBNAME OUT2" statements.

3. In CONTROL.SAS, specify the names of the input and output files in the "INFILE2" and "OUTFILA2" statements. The name that you specify with INFILE2 should be for the file that you created using the previous PQSASA1 program.

4. In CONTROL.SAS, specify the path (including the file name) for the ASCII population text file in the "FILENAME POPFILE2" statement. The file name will be QICTY97.TXT, QICTY98.TXT, QICTY99.TXT, QICTY00.TXT or QICTY01.TXT. These files are provided to you as part of the PQI module. The file used should be the one that is closest to the year associated with your particular data.

5. In CONTROL.SAS, specify the levels of aggregation for the observed rates. This is done by specifying numbers between 0 and 15 in the "%LET TYPELEVL =" statement. Each number from 0 to 15 corresponds to a different level or combination of levels as show in Table 5. The default values of 0, 8 will provide an overall total and area-level totals.

   ***WARNING:** TYPELEVL must include the values 0 (overall) and 8 (area) if you wish to subsequently run program PQSASA3 for risk-adjusted and smoothed rates.*

*If TYPELEVL includes the value 0, then the first observation in the output summary datafile will contain the overall totals and observed rates for the entire database created with the PQSASA1 program.*

| Table 5. PQSASA2 aggregation possibilities | |
|---|---|
| **TYPELEVL** | **Stratification** |
| **0** | **Overall** |
| 1 | Race |
| 2 | Sex |
| 3 | Sex * Race |
| 4 | Age |
| 5 | Age * Race |
| 6 | Age * Sex |
| 7 | Age * Sex * Race |
| **8** | **Area** |
| 9 | Area * Race |
| 10 | Area * Sex |
| 11 | Area * Sex * Race |
| 12 | Area * Age |
| 13 | Area * Age * Race |
| 14 | Area * Age * Sex |
| 15 | Area * Age * Sex * Race |

0 and 8 TYPELEVL (overall and area) are default values.

*If the MSALEVL parameter in CONTROL.SAS is set to 0 then each area will be defined as a county; if it is set to 1 then each urban area will be defined as a MSA and each rural area will be defined as a county.*

For example, if you are using a state inpatient hospital database, you might specify the following TYPELEVL values:

0 – provides overall rates for your entire state
4 – provides overall rates for your state, broken down by age groups
8 – provides rates for areas within your state
12 – provides rates for age groups within these areas

6. In CONTROL.SAS, if you wish to create an ASCII (comma delimited) text file that you can import into a spreadsheet, then set the TEXTA2 parameter to 1. Otherwise, the default value of zero for TEXTA2 will skip around the code that would generate this file.

7. In CONTROL.SAS, if you wish to create an ASCII (comma delimited) text file that you can import into a spreadsheet, then specify the path (including the file name) for this ASCII text file in the "FILENAME PQTEXTA2" statement.

## *Program 5: PQSASA3.SAS*

The PQSASA3 program calculates age and sex risk-adjusted rates for each PQI (overall rates and rates by area), and then calculates smoothed rates.

Input:

1. SAS format library (FMTLIB) created from the PQFMT program.

2. The discharge-level file that was created with the PQSASA1 program.

3. The SAS dataset with summary records that was created with the PQSASA2 program.

4. A text file with Census area residential populations, stratified by area, sex, and discrete age categories. Five such files are currently provided along with the PQI module software. The files are QICTY**A**97, QICTY**A**98, QICTY**A**99, QICTY**A**00, and QICTY**A**01. The user should select the file for the year that best matches the user's discharge datafile. The text file does not need to be converted to a SAS file for use with the software.

5. A text file (MEANSA00.TXT) containing overall means from the reference SID dataset for use in the risk adjustment process. This file is provided to you as part of the PQI module. The text file does not need to be converted to a SAS file for use with the software.

6. A text file (COVARA00.TXT) containing regression coefficients from a regression that was run on a reference SID dataset. These coefficients will be used in the risk adjustment process. This file is provided to you as part of the PQI module. The text file does not need to be converted to a SAS file for use with the software.

