Medicaid Program Integrity Manual
Chapter 9 – Data Analysis

Table of Contents

(Rev. 1, Issued: 09-23-11)

Transmittals for Chapter 9

9000 – Identifying Potential Audit Subjects

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

The DFRD identifies potential payment errors and trends related to fraud, waste, and abuse. This is accomplished by analyzing Medicaid claims data for potential overpayments, reviewing and identifying fraud, waste, and abuse trends, and conducting studies to support MIG activities and State Medicaid integrity programs. Identifying potential fraud, waste, and abuse related overpayments is accomplished in two phases (concept and algorithm development) and includes four steps:

1. A MIG staff person or the Review MIC proposes new algorithm concepts to DFRD for approval.
2. Subject Matter Experts (SMEs) within the DFRD review the concepts and either accept or reject the new algorithm concept.
3. Accepted algorithm concepts are then prioritized.
4. The DFRD authorizes the development and/or analysis of the algorithms by the Review MIC.

New algorithm concepts to identify potential fraud, waste, and abuse overpayments are initiated based on factors including, but not limited to: referrals from authoritative sources (Review MICs, Audit MICs, OIG, DOJ, States, etc.), specific State collaborations (high visibility collaborations or analysis based on known or perceived hot zones), CMS Medicare-Medicaid dual eligible crossover issues, previous experiences, and other environmental and mass media news sources.

Approval of new algorithm concepts are prioritized based on criteria including: return on investment potential, complexity, individual State policies, legal defensibility, data availability, and data analysis limitations.

Once an algorithm concept is identified and approved it is developed by the Review MIC. Once the algorithm is developed and it is accepted by the DFRD, it is available for assignment to the Audit MIC. The Audit MIC then reviews State policy and audits claims data to identify potential overpayments.

9005 – Algorithm Development, Acceptance, & Data Analysis

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

New algorithm development is dependent on the availability of quality data specific to the approved concepts process described above. During the validation process of the proposed algorithm, the DFRD determines whether each concept overlaps or complements any existing analysis, the level of effort needed and potential return on investment for the development of the algorithm, and the relevance to the Medicaid program on a national level.

Once a concept is finalized, the Review MIC develops the algorithm for the approved concept that will identify potential fraud, waste, and abuse payments. During development of the algorithm, the Review MIC runs it against the MSIS data in the Information Technology Infrastructure and the findings from the algorithm are reviewed by policy, clinical and technical subject matter experts from the DAA, the DFRD and the Review MIC. Claims information and other related data are analyzed to identify potential errors or potential fraud by claim characteristics (e.g., diagnoses, procedures, providers, or beneficiaries) individually or in the aggregate. The algorithm development and refinement process is an integrated, ongoing component of fraud, waste, and abuse detection and research and can be modified and rerun in a timely manner. Results are used to identify potential targets for audit.

Analysis of the data includes:

- Reviewing the data and conducting data investigation to run frequency distributions on certain variables and run validity checks on clinical codes;
- Looking at trends by quarter and annual intervals for each claim type to establish baseline and identifying areas of potential errors;
- Looking for adjustment indicators and missing values to ensure the variables needed for the algorithm are well populated;
- Conducting a data quality evaluation and making recommendations based on experience with the MSIS data;
- Having technical and clinical reviews of both algorithm specifications and output. Assisting in defining appropriate leads, removing false positives, and providing short, effective lead lists; and
- Establishing by claim and by provider minimum thresholds and recommend leads that are pursuable under relevant State Medicaid laws.

The Review MICs get assignments on algorithms from the DFRD on a monthly basis. They are tasked with developing new algorithms and data models that can be used to identify abnormalities and individual or group indicators that describe statistically significant outliers or aberrant trends. Examples of indicators or variables are:

- Standard deviation from the mean;
- Percent above the mean or median; and/or
- Percent increase in billing activity, payment charges and number of visits/services from one period to another.

9005.1 – Personnel

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

The DFRD and the Review MICs include staff with clinical expertise (e.g., registered nurses, clinical pharmacists) and a mix of technical and statistical skills in programming (SAS, SQL, and Oracle), data mining, statistics and Medicaid subject matter experts. The

DFRD and the Review MICs are responsible for the development of algorithms/models and are responsible for identifying potential audit targets for Audit MIC through their analytical work. They make use of available data and apply innovative analytical methodologies critical to the success of Medicaid Integrity Programs.

