Medicaid Payment Accuracy Measurement Project: Year 1 Pilot Results and Assessment

Centers for Medicare and Medicaid Services
Center for Medical and State Operations
Financial Systems and Quality Group

December 2002

Table of Contents

I. PURPOSE OF THE MEDICAID PAYMENT ACCURACY MEASUREMENT PROJECT

On July 3, 2001, CMS published a State Medicaid Director letter inviting states to apply to participate in a demonstration project to develop a payment accuracy measurement (PAM) methodology for Medicaid. CMS had already implemented a payment accuracy measurement program for Medicare, the Comprehensive Error Rate Testing (CERT) program. In addition, three states—Kansas, Illinois, and Texas—had conducted independent studies of the accuracy of Medicaid payments. A national measure of payment accuracy, obtained by appropriately aggregated rates from the state level, would provide an overall estimate of the accuracy of Medicaid payments to providers.

The demonstration was designed to provide states with the opportunity to work collaboratively to develop and test methodologies for measurement of accuracy of payments made for Medicaid services. During the first year, CMS planned to fund six to eight states but attracted promising applications from nine states: Louisiana, Minnesota, Mississippi, New York, North Carolina, North Dakota, Texas, Washington, and Wyoming.[1] These states represent the diversity of the 50 state Medicaid programs including differences in size and population density, use of managed care versus fee-for-service, practice patterns, and program integrity development.

The goal of the Year 1 demonstration was to develop a core model that could be further tested for use at the state and national levels. The initial experiences of the demonstration states, which are summarized in this report, informed the development of models for measuring payment accuracy in both fee-for-service and managed care Medicaid programs. While the activities to date are only the first step in developing a nationwide system for measuring Medicaid payment accuracy, the lessons learned from these demonstration projects are highly useful in formulating and refining workable approaches.

Section II of this report provides an overview of the methods pursued in the study. Section III summarizes the states' interpretation of this objective, outlines their activities during the first year of the project, and then presents a side-by-side comparison of them. Section IV highlights certain aspects of these unique methodologies to illustrate what they share in common and how they diverge with a focus later in the section on the breakdown of payment error rates reported by states. Section V discusses several of the issues that states encountered in the first attempt to engage in a project such as this and the strategies they used to overcome these difficulties. Section VI outlines the basic PAM models for the Year 2 studies involving both fee-for-service claims and capitated managed care payments. Finally, Section VII issues the conclusions that can be formed on the basis of what was learned in Year 1 of the project with an emphasis placed upon what can be improved in subsequent years.

II. METHOD

Each Year 1 state submitted a proposal outlining its approach to measuring payment accuracy in Medicaid. Although the CMS solicitation provided broad parameters for the basic expectations of each state's project, states were free to propose a variety of methods and approaches to measuring payment accuracy in Medicaid. Many states based their approaches on the methods employed by the small number of states that had developed their own PAM studies, the Medicare program, and/or other Medicaid audits.

Each participating state submitted quarterly reports, participated in regular teleconferences, and attended two two-day meetings in the Washington/Baltimore area to share their knowledge and experience with CMS and other participants. The states were also asked to submit final reports detailing the individual methods used to determine payment accuracy, and the quantitative and qualitative findings from their Year 1 studies. As of mid-December 2002, final reports had been submitted by five of the nine pilot states (Louisiana, Mississippi, North Dakota, Texas, and Wyoming). North Carolina and Washington submitted preliminary reports. Minnesota and New York experienced significant delays and will not complete their Year 1 projects until June 2003.

CMS, working with the pilot states and its consultant, used the experiences of the pilot states, other states that had explored Medicaid payment accuracy measurement, and methods used by the Medicare CERT and OIG audits to develop core models for PAM for fee-for-service and managed care. This inclusion of managed care reflects the fact that it represents a major provider of health care services within a large majority of the states' Medicaid programs. States selected to participate in Year 2 of the grant are evaluating the effectiveness and efficiency of these core models (see Chapter VI of this report for more information on the Year 2 model). In Year 2 CMS, the pilot states, and the consultant will continue to share, discuss, and refine study methodologies that represent best practices.

III. SUMMARY OF INDIVIDUAL STATE METHODOLOGIES AND RESULTS

As noted in the previous chapter, the first year of the PAM pilot project was designed to allow participating states to design and test a variety of methodologies, in hopes of identifying best practices that could be further refined in future project years. In designing their methodologies and following through on these research designs, the states offer much in the way of innovation and insight into the process of measuring payment accuracy. However, it is also important to recognize that for this reason, the payment accuracy rates reported by the states are not directly comparable to one another. This chapter summarizes each state's approach and key findings; most of this information has been taken directly from the reports submitted by the pilot states.

A. Louisiana

Louisiana's Medicaid Program Integrity Section assessed the state's Medicaid payment accuracy rate. Louisiana sampled a total of 600 claims, with a total value was $126,521, filed during the quarter beginning on July 1, 2001 and ending on September 30, 2001. The state divided these claims into two separate strata: paid claims and denied claims. Among the 600 claims sampled, 482 belonged to the paid claims stratum and were valued at $83,686 while 118 belonged to the denied claims stratum and were valued at $42,835.

An inter-disciplinary research team performed a three-part review to determine the accuracy of these claims by performing a claims processing review, a medical record review, and a contextual claim analysis of the paid claims and a claims processing review of denied claims.

The team found that of the 118 denied claims that were reviewed, all were correctly processed, and that 22 of the 482 paid claims had errors that affected payment. Of the 22 claims with errors, 8 had no documentation, 7 had insufficient documentation, 4 had dispensing problems, 2 had policy/coverage violations, and 1 had a coding error. The total payment error amount of these 22 claims was $3,034, which translated into a 97.80% accuracy rate across all 600 claims. Broken down by stratum, the accuracy rate for the denied claim stratum was 100.00% and the accuracy rate for the paid claim stratum was 92.92%.

B. Minnesota

The Minnesota DHS Encounter Data Validation Study was intended to validate the quality and completeness of the managed care encounter data that is used to develop risk-adjusted capitation rates. The validation process consists of evaluating the state data requirements, reviewing the health plans' capacity for producing accurate and complete data, analyzing DHS' submitted encounter data for completeness and accuracy, performing medical record reviews to validate the submitted encounters, analyzing the medical care episodes, and performing a value-based validity analysis. To assist in this process, DHS contracted with Mercer Government Human Services Consulting.

For this study, Minnesota generated a statistically valid sample using calendar year 2000 Adjusted Cost Group (ACG), encounter, and eligibility data. The sample size was 95 charts per plan but upon adding over-sample charts, the total number of charts pulled for review was 110 per health plan. A list of the recipients selected in the sample was distributed to each of the

health plans and the plans subsequently identified all of the providers that the selected recipients have received services from during the 6-month period of July 1, 2000 to December 31, 2000. Currently, the project is experiencing significant delays due to difficulties with the data exchanges involved and unresolved issues with the medical record review data abstraction requirements. The state applied for and received an extension from CMS through March 31, 2003 and is hopeful that it will be able to complete all tasks and have a final report by that time.