7. A text file (MSXA00.TXT) containing three arrays for use in the smoothing process. The arrays contain noise estimates, signal estimates, and mean area rates for each PQI. This file is provided to you as part of the PQI module. The text file does not need to be converted to a SAS file for use with the software.

Output:

1. A SAS dataset with an overall summary record and with area-level summary records that contain the three types of indicator rates, along with the components of the initial raw rates. Specifically, the file contains the observed rates (OAPQxx variables), the risk-adjusted rates (RAPQxx variables), the smoothed rates (SAPQxx variables), the counts of outcomes that formed the numerators of the observed rates (TAPQxx variables), and the residential population totals that formed the denominators of the observed rates (PAPQxx variables).

2. An optional ASCII (comma delimited) text file that the user can then import into a spreadsheet. This text file contains the same information as the SAS output dataset. The text file will contain the OAPQxx observed rates and their TAPQxx and PAPQxx components, the RAPQxx risk-adjusted rates, and the SAPQxx smoothed rates. Each record or row in the text file will correspond to a different area.

3. A PROC MEANS (with N, NMISS, MIN, MAX, MEAN, and SUM) of the area-level summary records that shows statistics for the OAPQxx observed rates and their TAPQxx and PAPQxx components, the RAPQxx risk-adjusted rates, and the SAPQxx smoothed rates. These means will only be generated if the user included a value of 8 for the TYPELEVL parameter discussed in item 5 in the "Changes" section for the previous PQSASA2 program.

4. A PROC CONTENTS of the output SAS summary dataset is output to hardcopy.

5. A PROC PRINT of the output summary dataset is provided at the end of the hardcopy printout. If the user does not wish to generate this printout, then the global “PRINT” parameter in CONTROL.SAS code should be set to zero.

Changes:

1. The “FILENAME CONTROL” statement at the beginning of the program must be changed to reflect the location of the CONTROL.SAS code.

2. In CONTROL.SAS, specify the locations (paths) of the input and output SAS files in the “LIBNAME IN3” and “LIBNAME OUT3” statements.

3. In CONTROL.SAS, specify in the “INFILEA3” statement the name of the discharge-level file that you created with the PQSASA1 program.

4. In CONTROL.SAS, specify the path (including the file name) for the ASCII population text file in the "FILENAME POPFILE3" statement. The file name will be QICTY**A**97.TXT, QICTY**A**98.TXT, QICTY**A**99.TXT, QICTY**A**00.TXT or QICTY**A**01.TXT. These files are provided along with the software. The file used should be the one that is closest to the year associated with your particular data. These files contain Census residential estimates by area, gender, and discrete age categories.

5. In CONTROL.SAS, specify in the "FILENAME MEANSA" statement the path (including the file name) for the ASCII text file that contains overall means from the reference SID dataset. The name of the file is MEANSA00.TXT.

6. In CONTROL.SAS, specify in the "FILENAME COVARA" statement the path (including the file name) for the ASCII text file that contains regression coefficients from a regression that was run on reference SID dataset. The name of the file is COVARA00.TXT.

7. In CONTROL.SAS, specify in the "FILENAME MSX" statement the path (including the file name) for the ASCII text file that contains the estimates to smooth the indicator rates. The name of the file is MSXA00.TXT.

8. In CONTROL.SAS, specify in the “INFILEA2” statement the name of the summary file that you created with the PQSASA2 program.

9. In CONTROL.SAS, specify in the “OUTFILA3” statement the name that you wish the output file to have.

10. In CONTROL.SAS, if you wish to create an ASCII (comma delimited) text file that you can import into a spreadsheet, then set the TEXTA3 parameter to 1. Otherwise, the default value of zero for TEXTA3 will skip around the code that would generate this file.

11. In CONTROL.SAS, if you wish to create an ASCII (comma delimited) text file that you can import into a spreadsheet, then specify the path (including the file name) for this ASCII text file in the "FILENAME PQTEXTA3” statement.

## Reviewing the printed output.

This section contains tips for reviewing some of the printed output from the PQI module. These tips are oriented toward explaining the interrelationships between printout items from different programs and hopefully will help to reveal the nature and structure of the module outputs. For guidance in interpreting the results, see the next section which begins on page 24.