The DFRD and the Review MICs have staff with appropriate training, expertise and skills to conduct systematic analyses and clinical evaluation of claims data for the development of algorithms for the new concepts. The DFRD and the Review MIC analysts use research and experience in the field to develop approaches and techniques useful in the data analysis of the algorithm. In addition, staff continually maintains communication with State Medicaid agencies concerning policies and data issues relevant to their data analysis activities.

The DFRD and the Review MICs are expected to provide State specific knowledge and apply State policy to algorithm and model development.

To date, the DFRD and the Review MICs have developed 105 algorithms covering service areas that include dental, durable medical equipment, inpatient hospital, lab and X-ray, nursing facilities, outpatient hospital, physician, prescribed drugs, and psychiatric services.

9010 – Sampling and Extrapolation

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

The MIG examined the feasibility of implementing a Sampling and Extrapolation strategy for the Medicaid provider audit program managed by the MIG. The MIG contracts with Audit MICs to perform audits of paid claims that have been identified by algorithms as potential fraud, waste, and abuse overpayments. The main goal is to establish a gold standard MIG sampling plan that can be used by all the Audit MICs so that there is no ambiguity in contractors' understanding their role and responsibility in conducting the sampling, extrapolation and audits. The MIG used sampling and extrapolation during test audits; however it is not currently being used in audits conducted as part of the National Audit Program.   The MIG plans to systematically pursue greater use of extrapolation in the future as the data is refined.

9015 – Sources of Data – The Information Technology Infrastructure, MMIS, MSIS

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

9015.1 – THE INFORMATION TECHNOLOGY INFRASTRUCTURE

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

The MIG has developed a scalable suite of data hosting, mining, and analysis services called the Information Technology Infrastructure. This infrastructure provides a high performance, clustered database system with terabyte-scale capacity and data mining software, hosted at the University of California San Diego (UCSD) campus. The database and support software operates on a scalable cluster of high memory servers that connect to an open architecture storage area network environment providing high bandwidth connectivity to expandable storage capacity. The system is configured to support data analysis and mining algorithms that allow the MIG to perform Medicaid fraud, waste, and abuse overpayment prevention and detection.

The Information Technology Infrastructure consists of a variety of Commercial Off-the-Shelf (COTS) software and hardware. From a user's standpoint, there are three major software tools:

1. **Statistical Analysis Software (SAS)**
   SAS is a statistical analysis application. SAS provides a user interface through which MIG users and Review MICs can analyze the Medicaid data using established CMS algorithms, generate end user statistical reports, and conduct basic data mining.

   SAS is used for routine analysis and reporting. SAS Analytics provides a wide range of statistical analysis tools from traditional analysis of variance to exact methods and dynamic visualization.

   SAS provides users with the ability to select the range of data on which they would like to perform analysis. Additionally, users may choose the type of analysis they would like to perform. Users can customize the information and format that is returned to them and designate if they would like to save the information as a report.

2. **Oracle Data Miner (ODM)**
   ODM is an advanced data mining application for identifying data anomalies and trends. MIG users and Review MICs will use ODM to generate end user reports to help identify suspected fraud, waste, and abuse. Through ODM's user interface, users can select the range of data on which they would like to perform analysis and choose the type of analysis they would like to perform. ODM allows users to manipulate the data analysis and searches, and compile and save reports on the Information Technology Infrastructure.

   ODM is used for mining larger data sets, increasing the performance of complex analysis tasks, or running data mining algorithms outside the scope of SAS.

   ODM algorithms that support solutions for classification problems include Decision Trees, Naïve Bayes, Generalized Linear Models (GLM) and Support Vector Machines (SVM). Regression problems can be solved using GLM or SVM. Text mining, feature extraction and anomaly detection utilize SVM and attribute importance uses Minimum Description Length (MDL). Associations employ Apriori and feature extraction uses non-negative Matrix Factorization (NMF), while clustering has several methods available, including hierarchical K and O-means.

3. **Oracle Business Intelligence Enterprise Edition (OBIEE)**
The Oracle Business Intelligence (BI) suite of applications provides a comprehensive collection of BI products, delivering the full range of BI capabilities including interactive dashboards, full ad hoc querying and reporting, proactive intelligence and alerts, precise reporting, real time predictive intelligence, disconnected analytics, and more. The Oracle Business Intelligence Suite is based on a proven, modern Web Services Oriented Architecture that delivers true next generation BI capabilities. As resources are available, Information Technology Infrastructure Team members will scope, develop, and utilize the Oracle BI suite to allow for enhanced web based access to BI tools and reporting capabilities.

