OMB Approval No.
0938-0786

VALIDATING ENCOUNTER DATA

A protocol for use in Conducting Medicaid External Quality Review Activities

Department of Health and Human Services
Centers for Medicare & Medicaid Services

Final Protocol
Version 1.0

May 1, 2002

Form CMS-R-305

VALIDATING ENCOUNTER DATA

I. PURPOSE OF THE PROTOCOL

Encounter data (i.e., data on the distinct health care services provided to each Medicaid managed care enrollee) can be a useful source of information for States, as well as managed care organizations (MCOs) and Prepaid Health Plans (PIHPs). Encounter data can be used to assess and improve quality, as well as monitor program integrity and determine capitation payment rates. However, in order for encounter data to effectively serve these purposes, it must be valid (i.e., complete and accurate). At present, completeness and accuracy of encounter data vary across States, MCOs, and PIHPs. This protocol specifies processes for assessing the completeness and accuracy of encounter data submitted by MCOs and PIHPs to the State. It also can assist in the improvement of the processes associated with the collection and submission of encounter data to State Medicaid agencies.

II. ORIGIN OF THE PROTOCOL

This protocol was developed from documents in the public and private sectors, as well as interviews with personnel from three State Medicaid agencies (Alabama, Arizona, and Oregon) experienced in the collection of encounter data. The documents reviewed included: 1) A Guide for States to Assist in the Collection and Analysis of Managed Care Data - second edition (draft), 2) The MEDSTAT Group (MEDSTAT)'s Final Design Report for Verification of Encounter Data (part of the evaluation of the Medicare Choices Demonstration), and 3) the National Committee for Quality Assurance (NCQA)'s 1999 HEDIS7 publication: *Volume 5, HEDIS Compliance Audit™ Standards and Guidelines*.

Beginning in 1995, the Centers for Medicare and Medicaid Services (CMS) (formerly the Health Care Financing Administration (CMS)) and MEDSTAT began developing a series of tools to help State Medicaid agencies collect, validate and use encounter data for managed care program management and oversight. The tools and approaches developed for this contract were further refined and narrowed as part of the pseudo-claims (encounter data) validation project CMS commissioned for the Medicare Choices Demonstration. For that project, CMS specifically requested the development and application of a statistically reliable encounter data validation process using medical records as the reference information. MEDSTAT has used similar approaches for validating Medicaid managed care encounter data in a number of States, although these approaches have varied depending on the sophistication of the MCO/PIHP and State information systems, the amount of encounter data collected, and each State's approach to improving the quality of encounter data. However, in all of these States, MEDSTAT has used an approach composed of three core activities:

(1) Assessment of the MCO or PIHP's information system (IS)
(2) Analysis of MCO or PIHP electronic encounter data for accuracy and completeness, including analysis of data reasonableness
(3) Review of medical records for additional confirmation of findings

As part of the validation process, MEDSTAT has developed data collection instruments designed to evaluate, troubleshoot, and facilitate improvement of encounter data and the information systems from which encounter data are produced. These instruments are intended to be completed by the MCO/PIHP. The information obtained through these instruments is to be confirmed through face-to-face interviews of MCO/PIHP staff by the staff assessing the MCO/PIHP IS. The tools also are designed to be sensitive to the burden placed on an MCO/PIHP by information collection, while maintaining the integrity of the information collected.

The NCQA HEDIS Compliance Audit tool was designed for auditing encounter data when encounter data are used to calculate certain performance measures. It also includes an IS assessment analogous to the first of the three elements of the MEDSTAT validation process. The NCQA tool is designed to gather information about MCO/PIHP's IS capabilities across all payors, rather than concentrating specifically on Medicaid or Medicare. The MEDSTAT tools were developed for Medicaid and Medicare only, although MEDSTAT also has companion approaches designed for commercial payors.

All the validation processes reviewed address the collection of information about the MCO/PIHP's IS capabilities as a first step. All also include medical record review as a component of the validation method. Each document provides a method to calculate a statistically valid sample size for the medical record review. Interviews with State Medicaid agency personnel found that their protocols also include medical record review.

The NCQA and MEDSTAT tools also are similar in that they: 1) contain pre-onsite visit questionnaires to be completed by MCO/PIHP staff, and site visit interview forms to be completed by staff performing the IS assessment; 2) require MCOs/PIHPs to provide information on the level of specificity of the diagnosis and procedure coding systems used; 3) devote considerable attention to medical record review; and 4) explore the issue of provider contracts and physician compensation. Provider contracts and compensation are significant elements in understanding the flow of encounter data from the providers to the MCOs/PIHPs, which influences the timeliness and completeness of encounter data.

III. PROTOCOL OVERVIEW

This protocol is based almost entirely on the guide for States developed by MEDSTAT for validation of encounter data. The elements contained in MEDSTAT's document are consistent with the other documents reviewed.

This protocol also makes the following assumptions:

1. For the purposes of this protocol, an encounter refers to the electronic record of a service provided to an MCO/PIHP enrollee by both institutional and practitioner providers (regardless of how the provider was paid) when the service would traditionally be a billable service under fee-for-service (FFS) reimbursement systems. Encounter data provides substantially the same type of information that is found on a claim form (e.g., UB-92 or CMS 1500), but not necessarily in the same format.

2. The State will further specify an operational definition of an "encounter" and the types of encounters (e.g., physician, hospital, dental, vision, laboratory, etc.) for which encounter data are to be provided. The State will also specify the information (data fields) to be submitted for each type of encounter.

3. Encounter data can be considered "complete" when they can be used to describe the majority of services that have been provided to Medicaid beneficiaries who are enrollees of a MCO/PIHP. [1]

4. Development of accurate and complete encounter data is an iterative process. Because encounter data are an outgrowth of MCO/PIHP IS and data policies, it is often not possible for MCOs and PIHPs to overcome all limitations in their IS and data policies in one year. As a result, in the first year that a State requires the submission of encounter data from its MCOs and PIHPs, the data may be significantly incomplete and contain errors. Improving the completeness and accuracy will take place through continuous quality improvement (CQI) processes implemented year after year. Because of this, States will need to develop a "phased-in" approach for using standards for encounter data accuracy and completeness. "Phased-in" standards acknowledge the start-up issues affecting both MCO/PIHPs and State Medicaid information systems receiving the encounter data.