C. Mississippi

The Mississippi Division of Medicaid (DOM) assessed the state's payment accuracy rate by selecting an unstratified, statistically valid sample size of 370 eligible beneficiaries who received Medicaid services between July and September of 2001 and incurred a total of $359,859 in Medicaid expenditures during this time. All claims for every beneficiary in the sample, which totaled 3,559 claims, were reviewed for medical necessity, policy adherence, systems payment accuracy, and beneficiary confirmation of services. Of the 370 beneficiaries audited, 202 (54.6%) had all their claims accurately paid and of the 3,559 individual claim items, a total of 3,024 (85%) were paid accurately.

Of the 535 errors found, the type of error most commonly recorded involved a lack of medical necessity for the service provided, which accounted for 283 (53%) of the total errors. With respect to the total cost of these payment errors, of the $359,859 in payments to the 370 beneficiaries in the sample universe, a total of $333,895 was paid correctly. The cost of the errors for the sample represented 7 percent of the total payments for the beneficiary universe in the study, meaning that the payment accuracy rate (percent of total dollars paid) for the Mississippi Division of Medicaid was 92.8%.

D. New York

For its Medicaid FFS Payment Accuracy Measurement study, New York State issued a Request for Proposals (RFP) in 2001, selected a vendor, Island Peer Review Organization (IPRO), and issued the final award letter on August 20, 2002. It then drew a stratified, random sample consisting of 1,004 claim lines from a universe of 13,619,650 claim lines, which were distributed among four strata: Inpatient (300 claim lines); Pharmacy (302 claim lines); Skilled Nursing Facility (200 claim lines) and Other services (202 claim lines). Each of the four strata was further subdivided by dollar values into sub-strata. Its selected vendor, IPRO, received the sample on September 25, 2002 and is currently working to develop forms, letters and systems for the project. As a result of significant delays in the RFP and contract process, the state received approval from CMS to extend the budget period of the Year 1 PAM grant through June 30, 2003.

Additionally, on July 8, 2002, CMS gave New York approval to use the remainder of the Year 1 PAM project funds to conduct a Medicaid PAM study of managed care. A sample was selected consisting of 1,400 capitation payments from a universe of 2,934,941 capitation payments for the period of April 1, 2002 to June 30, 2002. The sample, corresponding recipient profiles and Welfare Management System (WMS) and Medicaid Management Information Systems (MMIS) data, have been obtained for the managed care study and review of the sample claims has begun.

E. North Carolina

For the past five years, North Carolina's Department of Medical Assistance, Program Integrity (DMA-PI) in cooperation with the State Auditor has annually reviewed Medicaid claims for payment accuracy under the Single State Audit Act. Correspondingly, its PAM project consisted of two separate processes. The first involved a re-review of 300 claims from the prior year's sample (July 1, 2000 to June 30, 2001) using an enhanced process to show a comparison of the old process results versus the enhanced (PAM Grant) process. The second involved an initial review of 275 claims randomly drawn from the sample period between July 1, 2001 and June 30, 2002. These were pulled from a stratified, random sample of paid Medicaid claims that included Medicare crossover claims and were then stratified into ten strata based on their dollar range.

For each claim in the sample, medical records were requested and occasionally supplemented through phone interviews with recipients and/or caregivers to confirm receipt of services. Claims were then reviewed to determine if there was a duplicate claim, non-covered service, any third party or managed care responsibility and/or any other payment error. Additionally, claims were subjected to an enhanced review process involving (1) The contracted review of hospital and nursing home claims by North Carolina Professional Review Organization, the Medical Review of North Carolina (MRNC) and (2) An expanded medical record review of all paid claims and verification of receipt of services for home and community-based services.

The re-review of 300 claims from the prior year's sample using the enhanced process was completed with 14 claims being questioned totaling $7,498 in overpayments and 3 claims with underpayments that were identified but not included in the error rate. The result of the SFY 2000-2001 study was a finding of a 99.68% accuracy rate. The review of 275 claims for the most recent fiscal year was completed with 16 claims being questioned totaling $9,567 in overpayments and 2 claims with underpayments that were not calculated or netted against the overpayments. As of mid-December 2002, North Carolina had not completed and reported on its Year 1 project final results.

F. North Dakota

The Medical Services Division of North Dakota's Department of Human Services calculated the state's payment accuracy rate by choosing a random sample of Medicaid claims processed between October and December of 2001. The sample was stratified with 50 percent based on claims volume and 50 percent on expenditures. A total of 403 claims were selected, representing $440,595 in Medicaid payments. Detailed information from each of the claims was obtained from the state's data warehouse and a letter was sent to the providers requesting that they submit documentation supporting the claim. If the provider did not respond with the necessary information, follow up requests were made with a 96.5% rate of documentation submission. For those that submitted documentation, the documentation was reviewed to determine if the documentation supported the services reported on the claim and if the claim was properly adjudicated with errors being documented within one of the following six categories: documentation, coding, billing, processing, coverage issues, and medical necessity.

Of the 403 claims reviewed, a total of 29 had errors with 14 of those (valuing $20,989) attributable to a lack of documentation and 15 claims (valuing $5,700) classified as one of the other six error types identified, the most common of which was a documentation error in which the information that was sent by the provider did not substantiate the service provided. As a result, of the claims where a clear error was found, North Dakota calculated its payment accuracy rate to be 98.71% but this figure dropped to 93.90% (95.24% according to North Dakota's calculations) when expanded to include those claims where no documentation was submitted by the provider. With respect to both types of error, accuracy rates were lowest among claims for prescription drugs, outpatient professional services, and nursing facilities.

G. Texas

Texas conducted two studies for the PAM project, a Medicaid Fee-For Service (FFS) study and a Medicaid Vendor Drug Program (VDP) Pilot study. The samples for both studies were selected from the Medicaid claim data warehouse. The Comptroller's office contracted with professional consultants to perform the medical record reviews for both studies.

With respect to the FFS study, Texas randomly sampled health care claims for 800 Medicaid beneficiaries, which were selected from a universe of paid Medicaid claims during the timeframe of September 1, 2001 through November 30, 2001. All claims for each beneficiary on a particular date within that sample timeframe were pulled, for a total of 2,122 total claims. In order to confirm that a sample service occurred on the sample date and was performed by the billing provider, the review process included a beneficiary telephone interview, an online system review of the claim, and a utilization review of the provider's medical records for the sample service. The Medicaid FFS study had a total paid amount of $169,743 and a payment accuracy rate of 86.50%. The majority of the Medicaid errors were from providers, who either did not submit a medical record for the project team to review or submitted medical records without proper documentation of the sample service.

The Medicaid VDP Pilot study reviewed randomly selected prescription services paid on behalf of 225 sample beneficiaries from October 1, 2001 through December 31, 2001. This included 1,070 paid prescription services for a total paid amount of $46,936. The study used the same review criteria as the FFS study: beneficiary telephone interview, online system review of the related Medicaid medical records, and review of the original prescriptions. The Medicaid VDP Pilot study accuracy rate is 75.65%. The majority of the errors found were for pharmacies that did not submit a copy of an original prescription (written, fax, or telephone form) to confirm it met the Texas and federal prescription requirements.

H. Washington

Washington calculated a payment accuracy rate by using a dollar-weighted, stratified random sampling technique to select 560 claims from seven service categories, including inpatient hospitalization, outpatient hospitalization, physician, pharmacy, dental, medical vendor, and long-term care. These 560 claims were for services provided between July and September of 2001 and represented a total of $277,778 in Medicaid payments. For each of these claims, service recipients were contacted to verify service delivery and to compare service delivery claims with service delivery received. Additionally, claims were reviewed for accuracy by checking medical

documentation for type and level of service, medical necessity, and appropriateness of service. Claim payments were reviewed for conformance with policy, published rates, and schedules.