9020 – Sources of Data

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

The following is a list of data sources available on the Information Technology Infrastructure.

9020.1 – Medicaid Statistical Information System (MSIS)

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

The MSIS was developed in 1999 to provide the CMS with a detailed national database of program information capable of supporting a broad range of analytic and user needs. Using MSIS requirements, States supply the CMS with eligibility and paid claims information extracted from their Medicaid Management Information Systems (MMIS). The MIG Information Technology Infrastructure relies heavily on Medicaid data to conduct program integrity activities required by section 1936 of the Act. The CMS requires that States extract certain sets of raw Medicaid eligibility and claims data from their MMIS and submit them in a standardized format to the MSIS. The States submit five types of data to CMS on a quarterly basis:

- Eligibility actions;
- Inpatient hospital claims;
- Long term care claims;
- Prescription drug claims; and
- All other outpatient claims.

9020.2 – Social Security Administration (SSA) Death Master File

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

The Death Master File provided by SSA contains the death records for all individuals who registered with the SSA.

9020.3 – National Plan and Provider Enumeration System (NPPES)

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

National Provider Identifier (NPI)

HIPAA mandates the adoption of national, standard unique identifiers for health care providers and health plans. As a result, the CMS developed NPPES to assign unique NPIs for all registered providers and health plans.

9020.4 – Third Party Files

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

A. Drugs Files

The National Drug Data files provide prices, descriptions, and collateral clinical information on drugs approved by the US Food and Drug Administration (FDA), plus commonly used over the counter drugs.

B. National Correct Coding Initiative (NCCI)

The NCCI was developed by the CMS to promote national correct coding methodologies and to control improper coding leading to inappropriate payment. The Correct Coding Edits table and the Mutually Exclusive Edits table include code pairs that should not be reported together for reasons explained in the Coding Policy Manual.

C. The Current Procedural Terminology (CPT)

The CPT code set includes the codes, descriptions, and guidelines intended to describe procedures and services performed by physicians and other health care providers and are maintained by the American Medical Association.

D. Healthcare Common Procedure Coding System (HCPCS)

The HCPCS is a standardized coding system used primarily to identify products, supplies, and services not included in the CPT codes, such as ambulance services and durable medical equipment, prosthetics, orthotics, and supplies.

E. Diagnosis Related Group (DRG)

The DRG is a system used to classify hospital cases into one of approximately 500 groups expected to have similar hospital resource use. Developed for Medicare as part of the prospective payment system, DRGs are assigned based on ICD diagnoses, procedures, age, sex, discharge status, and the presence of complications or comorbidities.

F. International Classification of Diseases 9th Revision, Clinical Modification (ICD-9-CM)

The International Statistical Classification of Diseases and Related Health Problems (most commonly known by the abbreviation ICD) provides codes to classify diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances and external causes of injury or disease.

The CMS is working with the States to implement the conversion from ICD-9 to ICD-10 codes to ensure all HIPAA transactions, including outpatient claims with dates of service, and inpatient claims with dates of discharge on and after October 1, 2013 utilizes ICD-10 codes. Information and guidance regarding the conversion is available on the CMS website. In addition, informational bulletins, as well as other communication resources are being shared with the States to help facilitate and support the ICD-10 implementation.

9025 – SECURITY

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

9025.1 – SYSTEM SECURITY

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

The MIG works with a variety of information using a variety of system tools. The security level required for each operational task is determined by the type of information to be protected as categorized in Federal Information Processing Standard (FIPS) Publication 199. Security is integrated into business processes using an integrated Life Cycle approach based on the National Institute of Standards and Technology (NIST) 800 series government publications which contain government recommended procedures and criteria for assessing and mitigating threats. The CMS has created their own CMS integrated IT Investment and System Life Cycle Framework for Security (CMS ILC) adapting the NIST 800 series of publications to the unique CMS environment. The CMS ILC includes specific roles and responsibilities for personnel, reviews (4 Governance Reviews and 12 Projects Reviews), and documents (e.g., System Security Plan, Information Security Risk Assessment, Test Plan, Contingency Plan). For each identified security risk, a mitigating control must be implemented or in special cases, low risks may be accepted.