### *PQSASA1.SAS*

The initial printout from the PQSASA1 program contains **PROC MEANS** for all of the numeric variables in the output discharge-level dataset. It will contain information for the newly constructed TAPQxx flag variables that will later form the numerators for the indicator rates. For each TAPQxx flag variable:

- The SUM will contain the total number of observations in the dataset that were found to have the particular outcome of interest.

- For the majority (14 of 16) of the indicators, the MEAN, MINIMUM, and MAXIMUM will usually be the value one since the flag variables have either been set to missing ('.') or to a value of one. The two exceptions to this are the two indicators (#2, for Perforated appendix, and #9, for Low birth weight) that are based on subsets of the hospitalized population rather than the area residential population (see Table 1 on page 5). For these two indicators, a value of zero was assigned to the TAPQxx flag if a particular observation was part of the population for the rate denominator but did not have the particular outcome of interest to be included in the rate numerator. So for example, TAPQ02 = 0 implies a patient who had an appendectomy performed, but did not have a perforated appendix.

- For the two hospital-based indicators (#2 and #9), the MEANs will contain a close approximation of the eventual overall observed indicator rates. The values will change slightly after PQSASA2 has applied additional parts of the indicator definitions.

- N lists the number of observations in the dataset with non-missing values. For the 14 area-based indicators, N for TAPQxx will be the same as the SUM. For the two hospital-based indicators, N will contain the denominator for the observed indicator rate.

- There may be differences in the output from PQSASA1.SAS and PQSASA2.SAS and PQSASA3.SAS programs based on missing data. If any cases are missing the FIPS codes they will be included in the output from PQSASA1.SAS but will be excluded from the subsequent analyses (the second and third programs).

The PQSASA1.SAS printout also contains a **PROC CONTENTS** that lists all of the variables in the output dataset, including the character variables that are present, and a **PROC PRINT** of the first 24 observations in the output dataset.

### *PQSASA2.SAS*

The purpose of this printout is to provide the observed rates for all area-level indicators. The default printout shows results by area, but other levels of output (as shown in Table 5) can be specified as well. By default, each area will be defined as a county; however, if the user sets the MSALEVL parameter in the CONTROL.SAS program to 1 then each urban area will be defined as a MSA rather than a county and each rural area will be defined as a county. Note that risk-adjusted and smoothed rates are shown in the output from the next program.

The printout from the PQSASA2 program contains **PROC MEANS** on the area-level summary records for all of the numeric variables in the output summary dataset. It will contain information for the newly constructed OAPQxx rates, the PAPQxx denominators, and the TAPQxx numerators.

- The _TYPE_ variable described in the first row of the MEANS table identifies the stratification level for the records in the output data set. The _TYPE_ variable corresponds to the TYPELEVL parameter values that were specified (see item 5 in the "Changes" section for the PQSASA2 program on page 17). In this case, _TYPE_ always assumes the value of 8, since only the area-level records are selected.
- The N statistic for _TYPE_ contains the number of areas in the output summary data set. A TAPQxx numerator variable with a value for N that is lower than the N value for _TYPE_ indicates that there were no outcomes of interest for some of the areas.

- The MINIMUM value for the perforated appendix and the low birthweight TAPQxx numerators will usually be zero since values of zero were assigned for observations that were part of the population for the rate denominator but did not have the particular outcome of interest to be included in the rate numerator. For the other 14 indicators, based on the residential area population, the MINIMUM value will be one or higher.

*For the observed rates, the User should **NOT** report the MEANS displayed here, but instead should refer to the overall means in the output dataset that are also present in the subsequent PROC PRINT. The MEANS given here are means of area means; i.e., all areas are given equal weight, regardless of the populations in the different areas.*

- The SUMs for the counter variables TAPQxx and PAPQxx will contain overall file totals for these data elements. The SUMs for the observed rates have no intuitive meaning

If the "PRINT" parameter in the CONTROL.SAS program is set to one, the subsequent **PROC PRINT** output contains a complete dump of the output summary file. Listed for each record are the stratification values, the PAPQxx population denominators for the 16 indicators, the TAPQxx outcome numerators for the 16 indicators, and the 16 OAPQxx observed rates.