Washington experienced significant delays during its PAM project and was unable to complete its review of inpatient hospital claims by mid-December so the statistics included in its preliminary report did not include these reviews. Nevertheless, the state did report that of the 500 claims studied, 412 were accurately paid. The total dollar amount of these payments was $277,778 of which $273,411 was paid correctly. Correspondingly, the accuracy rate for these claims was 98.4%. Errors were most frequently attributed to insufficient or inappropriate documentation and accuracy rates were found to be highest among claims for long-term care and lowest for payments to medical vendors. However, project leaders noted that a sample size of 560 cases is likely too small to generate results that would be statistically significant at the 95 percent level of confidence and they recommended that it be increased to approximately 1,000 cases in order to reach this level of confidence.

I. Wyoming

The Office of Medicaid in Wyoming's Department of Health assessed the state's Medicaid payment accuracy rate by selecting a statistically valid, dollar-weighted, random sample of 300 claims from the 2nd and 3rd calendar quarters of 2001 (600 claims in sum), which represented a total of $141,335 in Medicaid expenditures. Wyoming reviewed claims paid for Medicaid services using several techniques: medical record review, telephone interview with the recipient, and a system accuracy review. Additionally, the medical record review was conducted with emphasis placed on checking both pre and post-payment procedures and compliance in addition to a check for correct keying, disposition of edits and audits, and whether the system and procedures in place were pointing to the correct payment rate for fee-for-service.

As a result, the research team found a total of 33 claims worth $2,599 that had been paid in error, creating a payment accuracy rate of 97.67%. The most frequent error type discovered was a lack of adequate documentation for services rendered and errors were most frequently found among claims for Home and Community Based Services.

J. Summary of State Findings

In the interest of helping to develop a national Medicaid payment accuracy rate, the individual state payment accuracy rates have been assembled and displayed in Table 1. When interpreting these figures, it is important to remember that flexibility was offered to states during the first year of the project. As discussed in more detail in the following chapter, states could choose how claims were sampled, what claims were included in the study, how they were stratified, and how they were reviewed for errors. This variability was useful in identifying best practices for the Year 2 core models, but also means that the differences in Year 1 payment accuracy rates must be interpreted with caution.

Table 1: Summary of State Payment Accuracy Rates

State	Sample Size	PAM Rate	95% Confidence Interval	Notes
Louisiana	600	92.9%	+/- 8.50%	Denied claims: 100%, overall: 97.8%
Minnesota	n/a	n/a		Study delayed
Mississippi	3,559	92.8%	[90.2%, 95.4%]	All claims for sample of 370 beneficiaries
New York	1,004	n/a		Study delayed, sample size for FFS study
North Carolina	300	99.7%		Rate is from SFY01, SFY02 rate pending
North Dakota	403	93.9%	[98.67%, 98.75%]	Rate includes undocumented claims
Texas	2,122	86.5%	[86.3%, 86.7%]	Sample size for FFS. VDP rate: 75.65%
Washington	500	98.4%		Rate will be updated with inpatient findings
Wyoming	600	97.67%		

IV. COMMON THEMES AND FINDINGS

While the particulars of each state's approach often varied, most followed the same three steps. First, states generated a sample from the universe of paid Medicaid claims (although a few also considered denied claims). Second, they gathered MMIS and medical record information on the sampled claims and reviewed those records in order to determine whether any errors were made. Third, they analyzed the results to identify trends, particularly among type of error and type of health care service involved. Within this general approach, participating states had the freedom to identify, create, and test a variety of approaches. Below we compare and contrast specific state approaches.

A. Sampling Methodologies

In choosing a sample of Medicaid claims, some states chose to stratify their samples. Since the payment process and cost for different types of health services often differ, some states chose to stratify their sample so as not to lump together services such as inpatient and outpatient care. Stratification also helped ensure that a significant number of claims from each service category would be chosen. However, it is worth noting that stratifying the sample is not the same as simply categorizing each claim as it was evaluated. First it affects the sampling methodology since claims are sampled randomly within each service type, which is different from sampling the entire population of claims randomly. Second, stratification helps states to develop separate payment accuracy rates for each type of service.

Some states also chose to use a completely random sample of Medicaid claims while others weighted the sample by dollar value. This is because a large number of Medicaid claims are for low-dollar services, such as prescription drugs. A random sample might result in a large number of those low-value services being selected, and fewer lower-volume, higher-cost services such as inpatient stays and nursing care. High dollar value claims are more likely to have high dollar errors, so the more claims with large dollar errors are sampled, the more accurate a payment accuracy rate can be attained. To ensure that sufficient high-cost services were included in the study, some states stratified by dollar value to ensure that a representative number of claims from a set of dollar ranges was selected. One state used a mixed method that weighted the sample partly by dollar value and partly by service type.

States could choose the sampling unit used to draw the sample. In this respect, states had the option of choosing a random sample from the universe of beneficiaries, claims, or claim line items during the sample period. Each option offered certain advantages and disadvantages and states pursued different paths. Louisiana, North Carolina, North Dakota, and Washington sampled claims, Mississippi sampled beneficiaries, and Texas used beneficiary-days as its sampling unit (see detail on Texas below). Wyoming initially chose a total of 600 claims to review but upon realizing that this amounted to 875 claim lines, decided instead to used the first line item per claim and calculate rates based on 600 claims.

States participating in the PAM project tested a variety of sampling approaches:

- Louisiana stratified its sample according to whether the claim had been paid or denied and then randomly sampled claims from each category by choosing a random starting

point and using a skip interval to select successive claims until a sufficient number were chosen.

- Mississippi randomly selected a sample of 370 beneficiaries and then reviewed all claims paid on behalf of those beneficiaries during the sample period between July 2001 and September 2001.

- North Carolina re-reviewed 300 claims that were sampled for its SFY 2000-01 study and then sampled 275 claims for the SFY 2001-02 study. Both samples were generated from a stratified, random sample of paid Medicaid claims that were drawn and then subdivided into ten strata based on their dollar range. The sampling universe included all fee-for-service claims including crossover claims with only $0 paid claims excluded.

- North Dakota also pursued a highly independent course by opting not to stratify the sample but weighing the sample with 50 percent based on the claims volume and 50 percent on expenditures.

- Texas pursued different strategies for its FFS study and VDP Pilot. For the FFS study, it identified eight strata of medical care services and then randomly selected 100 sample days for each of these strata between September 1, 2001 and November 30, 2001. For each date chosen, Texas randomly chose a beneficiary who received services on that day for a total of 800 beneficiaries, then reviewed all the services received by these beneficiaries on the day in question, which amounted to a total of 2,122 individual services considered. For the VDP Pilot, Texas selected a simple random sample of 225 dates from October 1, 2001 through December 31, 2001 and randomly chose one beneficiary who had filed a claim on each of those dates. It then identified all paid prescription claims within the sample time frame for each beneficiary within the study sample and came up with a total of 1,070 claims to be reviewed.

- Washington stratified its sample among seven service categories and dollar-weighted the claims in the sampling universe in order to determine the total number of claims it would randomly select from each stratum. Each stratum was subsequently divided into 4 sub-strata, from which the state selected 15, 25, or 30 cases depending on what type of service was being evaluated.