5. The State will establish standards for encounter data accuracy and completeness.

6. States will specify objective standards to which encounter data submitted by their Medicaid MCOs/PIHPs will be compared. These standards can be national, regional, or State standards, as discussed in ACTIVITY 3 of this protocol.

The protocol consists of five sequential activities:

(1) Review of State requirements for collection and submission of encounter data
(2) Review of each MCO/PIHP's capability to produce accurate and complete encounter data
(3) Analysis of MCO/PIHP electronic encounter data for accuracy and completeness
(4) Review of medical records, as appropriate, for additional confirmation of findings
(5) Submission of findings

IV. PROTOCOL ACTIVITIES

[1] A State may decide to use other sources of information, such as an immunization registry, to substantiate or complete the information on the provision of services to Medicaid beneficiaries when the encounter data format does not easily or accurately capture the required information.

ACTIVITY 1: Review State requirements for encounter data collection and submission.

Prior to performing encounter data validation, the External Quality Review Organization (EQRO)[2] needs to be familiar with the State's requirements for the collection, processing, and submission of encounter data by MCOs and PIHPs to the State. Some State requirements may be unique to a particular State program. States need to provide the EQRO with: 1) the State's requirements for collection and submission of encounter data by MCOs/PIHPs (these typically are found as specifications in the contracts between the State and the MCO/PIHP), 2) the data submission format specified by the State for MCO/PIHP use, 3) the State's data dictionary, 4) an explanation of the information flow from the MCO/PIHP to the State, 5) State standards for encounter data completeness and accuracy, 6) the time frames for data submission, 7) any historical problems experienced in this process, and 8) any other information relevant to encounter data validation.

The EQRO should also obtain from the State a listing of the types of encounter data to be validated. For each type of encounter data (e.g., office visit, inpatient, laboratory, et al.) to be validated, the State should specify the rates of missing, surplus, or erroneous encounters (as defined below) that it will find acceptable. The Acceptable Error Rates Specification Form below (or a similar form) can be used to summarize these specifications for each of the different types of encounter data.

The State also should specify acceptable rates of accuracy and completeness for each data field submitted for each encounter type. Attachment 1 contains MEDSTAT's recommendations for eventual accuracy and completeness standards for typical data fields. The EQRO will need to tailor this chart or generate a form or forms similar to Attachment 1 to identify accuracy and completeness standards specified by the State for all data fields the State requires for the different types of encounters. The standards should be more lenient in the early years of collecting encounter data, and more stringent as MCO/PIHP IS capabilities improve over time.

[2]It is recognized that a State may choose an organization other than an EQRO as defined in Federal regulation to perform encounter data validation. However, for convenience, in this protocol we use the term "external quality review organization" (EQRO) to refer to any organization conducting validation of encounter data.

Acceptable Error Rates Specification Form		
Type of Encounter	**Error Type**	**Acceptable Error Rate**
Office Visit (excludes dental and mental health / substance abuse visits)	Missing	< %
	Surplus	< %
	Erroneous	< %
Office Visit - mental health / substance abuse	Missing	< %
	Surplus	< %
	Erroneous	< %
Office Visit - dental	Missing	< %
	Surplus	< %
	Erroneous	< %
Inpatient admission - (excludes mental health / substance abuse visits)	Missing	< %
	Surplus	< %
	Erroneous	< %
Inpatient admission - mental health / substance abuse	Missing	< %
	Surplus	< %
	Erroneous	< %
Other types of encounters as specified by the State (e.g., laboratory, pharmacy, physical therapy).	Missing	< %
	Surplus	< %
	Erroneous	< %
The EQRO should add as many additional rows to this chart as needed to incorporate all types of encounters specified by the State.		

Definitions:

Missing - encounters that occurred but are not represented by an electronic record.
Surplus - encounters which are represented by an electronic record, but either did not occur or duplicated other records.
Erroneous - encounters that occurred and are represented by an electronic record, but contain incorrect data elements.
Acceptable Error Rate - the maximum percentage of missing, surplus, or erroneous records that the State is willing to consider acceptable.

ACTIVITY 2: Review each MCO/PIHP's capability to produce accurate and complete encounter data.

It is not feasible to review all encounters that beneficiaries have with MCO/PIHP providers to assess whether they are completely and accurately recorded. Therefore, efficiently assessing encounter data completeness and accuracy involves: 1) determining if the MCO/PIHP has structured its information system in a way that is *likely to* capture complete and accurate encounter data, and then, 2) examining more closely the data produced by the IS system to detect patterns that can indicate its completeness and accuracy. Activity 2 addresses the first of these two; the second is addressed in Activity 3.

Activity 2 attempts to answer the question: "To what degree is an MCO/PIHP's information system likely to produce complete and accurate information on all encounters between Medicaid enrollees and their providers (both institutions and practitioners)"? Reviewing the capability of each MCO/PIHP to do so is accomplished through two activities:

1. Reviewing a standardized assessment of each MCO/PIHP's IS capabilities, and

2. Interviewing personnel at each MCO/PIHP to augment information obtained through the standardized assessment.

Step 1: Review or conduct a standardized assessment of each MCO/PIHP's IS capabilities.

A standardized assessment (i.e., an assessment that does not vary by the individual performing the review or by the questions being asked) is necessary to promote reliable assessments of information systems. A standardized tool and approach for conducting such a review is found in Appendix Z (Information System Capabilities Assessment (ISCA) for Managed Care Organizations and Prepaid Health Plans).

An MCO/PIHP may already have undergone such an assessment of its IS. For example, assessment of IS is conducted when validating performance measures and performing accreditation reviews. The EQRO needs to determine if the MCO/PIHP whose encounter data are being validated has already undergone such a review, and if so, if the review findings are current. If a recent IS assessment has been conducted, the EQRO should receive a copy of the findings, review the results of the prior assessment, and seek more recent information where necessary. If the MCO/PIHP has not recently undergone an assessment, one will need to be conducted as part of encounter data validation consistent with the process described in Appendix Z.

Whether the EQRO reviews the results of an earlier IS assessment or conducts its own assessment, the content included in Appendix Z should be addressed. This content and the reasons for its significance include the following:

General Information

1. *Managed Care Model Type:* Encounter data completeness and accuracy are likely to be better in staff model than non-staff model MCOs/PIHPs.
2. *Year of incorporation:* Encounter data accuracy and completeness is likely to be better in more mature MCOs/PIHPs.
3. *Member enrollment:* A larger enrollment may indicate that the MCO/PIHP has more experience working with encounter data, but may also offer more opportunity for errors.