## *PQSASA3.SAS*

*The **PROC PRINT** at the end of this program provides your **final output**. (This printout appears if the "PRINT" parameter in the CONTROL.SAS program is set to one – the default setting.) It lists the numerator and the denominator, along with the observed, risk-adjusted, and smoothed rates for all indicators, as shown in the example output below.*

SAMPLE DATA
PROGRAM A3 PART III
AHRQ PREVENTION QUALITY INDICATORS: AREA-LEVEL MERGED FILES
FINAL OUTPUT
Indicator # 01: DIABETES SHORT TERM CX

| OBS | AREA | TAPQ01 (Numerator) | PAPQ01 (Denominator) | OAPQ01 (Observed) | RAPQ01 (Risk Adj) | SAPQ01 (Smoothed) |
|---|---|---|---|---|---|---|
| 1 | | 1258 | 25265496 | 0.000050 | 0.000049 | 0.000049 |
| 2 | 4001 | 6 | 41502 | 0.000145 | 0.000141 | 0.000213 |
| 3 | 4003 | 1 | 80446 | 0.000012 | 0.000015 | 0.000075 |
| 4 | 4017 | 8 | 59517 | 0.000134 | 0.000134 | 0.000189 |
| 5 | 6023 | 10 | 89183 | 0.000112 | 0.000112 | 0.000154 |
| 6 | 6025 | 19 | 94657 | 0.000201 | 0.000199 | 0.000229 |
| 7 | 6027 | 4 | 13797 | 0.000290 | 0.000297 | 0.000366 |
| 8 | 6033 | 6 | 41004 | 0.000146 | 0.000157 | 0.000225 |
| 9 | 6093 | 6 | 32246 | 0.000186 | 0.000194 | 0.000265 |
| 10 | 680 | 108 | 420693 | 0.000257 | 0.000255 | 0.000260 |
| 11 | 2620 | 22 | 80950 | 0.000272 | 0.000263 | 0.000287 |
| 12 | 2840 | 23 | 580571 | 0.000040 | 0.000037 | 0.000046 |
| 13 | 4120 | 22 | 936017 | 0.000024 | 0.000027 | 0.000033 |
| 14 | 4472 | 400 | 11156766 | 0.000036 | 0.000034 | 0.000034 |
| 15 | 6200 | 199 | 2072994 | 0.000096 | 0.000096 | 0.000098 |
| 16 | 6922 | 66 | 1215938 | 0.000054 | 0.000055 | 0.000059 |

Observation 1 (with AREA = *blank*) is the overall average for the entire dataset (TYPELVL = 0). The remaining observations are individual areas (TYPELVL = 8).

You may wish to express the results in more understandable terms:

- Multiply the area rates by 100,000 to express them as a rate per 100,000 population (e.g., 0.000494 * 100,000 = 49.4 cases of Congestive Heart Failure (CHF) per 100,000 population).

The MEANS table is generated from the permanent area-based output file. The means are provided just for the area-level summary records.

- The N for all but the TAPQxx counter variables should contain the number of different areas (MSAs and/or counties) in your database.

- The MEANs, MINIMUMs, and MAXIMUMs have their normal meaning and provide comparisons among the different areas in your database. Note that the MAXIMUMs for the counter variables (the TAPQxx and PAPQxx variables) are associated with specific areas, and therefore these MAXIMUMs may not match those in the prior PQSAS2 printouts since that run typically will include a record for the entire database.

*For the (observed, risk-adjusted, or smoothed) rates, the User should **NOT** report the MEANS displayed here, but instead refer to the overall means in the output dataset that are also present in the subsequent PROC PRINT. The MEANS given here are means of area means; i.e., all areas are given equal weight, regardless of the number of discharges in the different areas.*

- The SUMs of the counter variables (the TAPQxx and PAPQxx variables) yield the overall database totals.

## Interpreting the results.

After performing all of the steps as previously outlined, users will have three estimates of performance for each PQI – observed, risk-adjusted, and MSX smoothed. This section will outline how these estimates can be used.