- Wyoming stratified its sample by date, weighted claims by their dollar values, and chose a random sample of 300 claims from the 2nd and 3rd quarters of 2001, thereby creating a total sample size of 600 claims.

Each of the sampling approaches used by the pilot states had advantages and disadvantages. To help ensure the most precise estimate for the Year 2 pilot, the core model will use a proportional, stratified random sample. The sampling strata will be by major provider categories defined as: inpatient hospital services, long term care services, independent practitioners and clinics, prescription drugs, home and community-based services, and other services and supplies.

In the PAM core fee-for-service model, sample sizes by stratum will be proportional to the dollar value of the line items represented by each stratum for the most recent four quarters.[2] That is, if inpatient hospital services represent 30 percent of the dollar value of total Medicaid claims, 30 percent of the sample of line items should come from the inpatient stratum. This will result in oversampling in strata for which the proportion of Medicaid payments is greater than the proportion of Medicaid line items, and undersampling in those strata for which the proportion of line items is greater than the proportion of Medicaid payments. Therefore, when calculating the payment accuracy rate, states must take this oversampling and undersampling by strata into account in order to create an unbiased estimate of the overall payment accuracy rate. CMS will advise states on how to calculate their final accuracy rates in order to ensure that they are unbiased.

B. Claims Review Procedures

The process of reviewing records for claims sampled by states has two components: retrieving information from the beneficiary, the provider, and/or the state's own databases, then reviewing this information to determine the accuracy of payment. The nine pilot states used a variety of approaches in both collecting and reviewing information.

Most states used similar methods to obtain claims information, relying mainly on the MMIS for claims processing information and on providers for medical record information. Most states sent written requests to providers for medical records, and many followed up with repeated letters and/or phone calls when providers did not respond in a timely manner. Mississippi was the only state that conducted unannounced on-site visits at doctors' offices, although it switched to written requests after this process proved to be too time-consuming.

States varied in the use of beneficiary interviews: Louisiana and North Dakota did not contact beneficiaries; Mississippi, Texas, and Washington attempted to conduct telephone interviews with all beneficiaries in the sample; and North Carolina and Wyoming interviewed only a subset of the sample due to resource constraints. However, the states that contacted beneficiaries did not use the interviews to discount or establish the validity of any claims reviewed and simply considered them as supplementary information.

Once records had been received and interviews had been conducted, states evaluated the materials to determine the accuracy of payments. States could design their own methods for conducting these reviews, and used a broad range of techniques and guidelines:

- Louisiana relied on a three-part review process to determine the accuracy of the service selected: a claims processing review, a medical record review, and a contextual claim analysis.

- Mississippi reviewed claims for four separate elements: medical necessity, policy adherence, systems payment accuracy, and beneficiary confirmation of services.

- North Carolina reviewed claims to determine if there was a duplicate claim, non-covered service, any third party or managed care responsibility and/or any other payment error. Additionally, claims were subjected to an enhanced review process

involving (1) The contracted review of hospital and nursing home claims by North Carolina Professional Review Organization, the Medical Review of North Carolina (MRNC), and (2) An expanded medical record review of all paid claims and verification of receipt of services for home and community-based services.

- North Dakota conducted a review of all sample claims utilizing policy and system edits/audits, and requested provider documentation of services in order to check the following six characteristics: completeness, recipient information, service characteristics, coding, charges/billing, and operational response.

- For its FFS study, Texas reviewed three sources of information for each claim: beneficiary telephone interviews, online claim reviews, and medical record reviews. The beneficiary interview was intended to validate their sample services were indeed received, whereas the online claim review considered the appropriateness of the claim in light of other claims made on the beneficiary's behalf around that time, and the medical record review ensured that paid services were documented, coded, and billed correctly with the use of several reference materials to ensure inter-analyst reliability. Texas carried over the three review methodologies from the FFS study to the VDP Pilot, replacing a review of the prescribing physician's records with a review of the pharmacy's records.

- Washington contacted service recipients to verify service delivery and to compare service delivery claimed with service delivery received. Claims were also reviewed for accuracy by checking medical documentation for type and level of service, medical necessity, and appropriateness of service. Claim payments were reviewed for conformance with policy, and for conformance with published rates and schedules.

- The Wyoming study reviewed claims paid for Medicaid services using a medical record review, a telephonic interview with the recipient, and a system accuracy review with emphasis placed on checking both pre and post-payment procedures and compliance in addition to a check for correct keying, disposition of edits and audits, and whether the system and procedures in place were pointing to the correct payment rate or fee for the service.

Many states found that they had to revise their planned approaches over the course of their projects. For example, Wyoming originally intended to interview all beneficiaries but ended up interviewing only a subset of the sample due to resource constraints. Texas planned to compare physician orders with filled prescription records in its Vendor Drug Study, but was only able to review the prescription records due to time delays.

In Year 2, the core model allows that record documentation requests to providers via mail are sufficient. Based on the Year 1 results and discussions with participants, letter requests are sufficient. States may, at their option, conduct scheduled or "surprise" visits to provider offices to collect medical record documentation. States may also conduct beneficiary interviews for all or a subset of their sample if they choose.

The Year 2 review and audit should consist, at a minimum, of two components: processing validation and medical review. Each line item should be reviewed to validate that it was

processed correctly, based on the information that is on the claim. Specific processing issues to address in the review include: beneficiary eligible for billed Medicaid services on the date they were provided, duplicate item (claim), non-covered service, service covered by MCO (i.e., beneficiary is enrolled in managed care organization that should have covered the service), third party liability, invalid pricing (including unbundling of bundled services), and logical edits (e.g., incompatibility between gender and procedure). The comprehensive medical review should include at a minimum review of the guidelines and policy related to the claim, review of medical record documentation, medical necessity review, and coding accuracy review. CMS will provide states with guidelines for conducting the medical review but will allow states to use standards that are compliant with local Medicaid policies.

C. Sources of Payment Errors

In conducting their PAM studies, the Year 1 states came up with different ways of categorizing errors, which are not necessarily comparable with one another. Nevertheless, a discussion of the sources of the errors found in each of the studies is necessary and may potentially yield insight into how the states may improve their payment accuracy rates and how the PAM studies themselves may be improved. In general, the largest source of error in all of the studies emerged as the result of documentation problems. This, however, can be divided into claims where there was no documentation supplied by providers and claims where the documentation submitted was considered insufficient to substantiate the claim.

Discrepancies in the type and dollar amount of errors found by states may be a reflection of the study methodologies as much as state-specific Medicaid payment differences:

- Of the 22 paid claims ($3,034) that Louisiana found to be in error, 8 claims had no documentation, 7 had insufficient documentation, 4 had dispensing problems, 2 had policy/coverage violations, and 1 had a coding error.

- Mississippi reviewed a total of 3,559 individual claim items and found a total of 535 errors ($25,964), of which 76 resulted from no documentation, 40 had insufficient documentation, 15 involved no medical record, 60 related to incorrect upcoding or unbundling, 14 resulted from clerical errors, 2 were false claims, 16 constituted potential fraud, 29 involved systems errors, and 283 failed to demonstrate medical necessity.

- Of the 300 claims that North Carolina re-reviewed from its SFY 2000-2001 study, it found a total of 14 claims with errors and $7,498 in overpayments, of which 2 were the result of insufficient documentation, 2 had a payment change due to DRG coding, 9 nursing and psychiatric home claims documented a change in the level of care, and 1 claim could not be determined because of a lack of documentation submitted.