Information Systems Capabilities

1. *System descriptions:* This provides an indication of the MCO/PIHP's overall level of data management sophistication.

Data Acquisition Capabilities

1. *Forms used:* If the MCO/PIHP is not using standard claims forms (e.g., UB 92 or CMS 1500) or encounter forms, which are similar to the CMS 1500, or the UB 92, the forms used should be reviewed to ensure that they capture key data elements.
2. *Submission methods:* Processing paper forms is more prone to error than direct electronic data submission.
3. *Required data fields (data elements)*: Standard measures of plan performance typically require the availability of data on patient date of birth/age, sex, place of service, diagnoses, procedures, dates of service, revenue codes, and provider specialty.
4. *Number of diagnosis and procedure codes retained in data fields:* Data fields should allow a minimum of two diagnoses and two procedure codes to be retained.
5. *Coding schemes:* Knowledge of the coding schemes used by the MCO/PIHP is necessary to verify the accuracy of their use.

Claims/Encounter Processing

1. *Processing issues:* Points in the process where errors are particularly likely to occur should be identified
2. *Edit checks:* MCOs/PIHPs should have an established, standard set of edits, which verify field content and consistency

Enrollment Issues

1. *Type and frequency of updates:* Infrequent or inaccurate updates will result in invalid encounter data
2. *Use of unique identifiers:* Without reliable identification of enrollees and providers, encounter data validity is not possible

Vendor/Contractor Data

1. *Data submission policies:* Consistently applied policies that leave no room for variations in interpretation increase the accuracy of the data submitted
2. *Contract requirements:* Data are more likely to be complete when they are required as a condition of payment than when they are optional

Provider Contracting Arrangements

1. *Compensation arrangements:* Salaried providers will submit data on a timely basis if data submission is a parameter in their contract with the MCO/PIHP. Fee-for-service (FFS) providers have the greatest incentive to submit accurate and complete data, since their payment depends on it.
2. *Contract requirements:* Data are more likely to be complete when they are required for payment

After reviewing each MCO/PIHP's IS Capabilities Assessment, the EQRO staff will record their analytic findings on a standard form such as the Information Systems Capabilities Assessment for Managed Care Organizations and Prepaid Health Plans - *Reviewer Worksheet and Interview Guide* (Reviewer Worksheet and Interview Guide) found in Appendix Z. A form such as the Reviewer Worksheet and Interview Guide serves to document the findings of the EQRO staff when reviewing the IS Capabilities Assessment for each MCO, and to identify those issues to be addressed in Step 2, the follow-up interview with MCO/PIHP personnel.

Step 2: Interview personnel at each MCO/PIHP to augment information obtained through the standardized IS Capabilities Assessment.

Whether the EQRO is reviewing a previous assessment of MCO/PIHP IS capabilities, or it has asked the MCO/PIHP to complete a new Information Systems Capabilities Assessment (see Appendix Z), the written descriptions of IS capabilities submitted by the MCO/PIHP must be reviewed and supplemented by conversations with the MCO/PIHP staff to clarify or gather more detailed information regarding the MCO/PIHP's IS capabilities. The EQRO staff will interview appropriate MCO/PIHP staff using a standard interview protocol such as the MCO/PIHP IS Capabilities Assessment - *Reviewer Worksheet and Interview Guide* found in Appendix Z. However, not all information submitted by the MCO/PIHP on its Information Systems Capabilities Assessment might need further clarification. In addition, not all questions in the Interview Guide may need to be discussed with MCO/PIHP staff with respect to encounter data validation. This is because assessment of information systems can be conducted for different purposes (e.g., validation of performance measures and determining compliance with MCO/PIHP structure and operational standards). Further, some questions in the Reviewer Worksheet and Interview Guide will need to be reworded for newly formed MCOs/PIHPs. However, the following areas, at a minimum, should be fully described either in the MCO/PIHP's written documentation of its IS or through subsequent follow-up discussions between the MCO/PIHP and the EQRO.

Information Systems: Data Processing and Procedures

1. Data Base Management System (DBMS) Type
2. Programming language
3. Updating the program to meet changes in State requirements

Claims/Encounter Processing

1. Overview of the processing of encounter data submissions
2. Completeness of the data submitted
3. Policies/procedures for audits and edits

Claims/Encounter System Demonstration

1. Processes for merges and/or transfer of data
2. Processes for encounter data handling, logging and processes for adjudication
3. Audits performed to assure the quality and accuracy of the information and the timeliness of processing
4. Maintenance and updating of provider data

Enrollment Data

1. Verification of claims/encounter data
2. Frequency of information updates
3. Management of enrollment/disenrollment information

Based on this review of the MCO's/PIHP's IS capabilities, the EQRO should note for each encounter type listed in the Acceptable Error Rates Specification Form (described previously) whether or not there are concerns about certain types of encounter data, and note it in the fourth column in the chart below. This should trigger further investigation.

Acceptable Error Rates Specifications and Identified Areas of Concern Form			
Encounter Type	**Error Type**	**Acceptable Error Rate**	**Area of Concern (Yes / No)**
Office Visit - (excludes dental and mental health / substance abuse visits)	Missing	< %	
	Surplus	< %	
	Erroneous	< %	
Office Visit - mental health / substance abuse	Missing	< %	
	Surplus	< %	
	Erroneous	< %	
Office Visit - dental	Missing	< %	
	Surplus	< %	
	Erroneous	< %	
Inpatient admission - (excludes mental health / substance abuse visits)	Missing	< %	
	Surplus	< %	
	Erroneous	< %	
Inpatient admission - mental health / substance abuse	Missing	< %	
	Surplus	< %	
	Erroneous	< %	
Other types of encounters as specified by the State (e.g., laboratory, pharmacy, physical therapy).	Missing	< %	
	Surplus	< %	
	Erroneous	< %	
The EQRO should add as many additional rows to this chart as needed to incorporate all types of encounters specified by the State.	Missing	< %	
	Surplus	< %	
	Erroneous	< %	

ACTIVITY 3: Analyze electronic encounter data for completeness and accuracy.