The MIG handles information about persons (in the form of claims data), financial, budgetary, commercial proprietary information (provider information), internal administration (MIG operations), and other Federal Agency information (e.g., Social Security, law enforcement), all of which is categorized at the moderate security level in FIPS-99. Using the CMS ILC, the MIG has implemented a variety of mitigating controls to protect the data used in MIG operations and we are continuously reassessing security threats including mandated tri-annual security reviews. Examples of some of the mitigating security controls the MIG uses to protect data include 2-factor login authentication, cryptography, isolated network connectivity, firewalls, virus and intrusion detection, system software patching, and staff training.

9025.2 – Physical & Operational Security Requirements

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

To ensure a high level of security for the MICs, the MICs must develop, implement, operate, and maintain security policies and procedures that meet and conform to the requirements of the Business Partners Security Manual (BPSSM) and the Core Security Requirements (CSR) and its operational appendices (A, B, C, and D). The BPSSM is located at:
http://www.cms.hhs.gov/manuals/downloads/117_systems_security.pdf and the CSR is at
http://www.cms.hhs.gov/it/security. Further, the MICs must adequately inform and train all their employees to follow all security policies and procedures so the information the MICs obtain is confidential.

The MICs collect and use individually identifiable information on behalf of the MIG to routinely perform the business functions necessary for the administration of MIP activities. Any data the collected by the MICs, including sensitive information obtained as a part of the administration of their contracts is the property of the MIG. Consequently, any disclosure of individually identifiable information by the MIC without prior consent from the individual to whom the information pertains, or without statutory or contract authority, requires prior approval from the MIG.

9030 – Policy, Clinical, and Technical Quality Assurance Process

(Rev. 1, Issued: 09-23-11, Effective: 09-23-11, Implementation: 09-23-11)

Quality assurance is an integrated, ongoing component of MIG and Review MIC activities. With the DFRD as the lead, quality assurance is performed by the DFRD and the DAA for general surveillance and review of reports submitted by Review MICs and is a tool for identifying potential abnormalities or anticipating potential problems in audits. The quality assurance process analyzes claim information and other related data to verify potential errors in an algorithm or with its results.

During the development of the algorithm, a sample is sent to the State for validation. If the State finds issues in the sample, the Review MIC contacts the DFRD for guidance. The DFRD reviews the concerns and makes recommendations so a valid algorithm and an accurate Algorithm Findings Report (AFR) will be produced. The quality assurance process includes in-depth policy, clinical, and technical analysis used to confirm the findings contained in the AFR. The policy review looks at all Federal and State factors that may affect algorithm findings, while the clinical review analyzes the logic utilized to identify the medical diagnosis, treatment, services contained in the AFR. The technical review validates the programs in the header section, reviews the logic in the SAS code, and validates the Review MIC programming requirements.

Policy review of the AFR considers:

- State specific payment and coverage policies;
- State waivers, where applicable;
- State laws;
- Federal laws;
- Medical Coding or Classification policies; and
- State sample report validation or invalidation.

Clinical review of the AFR considers:

- Language;
- Medical coding and classification guidelines relating to the diagnoses and procedures within algorithms and or models under review;
- State specific payment policies and guidelines;
- Federal policies;
- Clarity;
- References;
- Citations;
- Congruency of the concept with State policy and regulations;
- Limitations and exceptions;
- Data anomalies; and
- Those recommendations are correlated with findings.

Technical review of the AFR consists of:

- Confirmation of the appropriate use of data based on concept description;
- Confirmation that the data is clinically based on concept description;
- Confirmation the output is consistent with defined policy in the concept description;
- Confirmation that the validity of the result findings coincides with what is written in the limitations, exclusions, and recommendation descriptions; and
- Confirmation of the accuracy of key fields in relation to the concept description such as:
    - National Drug Codes (NDC);
    - Current Procedural Terminology (CPT) codes;
    - Health Care Common Procedural Coding Systems (HCPCS) Codes;

- Current Dental Procedures (CDT) Codes;
- ICD-9-CM diagnoses and procedure codes;
- Adjudication dates;
- Medicaid Paid Amounts;
- Adjustment codes; and
- Algorithm review under review dates

Transmittals Issued for this Chapter

Rev #	Issue Date	Subject	Impl Date	CR#
R1MPI	09/23/2011	Initial Publication of Manual	09/23/2011	NA

Back to top of Chapter