The Prevention Quality Indicators are not intended as definitive quality measures. But they are useful, low-cost measures that can potentially illuminate differences across areas in hospital admission rates for ambulatory care sensitive conditions (ACSC).

Since there are no "right rates" established for most indicators, it is often best to compare area-level rates with other similar areas. These "peer groups" would ideally be as similar as possible in potentially important factors, such as socioeconomic status of the population, and urban or rural location.

Performance on a single PQI often cannot reliably indicate actual quality differences. For this reason, some indicators have been developed as measure sets. For instance, four indicators are related to diabetes – uncontrolled diabetes, diabetes short-term complications, diabetes long-term complications, and lower-extremity amputation among patients with diabetes. Examining these indicators together is likely to produce a more complete picture of overall quality of care for this condition.

### *Observed means.*

Interpretation of the observed (raw) PQI rates may be complicated by concerns over noise and bias. However, the observed rates do not require complex statistical methods as do the risk-adjusted and MSX smoothed rates. For this reason, they are the most intuitive estimates of PQI performance. Observed rates can be used as a baseline measure when comparing to risk-adjusted and MSX smoothed rates, to determine the impact of risk adjustment/smoothing.

### *Risk-adjusted rates.*

Simply put, risk-adjusted rates are the estimated performance of areas on the PQIs if those areas had an 'average' casemix, as defined by age and sex. This average casemix is estimated and included in the software using estimates from 29 States in AHRQ's State Inpatient Databases (SID). Estimates of the average casemix do not reflect only the areas in the dataset being analyzed, but rather the distribution in age and sex among the areas in the SID data.

All area indicators are risk-adjusted using age and sex, except for low-birth weight, which is not risk-adjusted. Ideally, one might like to adjust for the underlying health status of each area, but population-based measures of health status by demographic groups are not readily available.

Users should compare risk-adjusted rates with the observed rates. This will indicate the impact of risk adjustment and presumably casemix on indicator performance. Areas with large changes in performance between observed and risk-adjusted rates presumably have a more or less complex case mix.

The construction of risk-adjusted rates in the PQI module follows the method described in Greene, Econometric Analysis (2nd Ed), 1993, Section 16.4.2 on calculating fixed effects in a linear model. Two important features of this approach should be noted. First, the risk-adjusted rates are coefficients on area level covariates ('fixed effects') in a linear model for each PQI that also accounts for patient demographics (age and sex). Generally, these coefficients reflect the relative performance on the PQI relative to a reference or 'omitted' group (e.g., males age 65-69). As a result, the risk-adjusted rates for each area would reflect performance on the PQI for the reference group only, making interpretation and comparison with the observed raw rates problematic. Rather, the approach incorporated into running the model is to first subtract the overall PQI mean for each covariate (e.g., age category 1, age category 2, etc.) prior to running the model. The benefit of this approach is that the risk-adjusted rates (the 'fixed effects' in the model) now

reflect the estimated performance on the PQI for each area if that area had the 'average' casemix among all areas in the estimation sample.

Second, because the model is linear, rather than logistic, the estimated risk-adjusted rates may be negative (that is, the model does not require the rates to be between zero and one). This might happen, for example, if the observed rate is close to zero, and the area has a more severe than average casemix. The model would predict that the area's performance with an 'average' casemix would be less and subtract some adjusted amount from the observed rate. If that observed rate is already nearly zero, the adjusted rate may be negative. A negative rate means that the observed rate is close to zero, and that the area has a less-severe-than-average casemix. In order to facilitate the interpretation of the estimates, the PQI module imposes a lower bound of zero; i.e., negative values are set to zero.

## *MSX smoothed rates (optional).*

The PQSASA3 program also yields the estimated MSX smoothed rates for each PQI. These smoothed rates are estimated using multivariate signal extraction (MSX). MSX is a Bayesian smoothing technique. Generally, indicators defined on relatively small populations or for relatively rare events are very noisy measures. In other words, many other factors other than quality can influence the observed rate. For factors that we can observe, such as the patient's age or gender, risk adjustment can account for their influence. There are many other clinical and non-clinical factors that we cannot observe. However, the MSX method estimates how much of an impact random differences in these factors have on the observed rate. The program accomplishes this by adjusting toward the overall area mean according to the degree of reliability of the specific area indicator.