This activity represents the core of the process that the EQRO will use to test the validity (completeness and accuracy) of the encounter data. Once the steps in this activity have been completed, the EQRO and the State will have an excellent assessment of whether the data can be used for analysis. If the EQRO is unsure of the quality of the encounter data at the completion of Activity 3, then it should *not* proceed to the medical record review activity (Activity 4). Rather, it should either review the steps of this activity and identify areas where information did not satisfy the EQRO or it should seek additional assistance to determine why there is uncertainty about the quality of the encounter data. If the steps in Activity 3 are completed thoroughly and accurately, there should be little doubt about the quality and usefulness of the submitted encounter data.

In this activity, the EQRO undertakes analysis of each MCO/PIHP's encounter data through four steps. Information obtained from these four steps and the previously conducted IS Capabilities Assessment and the Structured Interview, should yield four classes of information available for each MCO/PIHP:

1. ***General magnitude of missing encounter data*****.** Evidence of whether the MCO/PIHP has been unable to submit any encounter data and reasons for failures, such as the inability to process the encounter data without edits.

2. ***Types of potentially missing encounter data*****.** MCOs/PIHPs, which have sub-capitated or sub-contractor relationships with providers often, experience difficulty in receiving information from those providers. Knowledge of the MCO/PIHP's contractual relationships with providers will help identify specific areas to investigate for missing services.

3. ***Overall data quality issues*****.** Identification of data quality problems such as inability to process or retain certain fields on the encounter data record. Some MCOs/PIHPs may not currently have room in their systems to maintain all the information, which is expected to be submitted.

4. ***MCO/PIHP data issues*****.** Problems with how the files are compiled and submitted to the States.

Step 1: Analyze information from the IS Capabilities Assessment and the follow-up structured interview with MCO/PIHP staff.

At the completion of Activity 2, the EQRO will have an excellent description of the MCO/PIHP's IS and should know what to expect when the MCO/PIHP's data files are investigated. The information from Activity 2 is incorporated into a plan for testing the quality of the data. This plan specifies the areas that will be tested in the data (the areas of investigation) and the expected results. Having such a plan ensures that parts of the data quality review are not overlooked. For example, it is expected that MCOs that pay a substantial portion of their primary care providers on

a capitated basis will have a lower encounter data submission rate than MCOs that pay their primary care providers predominantly on a FFS basis. As a part of a data quality test plan, provider groups would need to be identified by their type of payment and rates of outpatient visits per eligible beneficiary calculated. The rates of outpatient visits would then be compared to test the assumption that capitated providers have a lower rate of encounter submission than FFS providers do. If this assumption is not supported by the data, and the outpatient visit rates differ from the benchmark, then other questions need to be answered about the MCO/PIHP's processing of FFS claims. This is an example of how the information from Activity 2 is incorporated into Activity 3, Step 1. Other questions and issues can be addressed in the same way.

Using information provided by the State in Activity 1, and information obtained from the MCO/PIHP through the IS Capabilities Assessment and the Structured Interview, the EQRO staff will develop a data quality test plan that will:

- Adjust their own error detection programming specifications to reflect those used by the MCO/PIHP
- Adjust their own report specifications to match those used by the MCO/PIHP
- Create *ad hoc* data investigation specifications, based on the information gleaned from the IS Capabilities Assessment and the interview
- Create *ad hoc* report writing specifications
- Compile notes to assist in data interpretation.

Step 2: Inspect the MCO/PIHP's Encounter Data files

Step 2 and Step 3 of this Activity are closely inter-twined. To make the steps clearer, they have been broken into two parts because having two steps more accurately reflects the way a standard data quality review process would occur. When data are reviewed for accuracy and completeness, they are subjected to a macro and microanalysis. These steps are described separately to prevent the EQRO from rushing forward to generate a large number of reports and analyses before the basic integrity of the data have been verified. Step 2 represents the macro analysis. Step 3 represents the micro.

Step 2 describes a basic integrity check of the data files. It answers the questions: Are there data? Do they generally fit with expectations? Are they of sufficient basic quality to proceed with analyses that are more complex? In general, all of the analysis that is required in Step 2 should be highly automated and generated as a standard data review process. The analysis required in Step 2 can be separately performed on each of the different encounter data files (e.g., hospital, dental, ambulatory, etc.) for each of the data fields in those files, while the analysis described in Step 3 requires that encounter, eligibility, and provider data be linked together. The EQRO will obtain the encounter data to analyze either by accessing the State's information system or by receiving from the State an encounter data extract from the State's data system that replicates the data that the State has. In the latter case, the EQRO will access the data using its own analytical processes. The EQRO will inspect the files and perform the following activities:

1. Assure that the enrollment information that the State transfers to the MCO is accurately incorporated into the MCO information system and is being reported back to the State correctly. In many cases, MCO information systems do not use the State's Medicaid identifier as the means for tracking enrollees. When the encounter data are reported to the State, the Medicaid ID must be reattached to the data file. In this step, the EQRO will verify that the Medicaid IDs are being reported correctly. As an additional, but not required, step, the EQRO could compare the encounter data file to a State eligibility file and check for accuracy of the IDs, and other eligibility information (e.g., age, sex, and eligibility category). This step is optional since in the majority of situations, the encounter data file that the EQRO is reviewing has been edited in the State system - where eligibility checking is one of the core elements. This step should be considered in those situations where the encounter data have not been edited in the State system.[3] In addition, the EQRO will determine whether there are encounter data for the majority of beneficiaries, rather than a large volume of data but only a few IDs. The primary focus of this step is the verification of the correct eligibility numbers. In other parts of this activity, analysis will be done to ensure that the scope and volume of services are consistent with the eligibles.

2. Apply general edit and consistency checks, such as verifying that critical fields contain non-missing values, that values are consistent across fields (e.g., pregnancy and related diagnoses), and procedures are for individuals whose sex is coded as female.

3. Inspect the data fields for general validity (i.e., information for each critical field is within required ranges, and the volume of data is consistent with the MCO/PIHP's enrollment).

4. Capture more detailed validity information from the encounter data fields used for reporting purposes.

The analysis includes a review of each data element and a general review of the volume of data by type or place of service. This review concentrates on two areas: field validation and completeness:

1. *Field validation*

- Percent present: required data fields are present on the file and have information in that field
- Percent valid: data in the field are of the requested type (i.e., numeric fields have numbers, character fields have characters, etc.)

[3] In a few States, Medicaid programs have chosen to by-pass the MMIS and have MCOs submit data directly to an outside source. In these cases, the eligibility checking may not be as intense as that done within the MMIS. In those cases, the EQRO might choose to add this additional eligibility review.