- Of the 275 claims that North Carolina reviewed for its SFY 2001-2002 study, it found a total of 13 in error with $9,567 in overpayments of which 6 claims had insufficient documentation, 2 claims involved a provider's DRG coding, 4 nursing and psychiatric home claims had a level of care change, and the medical necessity of 1 claim was denied based on the provider's medical records.

- Of the 15 ($5,700) errors that North Dakota found in the records it reviewed 3 ($66) were coding errors, 8 ($4,791) had insufficient documentation, 2 ($769) involved billing errors, and 2 ($74) involved processing errors. It is worth noting that there were a total of 14 claims ($20,989) for which it did not receive any documentation at all.

- In Texas' FFS Study, it found a total of 462 services ($23,304) with errors, of which 2 ($12) were system payment errors, 20 ($0) were billing/clerical errors, 144 ($10,551) resulted from no medical record received, 90 ($4,227) resulted from a medical record that was received but included no documentation of the sample date or services, 129 ($5,767) resulted from a medical records that was received but included no documentation of the sample service on the sample date, 22 ($170) resulted from improper procedure in unbundling services (dental or laboratory tests), 26 ($979) were determined to be the result of improper upcoding, and 29 ($1,598) constituted potential fraud or abuse (services not medically necessary, etc.).

  In the VDP Pilot, Texas found a total of 351 services ($11,429) with errors of which 1 ($0) was attributed to a billing/processing error, 95 ($0) resulted from no prescription records having been received from the pharmacy, 184 ($8,402) in which the records received lacked the original copy of the sample prescription for the sample date requested, 1 ($62) involved a narcotic prescription received that was not compliant with the federal documentation requirements, 67 ($2,965) in which the prescription did not meet the general Texas Board of Pharmacy documentation requirements, and 3 ($0) in which there was no Medicaid medical record that corresponded to the prescription claim (not counted as an overpayment).

- Of the 500 claims worth $277,779 reviewed by Washington, it found a total of $4,367 in inaccurate payments of which $1,892 was the result of a claim where documentation didn't support the specific service, $243 where the documentation did not support the date of service, $296 where the reimbursement exceeded the appropriate rate or schedule, $32 for which documentation was non-existent, illegible, or indeterminate, $823 in which the documentation was not related to the claim, $556 resulting from an inappropriate level of service or modification on the claim, $302 caused by insufficient documentation, $18 in which the client was not Medicaid-eligible, $108 in which the prescription was not filled as written, and $97 where the service provided was not covered by Medicaid.[3]

- Wyoming found a total of 33 errors ($2,599) in its study but failed to offer a breakdown of where they had emerged beyond simply listing the types of errors that it coded such as service provided beyond program limitations or not provided at all, date discrepancy, absent or incomplete documentation, improper upcoding, and incorrect procedure code.

To help ensure the comparability of findings across states in the Year 2 pilot, a standardized list of error codes was developed based on the approaches used by the Year 1 states. While this list may evolve over the course of the Year 2 pilot, at a minimum states are expected to categorize processing errors into the following categories: duplicate item, non-covered beneficiary or service, MCO coverage of service, third party liability, pricing error, logical edit, or other. Medical review errors should be categorized as unsupported due to non-response to

documentation request, unsupported due to insufficient response to documentation request, medically unnecessary service, coding error, or other.

D. Payment Accuracy Rates by Service Type

Another integral part of each state's discussion of payment errors was a description of how the errors were distributed across different service types. This not only offered a basis for comparing accuracy rates across different types of services and providers, but also yielded the opportunity to explore whether there were areas with consistently low accuracy rates. These services could subsequently be targeted for improvement in the interest of improving the accuracy rates as a whole. It should be noted that, within any particular state, sample sizes by type of services were typically insufficient to obtain precise estimates of a service-specific error rate. However, the rates are suggestive and if all states adopt consistent service-type definitions (as in the Year 2 core model), the results, pooled systematically across states, will provide precise service-specific rates.

Similar to their division of Medicaid claims errors by error type, states had widely divergent ways of classifying claims by service type. Not all states used the same broad service type headings, and even those that used the same general categories may have assigned different specific services to the broader headings. Along these lines, some states divided their samples into a number of highly-descriptive strata that contained relatively few claims and could not be considered representative of this service type as a whole because of their exceedingly small sample size. For these reasons, the strata used by each state in the Year 1 pilot are not directly comparable but nonetheless offer a general idea of the differences in accuracy rates across different health care delivery areas:

- Mississippi found that its overall payment accuracy rate was 85.5% and the service-specific accuracy rates were as follows: 86.2% for dental services, 88.9% for pharmacy services, 59.8% for mental health services, 70.7% for ambulance services, 91.8% for non-emergency transportation services, 95.6% for physician services, 98.4% for inpatient hospital services, 88.4% for outpatient hospital services, 78.8% for Lab/X-ray services, 79.9% for DME services, 100.0% for home health services, 92.8% for vision services, 100.0% for long-term care services, 93.5% for clinic services, and 94.1% for other health care services.

- North Dakota calculated two separate payment accuracy rates for claims in which it found a clear error (98.71%) and for claims that were undocumented (95.24%). If one were to consider both these types of claims to be in error, its overall payment accuracy rate was 93.90% and this broke down across the 9 service areas that it established as follows: 94.09% for dental services, 91.00% for prescription drugs, 100.00% for inpatient professional services, 70.3% for outpatient professional services, 93.3% for nursing facility services, 100.00% in rural health clinic/FQHC payments, 100.00% in inpatient hospital services, 100.00% for outpatient hospital services, and 100.00% for crossover claims.

- In Texas' FFS Study, it calculated an overall payment accuracy rate of 86.50% which was divided as follows among the 8 service areas for which it randomly selected 100

beneficiary days: 94.73% for ancillary-outpatient services, 72.67% for dental services, 98.82% for home health services, 91.34% for hospital services, 94.08% for mental health services, 37.63% for other health care services, 41.95% for physician services, and 83.50% for supply/DME services. Additionally, Texas' VDP Pilot assessed the payment accuracy rate for prescription drug claims calculated to be 75.65%.

- Washington's study calculated an overall payment accuracy rate of 98.4%, and assigned accuracy rates to its 7 strata as follows: 95.1% for outpatient hospital services, 91.0% for physician services, 98.6% for pharmacy claims, 95.5% for dental services, 97.5% for payments to medical vendors, and 100.0% for long-term care services.[4]

- Wyoming assessed an overall payment accuracy rate of 97.7%, which broke down across its 8 service categories as follows: 100.0% for long term care services, 90.5% for transportation services, 100.0% for inpatient services, 100.0% for dental services, 90.9% for HCBS services, 96.5% for outpatient services, 96.1% for provider and ancillary claims billed on the HCFA 1500 form, and 99.2% for pharmacy claims. [5]

- Louisiana and North Carolina failed to offer a break down of payment accuracy rates by service type, with Louisiana stratifying its sample by paid and denied claims and North Carolina stratifying its sample on the basis of the claim values.