Just as risk-adjustment levels the playing field across areas by "equalizing" population characteristics, smoothing levels the playing field by removing random variation in the measures across areas and over time. In essence, smoothing tells us how persistent we would expect an area's rate to be from year to year.

*The smoothed rates are intended to help you look for how best to present the data. For example, if smoothed rates vary considerably from risk-adjusted rates, you might wish to examine risk-adjusted performance by combining data for several years rather than examining a single year's performance, or just compare areas within peer groups.*

The MSX method estimates reliability by looking at area performance among multiple indicators and across multiple years for a single indicator. If areas that perform high in one year also tend to perform high in the next, and areas that are low in one indicator tend to perform low in other related indicators, then the methods incorporate this information in forming an estimate of how much of an area's performance on a single indicator in a single year is "signal" and how much is "noise". PQSASA3 reports the "signal" estimate, after removing the "noise".

These estimates are referred to as "smoothed" estimates because when the raw data is reported annually on a graph, for example, the line connecting the points may show massive fluctuations, seemingly with no apparent pattern. Some of these fluctuations are due to random differences in factors that influence performance, so once the impact of these factors is removed, fluctuations are reduced, or smoothed out. Improvements in performance will still be observed, however, they will only be observed when it is likely that these improvements are actually true differences, and not simply random variation.

The estimates of how related the Prevention Quality Indicators are among measures and over time for a typical area are based on estimates using the 29 states in the SID database. Therefore, PQSASA3 does not require multiple areas, measures, and years of data to operate. Instead shrinkage estimates were calculated from the combination of 29 statewide HCUP datasets to which the MSX had been applied. These shrinkage estimates are the best estimates of the impact that MSX smoothing would have on a particular PQI if MSX could be applied to the dataset being analyzed. These shrinkage estimates are then applied to the adjusted rates for each PQI, giving 'smoothed' estimates of the rates.

The smoothed estimates represent the best guess of what the area performance would be for any given PQI if the data contained only a minimal amount of noise. This has several implications. First, users can be more certain that changes observed over time represent true differences in performance, rather than random variation. Second, in any given year, users can be more certain that the rate does not reflect only random noise. This can best be seen by comparing the smoothed estimates with the observed raw rates, which will indicate the impact of MSX on that indicator. A large impact means that the observed rate of the indicator is less precise, meaning it contains more noise relative to other indicators sustaining less impact. Low precision can result from small numbers of observations, or for relatively rare events. Therefore, the impact of the MSX estimate will in general be greater for infrequent events. Plotting the MSX estimates over time will reflect more persistent performance from year-to-year, and the correlations among indicators will generally be stronger.

## *Comparing observed, risk-adjusted and smoothed rates*

The purpose of the analysis determines which rates the user should look at in evaluating the performance of an area. If the user's primary interest is to focus on a particular area without any comparisons to other areas, simply examine the overall observed rate for the entire area, as well as further breakdowns by age, sex, and race/ethnicity which are the output created by the PQSASA2 program. Trend data are recommended for identifying changes over years.

If the purpose of the analysis is to compare the performance of a particular area with national, state, or regional averages or performances of other selected areas, then both observed and risk-adjusted rates should be examined. Variation in observed rates across areas is attributable to a variety of factors including differences in population demographics, disparity in access to and quality of care or other area characteristics ('systematic factors'), and random factors (non-systematic factors or 'noise'). Comparing observed and risk-adjusted rates can reveal if there is any difference between the area's population and the population of other areas. If the difference is minimal, one can compare the area's observed rate with the overall average across all areas. However, to account for differences in population demographics among different areas, risk-adjusted rates should be used for area-by-area comparisons.

After removing the impact of population demographics, if users want to examine potential existence of random factors, comparisons can be made between the risk-adjusted rates and the smoothed rates to determine if the differences in risk-adjusted rates across areas are due to systematic, as opposed to random, factors. The following two tables provide guidance on how to interpret differences in observed, risk-adjusted, and smoothed rates.