- Percent valid values: In those fields, the values are the expected values (e.g., are there valid ICD-9 codes in the diagnosis field, not just random numbers?). This review requires comparing the specific field to sources of information showing valid values. [4]

Field validation will require the use of State standards. Examples of eventual State standards are found in Attachment 1.

2. Data completeness

- Distribution by service type: determine whether there are data distributed as expected in the large data types: institutional, provider, pharmacy, dental, etc.
- Across time: determine the data volumes by month. Are all data types seen in all months? Is the data volume consistent across the months?

Step 3: Generate and Review Analytic Reports

Using simple statistical procedures such as measures of central tendency, univariate descriptive statistics, and bivariate distributions, the EQRO will analyze the data to obtain a "data validity" overview of each MCO/PIHP's encounter data. This process will analyze and interpret data on: 1) submitted fields, 2) volume/consistency of encounter data, and 3) utilization rates.

Analyzing and Interpreting Data in Submitted Fields

There are three questions to be addressed in a field-specific review:

1. **Is there information in the field, and is that information of the type requested?** Each field will have a definition, which will include data type (e.g., alpha, numeric, mixed) and size. The fields must be checked to determine whether the information is of the correct type and size. For example, if ICD-9 diagnosis codes have been requested, the field should have 5 digits. If CPT-4 codes are requested, the field should have 4 digits. If the State's Medicaid beneficiary ID is requested, the field should contain the correct number of letters and digits.

2. **Are the values valid?** When compared to an external standard, are the values in the field valid for that standard? For instance, if ICD-9 diagnosis codes have been requested, are the values in the diagnosis field valid ICD-9 diagnosis codes? A field could have 5 digits in it but those digits might not represent valid codes.

[4] This is a place where the information from the IS Assessment becomes critical. One of the things that will be asked is which version of the ICD-9 codes is used by the MCO. Often the State and the MCO maintain different versions of these files resulting in values being considered invalid when they really are valid.

Findings for questions 1 and 2 could be recorded on a standard form such as that below:

Sample Form for Recording Evaluation of Submitted Fields								
Required Field	**Information present**		**Correct type of information**		**Correct size of information**		**Presence of valid value ?**	
	#	**%**	**#**	**%**	**#**	**%**	**#**	**%**
Enrollee ID								
Plan ID								
Provider ID								
Principal Diagnosis								
Procedure Code								
Date of Service								
Units of Service								
Others (continue adding fields as appropriate). . .								

3. **Are the values reasonable?** A frequency distribution of the values needs to be developed and then compared to an external standard to determine whether the values make sense for the submitted population. For instance, if one of the required fields is "place-of-service," there should be a reasonable distribution between inpatient hospital, outpatient hospital, emergency room, and physician office. In this data review, the values in the fields could have passed the test in steps 1 and 2, but failed at this stage. A plan could submit data with a single valid value in a field; these data would pass steps 1 and 2 but fail step 3.

Analyzing and Interpreting Volume / Consistency of Encounter Data

This type of evaluation provides basic statistics on the encounter data. It describes, among other things, the number of Medicaid enrollees, the number of encounters, and counts and totals for various demographic subgroups, diagnoses, and types of services. The EQRO should run frequency distributions on specific fields as well as on the variables created explicitly for data validation reporting purposes. The EQRO may also run distributions on subsets of variables and observations where the result indicates potential data validity concerns. For instance, a subset of rates of outpatient services by provider zip code might highlight missing zip codes in the edit file,

which results in the rejection of all encounters with that zip code. This initially looks like an overall low rate of services but by looking at a subset of a specific field and checking for reasonableness, a different problem is detected. The EQRO also should generate univariate statistics (e.g., means, medians, and modes) as appropriate on continuous and discrete data fields. The output produced for these reports should be checked for reasonableness, and to detect specific problems such as entire categories of data missing from the regular data submissions.

The EQRO should also analyze encounter data for other volume/consistency dimensions. These dimensions can include time, provider type, type of service, and demographic groupings, but States may have additional dimensions (aid category) which also need to be included. This information allows the EQRO to look for trends such as the following:

- **Time** - This analysis would examine encounter data both by service date and by processing date to check consistency. MCOs/PIHPs often have problems processing encounter data and in many cases, these claims are processed sporadically. When such a situation is present, it can often be an indication of other problems within the MCO/ PIHP's information system. After establishing the length of time between service dates and processing dates by MCO/PIHP, the EQRO could compare these with existing benchmarks for data submission and processing.

- **Provider** - Encounter data validation can verify the presence of encounter data for all provider types and determine if there are significant fluctuations in patient visits per time period. In addition, information collected during the capabilities assessment will be used to identify missing encounter data for specific provider types. The distribution of encounter data will be compared by provider type with the benchmark information described above.

- **Service Type** - The EQRO should verify whether ancillary services (e.g., labs, x-rays, therapy, etc.) are evenly represented as visits. The clinical connection between the use of services is being evaluated. If a Medicaid beneficiary were receiving x-ray or lab tests, one would expect to see an office visit (or perhaps a hospital admission) in the same time frame. Other areas where reasonableness between the encounter data should be tested include: 1) relationship of outpatient visits to number of prescriptions, 2) relationship of primary to specialty care visits, and 3) outpatient services associated with inpatient admissions. Examination of the data in this way will reveal whether there are missing encounters.

- **Age- and Sex-Appropriate Diagnoses and Services** - The EQRO will determine if the diagnoses and services reflect expected care by age and sex. As an example, one would expect sex-specific diagnoses (such as endometriosis or undescended testes) and procedures (such as deliveries or hysterectomies) would have the patient's sex coded correctly. Conversely, one would expect that men and women would not have procedures coded that cannot be performed on a person of that sex.

As part of the review, the EQRO will find it helpful to display the data quality findings graphically. It is nearly always true that in these situations, "a picture is worth a thousand words." These graphs will be useful internally for identifying issues, and externally for conveying the results of the data quality review. An example of the type of chart that might be generated is a frequency distribution of the number of encounters per Medicaid enrollee. This chart shows how many enrollees had zero, one, two, three, etc. encounters. This chart can also be replicated by different age, sex, and race groupings, to see whether there are differences along these dimensions. These charts should be generated for all MCOs/PIHPs in the aggregate, and for each MCO separately so that issues with a specific plan can be identified.