In the Year 2 study, all states will stratify claims into six strata: inpatient hospital services; long term care services, independent practitioners and clinics, prescription drugs, home and community-based services, and other services and supplies. States will also be given detailed crosswalks indicated which specific Medicaid services (based on the Medicaid Statistical Information System categories) map into which strata. This will help ensure consistency and allow comparisons among states. It may also allow CMS to develop statistically valid national estimates of payment accuracy for each stratum by combining the findings of individual pilot states.

V. ISSUES AND RESOLUTIONS

In the final reports submitted to date, participating states identified a number of areas where they ran into difficulties. Generally, these issues fell within six primary areas: a lack of time and/or resources, claim retrieval and documentation, record review and error coding, prescription drugs, crossover claims, and quality control. Most states described the approaches they used to overcome these obstacles. These "lessons learned" will be used to help states participating in the pilot project in Year 2 and beyond to avoid or mitigate these challenges.[6]

A. Time and Resource Issues

Nearly all of the states reported unexpected delays with respect to certain parts of the project that forced them to modify their originally stated timelines. Specifically, Louisiana addressed this type of difficulty as follows:

> Beneficiary interview were not conducted due to the length of time from date of service of the sampled claims and the medical record review of the nurses. Interview forms were designed but not used.[7]

Along these lines, Mississippi was also forced to modify its study methodology in response to significant delays and a lack of sufficient resources:

> The initial fieldwork process of personal, unannounced visits to obtain records was too time consuming and logistically difficult. The process was modified accordingly.[8]

Furthermore, the state's own unique political circumstances caused delays in performing certain functions integral to the study:

> Because of a change in the fiscal agent and additional legislative/political review requests on DOM staff, coupled with an already overburdened and undermanned staff, the PAM project experienced delays in responding to requests. The change of fiscal agent also created a delay in pulling the sample.[9]

The major difficulties identified by the state of Texas in completing its Medicaid FFS study pertained to the unexpected delays it encountered, the first of which dealt with its selection of a sample:

> Delays occurred during each phase of this study, some expected, others not. The expected delay was with the sample selection. The project team recognized that the sample must represent the universe even though three strata were purposefully over-sampled: Home Health, Mental Health, and Dental Services. These three strata were over-sampled with a larger quantity of services than normally found in the universe of claims because historically these providers have more problems with overpayments. The Mental Health and Home Health strata samples had to be re-selected to acquire a statistically representative sample for a 95 percent confidence level prior to starting the study reviews.[10]

A second source of delay concerned the logistical issues involved in retrieving and reviewing medical records:

> The project team encountered unexpected delays in getting the provider medical record request letters mailed promptly, especially the second and third requests. This process will be monitored more clearly in the Year 2 PAM project. Additionally, finding qualified professional utilization review nurses with Texas Medicaid experience to contract for the medical record reviews provide to be an issue. The Comptroller's office contracted with only one consultant which required the Comptroller's project manager to perform review with the contractor to meet the project deadlines.[11]

Wyoming was one of the few states that did not mention running short of time but did bring up a resource issue that it had not accounted for in its budgetary figures:

> In our original proposal we failed, to allow for the cost of postage and for the reimbursement for providers submitting medical records. This cost amounted to $1530 and was offset by reduced costs in other areas.[12]

B. Claim Retrieval and Documentation Issues

Like the time and resource shortages that most states encountered, nearly all states reported difficulties in retrieving medical records for Medicaid claims (these problems also caused delays and timeline adjustment). For example, Louisiana was forced to increase the amount of time that it allotted for medical record retrieval as follows:

> Considerable effort was expended by project staff in obtaining the medical records necessary to conduct the above described review. The medical record review period consequently covered early February through early July 2002.[13]

Mississippi identified the following as problems that related to its ability to obtain the relevant medical records, communicate with Medicaid beneficiaries, and track these documents:

> Experienced difficulty with getting MMIS records on the sample unit of beneficiaries.[14]

> Inability to contact/locate many of the sample population for the beneficiary interview and verification of service. Resolution included follow-up by letter asking beneficiaries to contact DOM which increased the contact by a very small number.[15]

> Difficulty with maintaining a timely "tracking" of the sample claims through the process.[16]

The majority of problems that North Dakota experienced with its PAM project related to its requests for medical records and a lack of response on the part of providers to these requests. Correspondingly, it was the only state to have calculated payment accuracy rates separately for claims that had and had not been documented. Regarding its requests for medical records that were sent to providers, it encountered the following two difficulties:

> When we first selected the data, we selected all the information relating to the claim. That provided too much information, making the sampling process inefficient. The original information was passed to our Research and Statistical Unit in its entirety. The information sent to the statistical department had too much information attached and slowed the ability to compile a sample size. This was then replaced by information, which only included the paid amount, number of line items per claim, and ICN number. From this, a sample was generated. The sample was then used to obtain the detailed information for the claims selected.[17]
>
> The request for documentation letter we utilized during the first request for information needs to be more specific as to what documentation was required from each provider in the stratum. We initially developed a universal letter to be sent out to all provider types. We found that this caused a lot of confusion and resulted in providers submitting information that did not suffice for the review or more information than was needed. For future reviews, it is imperative that the requests for information are provider-specific and specifically identify what documentation is required to be submitted. [18]

Concerning the responses from providers to these requests, there were two primary issues that were addressed:

> Providers are becoming aware of the Health Insurance Portability and Accountability Act, and on occasion refused to provide the information without a release from the recipient. We addressed this issue in Year 1 by talking directly with the providers. For the second year, a blanket statement on all requests should be included in the letter addressing this issue.[19]
>
> Providers did not always respond to requests for documentation. We had a number of providers that did not respond to our request and follow-up requests for information. On-site visits and adjusting these claims and recouping the money would help address this issue.[20]

In its discussion of issues encountered, Wyoming brought up the question of how to separate Medicaid claims and line items in the interest of maintaining its sample size and came up with the following resolution to this difficulty:

> Each "claim" to be reviewed originally was to represent the entire claim, however, during the course of the review we determined to have each claim represent a line item. Therefore the original proposal to review 600 claims in essence amounted to 875 claims which were reviewed. For purposes of the project, we used the first line item per claim and calculated rates based on 600 claims.[21]

Like other states, Wyoming also had difficulties obtaining current contact information for its Medicaid beneficiaries and this impeded its ability to conduct telephone interviews with them regarding the Medicaid services that were delivered:

> For the telephonic interviews it was extremely difficult to locate many of the recipients. The Wyoming MMIS does not capture phone numbers. A significant amount of time

was spent trying to obtain these numbers. However, we felt that it was important to attempt a certain number of telephonic interviews in order to determine if there was a benefit to performing that type of review. We were finally able to gain access to the eligibility system where we did obtain a sufficient amount of telephone numbers. Even with the number it was still often difficult to get in touch with the appropriate parties who could answer the questions. And in many cases, so much time had elapsed since the services were provided, the recipients and/or their guardians could not remember specifics of the appointments or services.[22]

C. Record Review and Error Coding Issues

Once medical records had been retrieved, some states encountered additional difficulties that hindered them in reviewing the records and appropriately coding errors. Mississippi noted a general review problem:

> Lack of formal cross training in investigating the full range of claim types covered in the study.[23]

North Dakota had more extensive and technical difficulties. The first emerged very early in the sampling process, when its proposed methodology was found to be incompatible with the Medicaid data that were maintained:

> The strata categories originally proposed in our grant application were not in the same format that we house our data in our dataprobe warehouse. Therefore, we changed our strata to match the method of storing our data in our data warehouse for the case of pulling the data and the ability to relate the selected claims with the detailed information maintained in the program. We pulled the sample based on the following strata: dental, drug, inpatient/outpatient UB-92, inpatient/outpatient HCFA 1500, Nursing, RHC/FQHC, and crossovers.[24]

The second difficulty stemmed from a lack of clarity with respect to the error codes that it had established:

> During the review, we found an error that could be classified into separate categories. During our testing, we ran across this problem and categorized it into the first error that happened, during the billing and adjudication process. We would recommend that a clear definition of each category is developed and a hierarchy established for all future testing scenarios.[25]

Wyoming encountered similar problems regarding the coding of errors and made the following two points in its discussion:

> There were instances where we had originally coded a claim as an error, but later found no basis in policy or rule to do so and then removed them from the error category. It appears that over time, some policies have not been put into place. These issues were referred to the appropriate policy areas for review and follow-up.[26]

> The number of potential error codes/descriptions that we used was too large. We will follow the guidelines set forth for the project year 2 and perhaps have sub-groups of errors within each of the established error codes for the next review.[27]

D. Prescription Drug Review Issues

Prescription drugs and pharmacy claims were a frequent source of problems for the pilot states (some of these problems overlap with the other categories established here).

Mississippi stated the following in its final report regarding its review of pharmacy claims:

> Because of the sheer number of pharmacy claims involved, the pharmacy unit was overwhelmed with their review responsibilities. Program Integrity staff assisted with Pharmacy reviews.[28]

Texas identified the following during its VDP Pilot Study, which considered only pharmaceutical claims:

> The Medicaid Vendor Drug Program Pilot study encountered some significant delays that required the project team to alter the format and objective of the study. Since this was an exploratory study to develop appropriate review methods, the project team considered these delays as part of the development process. The most significant delay occurred early in the study with acquiring the sample from the Vendor Drug Program's claim history data. The data stored in the VDP database is not an exact match nor does it contain all the Medicaid demographic information on the beneficiary or provider required for the study reviews. These deficiencies required the National Heritage Insurance Company (NHIC) system staff to write special programs to cross reference the VDP data with the Medicaid claim history file data to acquire the information necessary to perform the study reviews, causing a four month delay. A designated sampling program developed specifically for this study will resolve this delay for the next study.[29]
>
> The sample selection delay caused another issue with a proposed review for the VDP pilot study. The project team had proposed to request medical records from the prescribing providers to review the prescription orders and to establish medical necessity of the drugs prescribed for the beneficiary's diagnoses. The project team needed to receive the pharmacy's original prescription copies to confirm or in some cases identify the prescribing provider before a medical record could be requested. In most cases, the related Medicaid medical claims were not billed by the prescribing provider and processed until late in the study due to the difference in claim filing between the two programs. The majority of prescription claims are paid by 'point of sale' or at the same time the prescription is filled, while the Medicaid program gives the providers 90 days to submit their claims for payment. Consequently, the project team compared the prescription information with the Medicaid medical claims submitted by the prescribing providers to first determine if the drugs prescribed were related to a medical claim by date and then if the drugs were appropriate for the diagnoses treated. This research found that all but two of the sample drugs did have a related corresponding Medicaid medical claim.[30]

Wyoming also found the need to modify record requests and other printed materials when following up on pharmacy claims upon realizing that its pharmacies were not at all familiar with this sort of inquiry:

> Following the first mailing of requests for records, it became apparent that our pharmacies were not regularly reviewed and in fact not at all familiar with the letter we had sent. Therefore, we revised the letter for pharmacy requests only and re-sent it. This second letter was more specifically designed for pharmacies and received a much better response as was noted by the number of records received. Some pharmacists simply said that they didn't have time to "dig up" the records. These were recorded as an error and referred to the Program Integrity Unit for follow-up.[31]

E. Crossover Claim Issues

Crossover claims made under both Medicaid and Medicare constituted only a very small fraction of the total claims that were reviewed by each state. Correspondingly, Louisiana was the only state to have brought up any issues with respect to these:

> Medicare crossover claims that resulted in zero payment from the Medicaid program were not considered discrepant, even if there were post-payment review documentation errors that would have made the Medicaid payment an error if payment had been made. Six such claims identified in the claims detail report are not considered discrepant for the purpose of this study. Likewise, documentation deficiencies that did not impact payment were not considered discrepant.[32]

Crossover claims have been excluded from the Year 2 pilot design.

F. Quality Control

Finally, with respect to quality control for the project, all states noted that it was a daunting task to ensure the consistent and accurate review of hundreds of Medicaid claims. For this reason, a number of states established certain quality control procedures, such as re-review of a sub-sample of claims or re-evaluation of those claims that were coded as errors. A few states did not have a comprehensive quality control mechanism or used one that encountered difficulties that it could not easily address. For example, Wyoming noted that its lack of a formal quality control mechanism most likely hindered its ability to reach solid conclusions:

> We did not have a formal level of quality control for this project; however state staff did review each claim for appropriateness of error coding. For Year 2, we will have a defined level of quality control as part of the process.[33]

VI. BASIC PAM MODELS FOR YEAR 2

Incorporating best practices from these nine pilot projects, and groundbreaking earlier studies in Illinois, Texas, and Kansas, CMS and its consultant, The Lewin Group, have developed the core requirements for the CMS PAM Model.[34] This model has been designed to produce both state and national level estimates, providing CMS with both the uniformity and precision to report payment accuracy estimates across the Medicaid program at the national level, while maintaining sufficient flexibility to enable states to produce state-specific estimates. CMS is funding twelve states to participate in the second year of the Medicaid PAM Project, using grant money allocated from the Health Care Fraud and Abuse Control (HCFAC) program budget.

A. Payment Accuracy Measurement Model: Fee-for Service Claims

The basic fee-for-service payment accuracy measurement concept is:

1. Draw a random sample of claim lines from the universe of all fee-for-service Medicaid claims *paid* to providers for which there is Federal financial participation (exclude any non-service based payments such as disproportionate share payments, crossover claims, and aggregate cost settlement payments). The overall sample size will be drawn to obtain an estimate of the accuracy rate that is within ± 3 percentage points of the true population accuracy rate, with 95 percent confidence: sample sizes necessary to achieve this level of precision will typically be greater than 800 line items.

2. Subject that sample to review and audit to determine the validity of the payments made, consisting, at a minimum, of two components: processing validation and medical review. Each line item should be reviewed to validate that it was processed correctly, based on the information that is on the claim (e.g., beneficiary was eligible for Medicaid at the time the service was rendered, it is not a duplicate claim, non-covered service, etc). The comprehensive medical review should consist of a review of the guidelines and policy related to the claim, a review of medical record documentation, and a review of medical necessity and coding accuracy.

3. Compute an accuracy rate based on the sample, where the accuracy rate is defined as the ratio of the expected dollar value of payments paid accurately to the dollar value of total payments made.

B. Payment Accuracy Measurement Model: Capitated Managed Care

The basic managed care payment accuracy measurement concept is:

1. Draw a random sample of "claims," which for these purposes are defined as a capitation payment (or premium) made on behalf of an individual beneficiary for an individual month or partial month. Payments made by MCOs to individual providers for services are not included in the sample—only payments from the state to the MCOs. The sample of capitation claims should be drawn to obtain an estimate of the accuracy rate that is within 3 percentage points of the true population accuracy rate for capitation payments, with 95 percent confidence.