**Comparing Observed Rates with Risk-adjusted Rates**

*Purpose:* *To identify if there is any difference in the demographic composition of the area relative to the demographic composition of all areas combined*

| | |
|---|---|
| Observed rate **>** Risk-adjusted rate | The area's population has a ***higher*** risk of being admitted to hospitals for the condition due to its demographic composition (for example, older or a greater proportion of a higher-risk gender). |
| Observed rate **<** Risk-adjusted rate | The area's population has a ***lower*** risk of being admitted to hospitals for the condition due to its demographic composition (for example, younger or a greater proportion of a lower-risk gender). |
| Observed rate **=** Risk-adjusted rate | The area's population is similar to other areas in demographic composition, suggesting that the demographic composition is not a contributing factor to the area's performance for the indicator. |

**Comparing Risk-adjusted Rates with Smoothed Rates (optional)**

*Purpose: To determine if differences in risk-adjusted rates across areas are due to systematic or random factors after demographic composition is taken into account.*

| | |
|---|---|
| Risk-adjusted rate ***similar to*** Smoothed rate | The area's risk-adjusted rate is more precise, an accurate reflection of the area's systematic ('true') rate. If there is any difference between the risk-adjusted rate and the overall area average[1], such difference is likely to be due to systematic factors. |
| Risk-adjusted rate ***different from*** Smoothed rate | The area's risk-adjusted rate is less precise[2]; that is, some of the difference between the risk-adjusted rate and the overall area average is due to random ('noise') factors. The smoothed rate is thus a more accurate estimate of the systematic rate than the risk-adjusted rate. Comparing the smoothed rate to the overall area average reveals more accurately the 'true' performance of the area. |

Notes:
1. The overall area average refers to the weighted average risk-adjusted rates across all areas.
2. The risk-adjusted rates for areas with small population are generally less precise.

Generally, subtracting the observed, risk-adjusted and smoothed rates can help to reveal differences among these rates. To get a sense of the magnitude, the most straightforward approach is to estimate the difference in terms of the number of cases in the numerator. For example, for an area with 35,000 pediatric population (under 18 years of age), a difference of 0.0008 between the observed rate and risk-adjusted rate would mean 28 hospital admissions. Whether the difference between rates is statistically meaningful will depend on the size of the denominator, and the sampling variability for each indicator. Users can apply a rough rule of thumb to estimate the required difference: 1.4* SQRT(p*(1-p)/N), where p is the rate and N is the denominator of the indicator.

For example, if the observed rate for an indicator is p=0.0045 and the denominator is N=35,000, applying the formula to these numbers will produce: 1.4* SQRT(0.0045*(1- 0.0045)/35,000) = 0.0005. That is, for the risk-adjusted rate to be significantly different from the observed rate, the absolute difference between the two rates has to be equal to or greater than 0.0005.

## Benchmark timings

The benchmark runtimes given below are from runs made on a Windows NT 4.0 workstation, with a Pentium II 400 MHz processor, 256 MB of RAM, and a SCSI hard drive. The dataset used contained 763,363 observations with 15 diagnosis fields and 15 procedure fields.

| Step | Run time (in seconds) |
|---|---|
| Run PQFMTS to create the format library | 002 |
| Run PQSASA1 | 367 |
| Run PQSASA2 | 022 |
| Run PQSASA3 | 244 |

## User Support

We would like to receive your feedback on the AHRQ Quality Indicators.

Our Internet address for user feedback is: support@qualityindicators.ahrq.gov

We offer a listserv to keep you informed on the Quality Indicators. We encourage you to sign up for this free service. All you need is a computer, Internet access, and an E-mail address. It works just like other electronic distribution lists.

Here's how to register:

1. Send an E-mail message to: listserv@qualityindicators.ahrq.gov.
2. On the subject line, type: Subscribe. For example:

   Subscribe
3. In the body of the message type: sub Quality_Indicators-L and your full name. For example:

   sub Quality_Indicators-L John Doe
4. You will receive a message confirming that you are signed up.

If you have any questions, contact AHRQ QI Support at the e-mail noted above. If you do not receive a confirmation message call (888) 512-6090.