Analyzing and Interpreting Utilization Data

The EQRO also should routinely compile and review, on a periodic (e.g., monthly, bi-monthly, quarterly or other periodicity as directed by the State) basis, statistics displaying information on utilization rates overall and by specific diagnosis, procedure, service and provider types when appropriate. These reports initially should be generated both for each MCO/PIHP and on the entire encounter data set for all MCOs/PIHPs together to account for problems associated with small numbers of encounters for certain MCOs/PIHPs. During the program start-up phase, many MCOs/PIHPs may have very few encounters for some diagnoses and services, which would make their rates statistically imprecise.

One method for estimating the completeness of the encounter data for each MCO/PIHP is to benchmark the utilization rates for a given MCO/PIHP against utilization statistics from other sources. These benchmarks would be incorporated into the MCO/PIHP-specific utilization reports. Utilization rates could be broken-out by patient demographics, diagnosis, type of service, and type of provider.

Step 4: Compare findings to State-identified standards

The EQRO will next compare the encounter data submitted by each MCO/PIHP to standards and benchmarks that are identified by the State. These standards can be obtained from a number of different sources, including: aggregate encounter data from all Medicaid MCOs/PIHPs in that State or other States which are considered to be comparable, historical FFS Medicaid data in that State, or data from a State's Primary Care Case Management (PCCM) program. The State may also look to commercial managed care plans, national standards, or other benchmarks. The State will need to identify such standards and document them using a form such as that found in Attachment 2: Table of Benchmark Utilization Rates. States will also need to specify acceptable variation from these standards. Both the standards and acceptable variations from the standards can be made more stringent over time.

For instance, when comparing encounter data utilization rates to a Medicaid FFS benchmark, one might expect to see a drop in emergency room utilization per member month under managed care. However, large, downward swings in other types of utilization (e.g., >30 percent drops in

ambulatory care) may indicate incomplete encounter data rather than a change in provider practice patterns. The EQRO should test their assumptions about changes in utilization under Medicaid managed care with the MCOs/PIHPs and with the State, as a further test of the completeness of the encounter data.

The results of ACTIVITY 3 are used to form the basis for a long-term monitoring strategy for assessing the quality of the encounter data. As the data improve over time due to monitoring and validation, the EQRO will be able to design targeted validation strategies by using analytic testing on the encounter data files to identify problem areas requiring medical record review validation, thus allowing the conservation of resources by avoiding unfocused "broadside" medical record review.

ACTIVITY 4: Review of medical records for confirmation of findings of analysis of encounter data

Medical record review can provide additional verification of the information obtained from the preceding analysis of electronic encounter data for accuracy and completeness (Activity 3). However, medical record validation is a complex and resource-intensive process. While it can easily be used to validate specific areas of concern, it is not an efficient method for performing a more general validation of encounter data. As stated at the beginning of Activity 3, if the EQRO is unsure of the quality of the encounter data at the completion of Activity 3, then it should *not* proceed to medical record review. Rather, it should either review the steps in Activity 3 and identify areas where information did not satisfy the EQRO or it should seek additional assistance to determine why there is uncertainty about the quality of the encounter data. If the State is in the initial phase of collecting encounter data from its MCOs/PIHPs, there may be a time lag of as much as three years before the encounter data are of sufficient completeness and accuracy to warrant the investment in medical record review

Further, medical record review should not be used to validate information that is collected at another source and considered more accurate. For example, information collected as part of eligibility determination is considered the primary source for demographic information such as patient age, sex, and race. This eligibility data should be used as the source of demographic information when validating encounter files.

In this protocol, the following assumptions are made with respect to medical record review:

- Medical record review for encounter data validation is being performed independently of medical record review for evaluation of performance measures or other purposes, however these reviews could be coordinated in an effort to reduce the burden on the MCOs/PIHPs.

- When medical record review should begin, and how frequently it should be performed, will be decided by the State.

- One medical record review is determined by the State to be appropriate, the EQRO will draw a sample of medical records for validation on a regular and periodic basis specified by the State.

Validating encounter data using medical records must be approached as if it were a research question. There needs to be clear hypotheses, well defined populations, and stated error tolerances. A rigorous research design ensures that the results of this resource intensive effort will be meaningful and useful.

The approach to medical record validation depends on the questions/hypotheses to be addressed. Depending on the stated hypotheses, one would begin with a sample of encounters, enrollees, or both. The table below illustrates how two different questions are answered by using different sampling universes.

Examples of questions to be addressed	**sampling universe**	
	Encounters	Enrollees
Is the information found on the encounters accurate when compared to the medical record?	Use sample of encounters	
Are there electronic encounters for all the services that were provided to enrollees?		Use sample of enrollees

Another possible question may be "Are the encounters fully coded for all diagnoses?" Often when providers submit encounters, they only provide the primary diagnosis code, since that is usually sufficient for payment. Another question might be "Are the procedures fully coded?" Providers might record on the encounter form only those procedures that historically have affected reimbursement. In such cases, additional procedures performed can only be discovered by reviewing the medical record. In this type of medical record validation, a sample of *encounters* is selected, and the medical records for those encounters are reviewed.

When one wants to determine whether all encounters have been received for services that were delivered, then a sample of *enrollees* is selected, and their medical records are reviewed against encounters. In this case, the EQRO would sample MCO/PIHP enrollees rather than encounters for a specific time period. Services recorded in the medical record would be matched with those found on the encounters to assess the completeness and accuracy of encounter data. To reduce the number of medical records that must be located for each enrollee, the EQRO should consider limiting medical record review to a specific type of encounter, such as inpatient admissions or physician office visits.

Where medical record review is performed is a complicated question and each solution has its own strengths and weaknesses. In all cases, the medical record must be located, the reviewers

must be trained and experienced, the confidentiality of patient records must be maintained, and costs must be minimized. Balancing these four requirements is difficult and results in varied solutions that fit with the specifics of each particular situation. In many cases, reviewers decide to look at medical records away from the site of health care delivery and request that providers copy charts and send them to a central location. This approach places a large amount of the responsibility for locating the charts on the plans and providers, but also increases the number of charts that are located. Providers should be compensated for the direct costs of copying and submitting the files.

Step 1: Sampling for medical record review

When medical record review is to be undertaken, the EQRO may initially choose to conduct "trial" reviews of the encounter data using medical records previously obtained for other purposes, such as focused studies of clinical care topics. This trial review would provide an opportunity for the EQRO to acquire experience in using medical records specific to individual MCO/PIHPs for encounter data validation, as well as highlighting community practice patterns with regard to developing encounters from medical records, without adding unnecessary administrative burden to the MCOs/PIHPs.