2. Subject that sample to review and audit to determine the validity of the payments made. First, reviewers must determine whether a capitation payment made by the state to a specific health plan on behalf of an individual beneficiary for a given month was, in fact, warranted under the rules of the state's managed care program. Reviewers should analyze the information available from eligibility and enrollment transaction records and make an independent determination of whether that enrollee was eligible to be enrolled, and was actually enrolled, in a specific health plan in the month in question. Second, in the event that a payment by the state to the health plan for that beneficiary is determined to be appropriate, reviewers must determine whether the proper amount of payment was actually made (e.g., proper rate cell used).

   If the state carves any services out of the capitated rate and pays them on a fee-for-service basis, the state must assess whether any fee-for-service claims were paid on behalf of beneficiaries during their period of enrollment in a capitated health plan, and whether such payments were accurate. Reviewers should obtain from the state's paid claims history all of the fee-for-service paid claims for services *incurred* on behalf of the sampled beneficiary during the enrollment month. Each fee-for-service claim for a service that is determined to have been incurred during the period in which a beneficiary was actually enrolled in a capitated managed care organization should be reviewed for appropriateness of fee-for-service payment in light of the terms of the contract with the specific managed care organization in which the beneficiary was properly enrolled. Where a determination regarding contractual coverage turns on clinical issues, that determination should be made based on a medical record review process.

3. Compute an accuracy rate based on the sample, where the accuracy rate is defined as the ratio of accurate payments made to total payments made.

VII. CONCLUSIONS

Prior to Year 1 of the PAM project, CMS and most states had no experience in trying to measure Medicaid payment accuracy on a systematic basis. The several states that had conducted studies in this area – notably Illinois, Kansas and Texas – had sampled and reviewed claims in very different ways, and no semblance of an 'industry standard' for Medicaid PAM existed. A number of states had some interest in pursuing the PAM project, but it is clear that the offer of 100 percent federal funding was critical in securing the participation of most of the Year 1 states. In Year 1, each state designed and conducted its PAM study however it thought best. This maximized the collective federal-state experience and learning opportunity, though it obviously minimized the ability to make comparisons across the states. As a result of work done last year, there is now a model PAM methodology which all twelve Year 2 participants have agreed to test. That testing will help not only in refining the methodology, but also in projecting the cost of Medicaid PAM should it be implemented on a national basis. The cost data for Year 1 is not indicative of what a national program would cost, as each state's project was unique, and few were as extensive as what the draft PAM model requires. Moreover, several states have not yet completed their Year 1 projects. This includes New York, which had the largest Year 1 budget, but was unable to launch its pilot until late summer, so to date has reported no Year 1 costs. Minnesota and several other states likewise received no-cost extensions from CMS, and are still incurring Year 1 costs. And most other Year 1 participants are still accumulating and reporting project costs incurred in FFY 2002.

The CMS PAM model that the Year 2 states agreed to test captures overpayments only. Since OMB and the Congress now define 'improper payments' as also including underpayments, CMS is modifying the Medicaid PAM methodology to do so. Year 2 states will be asked to use the modified methodology in their PAM studies this year. Doing so highlights the fact that state participation in the PAM project is strictly voluntary. Most states are facing severe budget shortfalls, and Medicaid directors report staffing and administrative budget cuts that make it increasingly difficult to perform their basic program management functions. They are not looking for additional work even where, as currently with CMS' PAM project, 100 percent federal reimbursement is offered. It will be a challenge to entice more states to test the refined Medicaid PAM methodology in FY 2004.

---

[1] Participating states report project costs as Medicaid Administration and receive the applicable federal match rate (typically 50 percent). What would normally be the state share of project costs will be paid from the grants CMS awarded the states in September 2001. Funding for these grants came from the FY 2001 Health Care Fraud and Abuse Control (HCFAC) Program. Approved budgets for the nine Year 1 states were as follows: Louisiana ($381,000), Minnesota ($347,281), Mississippi ($891,363), New York ($1,000,000), North Carolina ($440,592), North Dakota ($58,021), Texas ($158,220), Washington ($224,647), and Wyoming ($37,224).

[2] This improves the precision of the estimate if the variance of the accuracy rate across strata is proportional to the Medicaid payment share represented by the stratum.

[3] Washington did not break down the error coding by claim numbers, opting instead to portray them by line numbers and value. Furthermore, the figures given here do not include the 60 inpatient hospital claims that Washington was unable to analyze within the specified timeline.

[4] Payment accuracy rates for inpatient hospital services were not available due to delays experienced by Washington in analyzing those claims.
[5] "HCFA claims" refer to claims billed under the HCFA 1500 form for ancillary or physician services
[6] The issues and resolutions presented here do not include difficulties encountered by North Carolina and Washington, whose preliminary reports did not specifically address this topic.
[7] Louisiana Year 1 Medicaid PAM Final Report, p. 6
[8] Mississippi Year 1 Medicaid PAM Final Report, p. 6
[9] Mississippi Year 1 Medicaid PAM Final Report, p. 6
[10] Texas Year 1 Medicaid PAM Final Report, p. 12
[11] Texas Year 1 Medicaid PAM Final Report, p. 12
[12] Wyoming Year 1 Medicaid PAM Final Report, p. 6
[13] Louisiana Year 1 Medicaid PAM Final Report, p. 6
[14] Mississippi Year 1 Medicaid PAM Final Report, p. 6
[15] Mississippi Year 1 Medicaid PAM Final Report, p. 6
[16] Mississippi Year 1 Medicaid PAM Final Report, p. 6
[17] North Dakota Year 1 Medicaid PAM Final Report, p. 5
[18] North Dakota Year 1 Medicaid PAM Final Report, p. 5
[19] North Dakota Year 1 Medicaid PAM Final Report, p. 5
[20] North Dakota Year 1 Medicaid PAM Final Report, p. 5
[21] Wyoming Year 1 Medicaid PAM Final Report, p. 6
[22] Wyoming Year 1 Medicaid PAM Final Report, p. 6
[23] Mississippi Year 1 Medicaid PAM Final Report, p. 6
[24] North Dakota Year 1 Medicaid PAM Final Report, p. 5
[25] North Dakota Year 1 Medicaid PAM Final Report, p. 5
[26] Wyoming Year 1 Medicaid PAM Final Report, p. 6
[27] Wyoming Year 1 Medicaid PAM Final Report, p. 6
[28] Mississippi Year 1 Medicaid PAM Final Report, p. 5
[29] Texas Year 1 Medicaid PAM Final Report, p. 17.
[30] Texas Year 1 Medicaid PAM Final Report, p. 17
[31] Wyoming Year 1 Medicaid PAM Final Report, p. 6
[32] Louisiana Year 1 Medicaid PAM Final Report, p. 6
[33] Wyoming Year 1 Medicaid PAM Final Report, p. 6
[34] Illinois Division of Medical Programs, Office of the Inspector General. *Payment Accuracy Review of the Illinois Medical Assistance Program: A Blueprint for Continued Improvement.* 1998; Kansas Medical Policy Department, Social and Rehabilitation Services, *Payment Accuracy Review of the Kansas Medical Assistance Program*. April 2000; Texas Comptroller of Public Accounts, *Texas Health Care Claims Study, 2000*, January 2001.