Once medical record review is to be conducted on a routine basis, the size of the sample must be determined for each MCO/PIHP. The size of the sample of medical records that must be drawn in order to make statistically valid inferences regarding the validity of the encounter data depends on a number of factors, including:

- The minimum error rate the State wants to be able to detect
- The frequency with which the State wants to perform the review
- The subsets of encounter data the State intends to validate

Because of these and other factors(e.g. the size of an MCO/PIHP's enrollment and previous validation results), it may be statistically appropriate to determine different sample sizes for each MCO/PIHP. However, it may be operationally more efficient (and statistically valid) to specify the same sample size for all MCOs/PIHPs. The EQRO will need to determine the sample size for each MCO/PIHP either as directed by the State, or, if the State allows the sample size to be determined by the EQRO, in consultation with a qualified statistician. Whether one sample size or multiple sample sizes will be used will be determined by the State, in consultation with the EQRO.

Once the sample size is determined, the EQRO will select this number of enrollees from each MCO/PIHP for medical record review in order to calculate fault rates for each MCO/PIHP. The sampling must be performed using methodologically sound techniques that defend against sampling bias. The fault rate is the ratio of missing and erroneous records to the total number of encounters that took place during the time period being examined. A fault rate can be calculated for each encounter type.

It is anticipated that fault rates initially will be at least 30 percent. However, each MCO/PIHP's *targeted* fault rate should be below 5 percent ($f < 0.05$) for each time period examined, and if possible, demonstrate a decrease over time. Consequently, for each time period examined, the EQRO will test the hypothesis that there is no difference between the services recorded in the medical record and those found in the encounter data (i.e., the "null" hypothesis). The ability of this test to reject the null hypothesis (H_o) when it should be rejected will depend on the size of the encounter sample, the size of the total population of true encounters, and the MCO/PIHP's actual fault rate. The EQRO should set sample sizes sufficient to estimate the fault rate for each type of encounter within each MCO/PIHP, with equal precision for each time to be studied.

Step 2: Review medical records and record findings on a standardized worksheet.

Fields to be validated through medical record review should include a few socio-demographic fields (e.g. date of birth, sex) that may be needed to identify the correct beneficiary. Typically, when medical records are selected for review, the provider is given the patient's name, age, and sex, the provider's name, and the target dates of service. This information helps the provider find the correct medical record. When the comparison is done between the encounter data and the medical record, this information is included in the data elements that are reviewed to further ensure that the correct record was actually reviewed. The demographic information on the medical record is not considered the definitive source of demographic information. It is included here to support the medical record review process. Other fields on the encounter data should be assessed so that it is possible to create measures of access and quality of care. These data elements include date of service and the clinical codes (such as ICD-9-CM, HCPCS, and CPT codes) that define diagnoses, procedures, and other services.

For diagnoses, the medical record review staff should review codes based on the diagnoses stated by the provider in the medical record, (not based on diagnoses indicated by their own judgement). Because of this, medical record review staff should be experienced clinical coding validators. Clinical coding validators should have substantial clinical background, including anatomy and physiology, pathology, microbiology, pharmacology, and disease process. They use this clinical understanding, combined with their knowledge of appropriate coding guidelines, to assign the right codes to a record. When a medical record lacks sufficient documentation to select the most specific code(s), clinical coding validators may consult with each other and other health care professionals to answer clinical questions.

Sources of coding errors, described by the Institute of Medicine in a study of the reliability of coded data in the National Hospital Discharge Survey included:

1) Incorrect selection or sequencing of principal diagnosis codes
2) Incorrect selection of other individual codes
3) Coding diagnoses or procedures not documented in the medical record
4) Errors caused by mistakes in entering the data into a database
5) Failure to review the entire medical record

Building on these ideas, the EQRO should develop an error categorization scheme to identify areas of incompleteness and inaccuracy. Codes assigned by the clinical coding validators will have three components, as separate data elements. Data can be reported for any single error component, or by any combination of these codes. These error components are:

> **Level:** Identifies whether an encounter is present in the database. The presence or absence of an encounter determines the strategy followed by the reviewer to complete the review.
>
> **Type:** Describes if codes or other data are correctly or incorrectly present or absent.
>
> **Source:** Assigns the most likely reason for the type of error found. While this may be a subjective determination, it will be helpful when giving feedback to MCOs/PIHPs so that they can target their data quality improvement processes.

Certain errors may be designated as "critical" for the purposes of an audit. These can be defined differently at varying points in time, because issues that are critical in encounter data validation change over time. For example, in the early rounds of validation, a State may wish to focus on diagnosis and procedure codes, and not on physician specialty or place-of-service. These fields are important, but are of little value if the MCOs/PIHPs are not able to produce accurate clinical coding. Once accurate diagnosis and procedure coding is taking place, knowing accurate specialty and place of service greatly increases the value of the data. Having "tiers of errors" (e.g., critical, serious, moderate, etc.) allows the State to move ahead with using encounter data that are not totally complete and accurate. This is an important goal. In future years, another tier of more refined error types will be added to the critical error types. Assigning the label of "critical" to the errors will be determined by the State.

The findings of each medical record review should be documented on a standardized form, such as the Medical Record Review Findings Tool for Encounter Data Validation found as Attachment 3. It is also important to provide documentation guidelines to the staff performing the medical record reviews. These guidelines should describe exactly how to document the findings of the medical record review. The guidelines also should be closely linked to the reporting requirements and the data elements chosen for validation, so they should be written after the error classification scheme is set. Written policies for interpretation of the documentation guidelines should include the following:

- Directions for reviewing medical records
- Instructions on what to do when faced with conflicting documentation
- Instructions on what to do when no code can be readily assigned
- Use of optional codes
- Definitions of what constitutes errors, and how to document them
- List and location of approved reference materials (e.g., coding manuals, medical textbooks, etc.)
- Whom to consult for additional assistance

ACTIVITY 5: Submission of findings

After the performance of Activities 1- 4, the EQRO will create data tables that display summary statistics for the information obtained from these activities for each MCO/PIHP. Summarizing the information in tables makes it easier to evaluate, and highlights issues with respect to the accuracy and completeness of encounter data. A narrative will accompany these tables, which will highlight MCO/PIHP-specific issues.

In addition, a State at its discretion, may direct its EQRO to undertake technical assistance to its MCOs/PIHPs to improve the accuracy and completeness of encounter data.

END OF PROTOCOL

EXAMPLES OF RECOMMENDED DATA QUALITY STANDARDS FOR EVALUATION OF SUBMITTED ENCOUNTER DATA FIELDS (Physician and Other Provider)		
Data Element	**Expectation**	**Validity Criteria**
Enrollee ID	This should be a valid ID as found in the State's eligibility file. Use the State's ID unless State also accepts SSN	100% valid
Enrollee Name	This should be captured in such a way that makes separating pieces of name easy. There may be confidentiality issues that make this difficult to obtain. If collectable, expect data to be present and of good quality.	85% present. Lengths should vary and there should be at least some last names >8 digits and some first names < 8 digits. This will validate that fields have not been truncated. Also verify that a high percentage have at least a middle initial.
Enrollee Date of Birth	This should not be missing and should be a valid date	< 2% missing or invalid
MCO/PIHP ID	This is a critical data element	100% valid
Provider ID	This should be an enrolled provider listed in the provider enrollment file	95% valid
Attending Provider ID	This should be an enrolled provider listed in the provider enrollment file (also accept the MD license number if listed in provider enrollment file)	> 85% match with provider file using either provider ID or MD license number
Provider Location	The minimal requirement is a county code, with zip code being strongly advised	≥ 95% with valid county code ≥ 95% with valid zip code (if available)
Place of Service	This should be routinely coded, especially for physicians	≥ 95% valid for physicians ≥ 80% valid across all providers
Specialty Code	This is coded mostly on physician and other practitioner, optional on other types of providers	Expect > 80% non-missing and valid on physician or other applicable provider type claims (e.g. other practitioners)
Principal Diagnosis	This is well coded except by ancillary type providers	> 90% non-missing and valid codes (using ICD-9-CM lookup tables) for practitioner providers (not including transportation, lab and other ancillary providers)
Other Diagnosis	This is not expected to be coded on all claims, even with applicable provider types, but should be coded with a fairly high frequency	90% valid when present

Data Element	Expectation	Validity Criteria
Date of Service	Dates should be evenly distributed across time	If looking at a full year of data, 5-7% of the records should be distributed across each month
Unit of Service (Quantity)	The number should be routinely coded	98% non-zero < 70% should be one if CPT code in range 99200-99215, 99241-99291
Procedure Code	This is a critical data element and should always be coded	99% present (not zero, blank, 8- or 9-filled). 100% should be valid, State-approved codes. There should be a wide range of procedures with the same frequency as previously encountered.
Procedure Code Modifier	This is important to pick up to separate out surgical procedures/anesthesia/asst. Surgeon. It is not applicable for all procedure codes	> 20% non-missing. Expect a variety of modifiers both numeric (CPT) and Alpha (HCPCS). The more common codes, which should appear with at least a minimal frequency, are 47 (anesthesia) and 80 (asst. surgeon).
EPSDT Indicator (All States might not have this.)	If this field is used, the beneficiary should be < 21 years of age and the provider should be certified to administer EPSDT screens	95% enrollees < 21; 85% providers certified as EPSDT
Patient Discharge Status Code (Hospital)	This should be valid code for inpatient claims, with the most common code to be a "Discharged to Home". For outpatient claims, it can be coded as "Not Applicable".	For inpatient claims, expect >90% "Discharged to Home". Expect 1-5% in all other values (except "not applicable" or "unknown").
Revenue Code (Hospital)	If the facility uses a UB92 claim form, this should always be present	100% valid

Source: The MEDSTAT Group

Attachment 2

TABLE OF BENCHMARK UTILIZATION RATES (for services incurred between XX/200x and YY/200x)			
Measure	**Value from MA FFS or PCCM**	**Value from Comparable State or States**	**Other Comparison Value**
Inpatient Discharges			
Inpatient LOS Overall By high volume DRGs By eligibility category/patient cohort			
Ambulatory Surgeries Total # surgeries By high volume CPT codes or by ambulatory surgery categories Total # surgeries/1,000 enrollees By high volume CPT codes or by ambulatory surgery categories			
Number of Providers Primary care physicians Specialists Other (e.g., mental health providers)			
Number of Enrollees Total # By eligibility category By age/sex categories			
Number of Users (i.e., enrollees who used services) Total # By eligibility category By age/sex categories			
Visits Total # #/enrollee #/user by visit categories (e.g., well child, well adult, ob/gyn, mental health, substance abuse, etc.)			
Other Services (e.g., prescription drug) Total # #/enrollee #/user by service category			

Attachment 3

Sample Medical Record Review Findings Tool for Encounter Data Validation

Patient ID Number:_____ Medical Record Number:_____

Patient Name:_____ Completion Date:_____

Provider Name:_____ EQRO Reviewer:

Coder Reviewer:_____

Attending Physician Name:

Visit Dates: Begin Date:_____ End Date:_____

Required Review: (Check one)

[ ] Office Visit - (excludes dental and mental health / substance abuse visits)

[ ] Office Visit - mental health / substance abuse

[ ] Office Visit - dental

[ ] Inpatient admission - (excludes mental health / substance abuse visits)

[ ] Inpatient admission - mental health / substance abuse

[ ] Other types of encounters as specified by the State (e.g., laboratory, pharmacy, physical therapy). Specify:_____

Diagnosis Codes and Descriptions				
	Diagnosis Code	**Match**	**No Match**	**Diagnosis Description**
a.				
b.				
c.				
d.				
If the diagnoses in the record do not match the billed information, write the correct diagnosis description(s) on the lines provided.				

Procedure Codes and Description				
	Procedure Code	**Match**	**No Match**	**Procedure Description**
a.				
b.				
c.				
d.				
e.				
f.				
If the procedures in the record do not match the billed information, write the correct procedure description(s) in the spaces provided.				

Revenue Center Codes and Descriptions						
	Revenue Center Codes	**Revenue Center Descriptions**	**Match**	**No Match**	**Correct Revenue Center Description**	**Correct Code**
a.						
b.						
c.						
d.						
e.						
If Revenue Centers in the record do not match the billed information, write the Correct Revenue Center Description(s) in the spaces provided.						

NOTE: The EQRO should tailor and add to this form to address all data fields under review.

END OF PROTOCOL ATTACHMENTS