



PharMetrics Health Plan Claims Data User Guide & Data Dictionary

February 2014

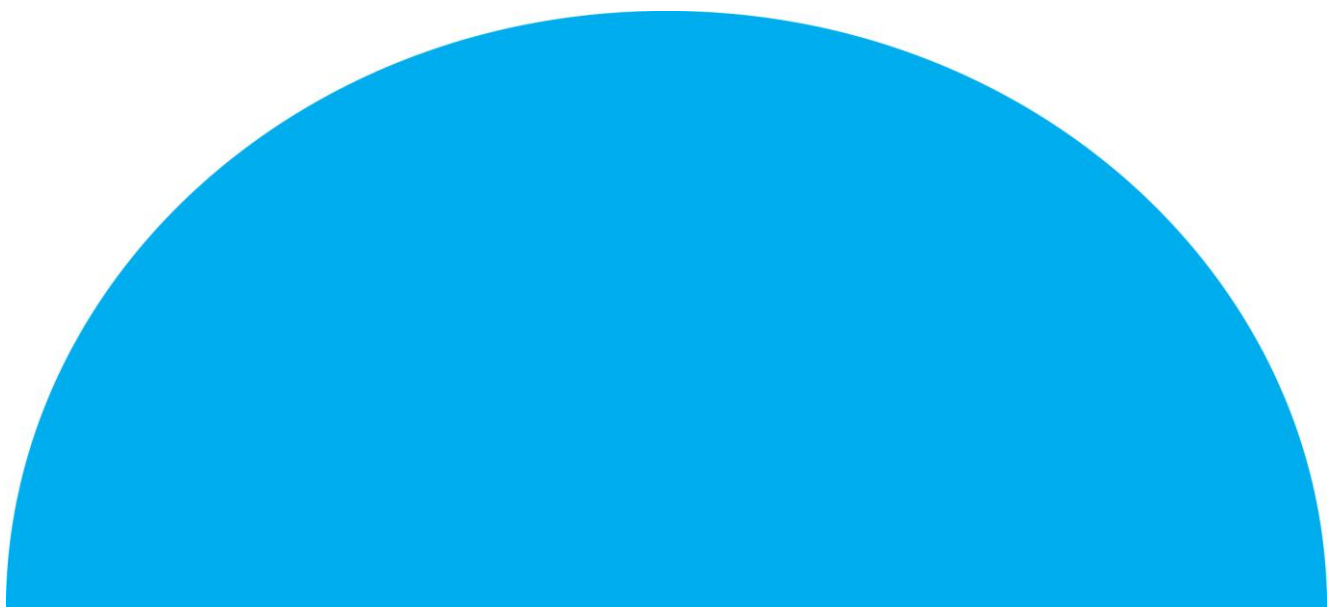




Table of Contents

| | |
|---|----|
| Table of Contents | 2 |
| Content Summary | 3 |
| 1. General Information | 4 |
| 2. File Details | 6 |
| 3. Detailed Variable Descriptions | 17 |
| 4. Setting Level Segmentation Approach | 22 |
| 5. Frequently Asked Questions | 26 |
| 6. Standard Field Values | 31 |
| 7. U.S. Health Claims Coding | 37 |
| 8. Data Quality Overview | 41 |



Content Summary

Section I: General Information

This section provides general information about the data, its source, and basic descriptions about many of the positive aspects and limitations of claims data.

Section II: File Details

This section contains a complete listing of all file fields and a brief description of their contents. Three tables are present in this section, with one for the Claims Detail file, one for the Eligibility file, and one for the optional Lab Results file (Please note, the Lab Results file is an additional purchase not included with a standard extract).

Section III: Detailed Variable Descriptions

This section provides a thorough and more detailed explanation regarding many of the variables catalogued in Section II: File Details. Further detail can also be found in this section describing our methodology used in addressing general data limitations found throughout the industry.

Section IV: Setting Level Segmentation Approach

This section provides recommendations on how best to group claim detail lines for accurate counting of services and accurate reflection of costs per service.

Section V: Frequently Asked Questions

This section provides brief responses to many of the more common questions with references back to any more detailed sections throughout the document.

Section VI: Standard Field Values

This section contains the values for several variables that are standardized by IMS (e.g. specialty, provider type, etc.).

Section VII: U.S. Health Claims Coding

This section provides detailed information about the various coding schema used in the reimbursement of health care services.

Section IX: Data Quality Overview

This section provides an overview of the data processing and quality review of data contributors to the health plan data.



1. General Information

The LifeLink® Health Plan Claims Database is comprised of commercial health plan information obtained from managed care plans and other sources throughout the United States. It is paid claims data, which by definition is information collected by the medical plans from medical service providers to facilitate the adjudication and payment of health insurance benefits on behalf of the plan's enrolled members. It is fully adjudicated medical and pharmaceutical claims for over 75 million unique patients from over 80 health plans across the U.S. (approximately 25 million covered lives per year).

The database includes both inpatient and outpatient diagnoses (in ICD-9-CM format) and procedures (in CPT-4 and HCPCS formats) as well as both retail and mail order prescription records. Available data on prescription records include the National Drug Code (NDC) as well as the quantity of the medication dispensed. The Charge, Allowed and Paid amounts are available for all services rendered, as well as dates of service for all claims. Additional data elements include demographic variables (age, gender, geographic region), product type (e.g., HMO, PPO), payer type (e.g., commercial, self-pay), provider specialty, and start and stop dates of health-plan enrollment.

Records in the LifeLink® Health Plan Claims Database are generally representative of the national, commercially insured population in terms of age and gender. The data are also longitudinal with average member enrollment duration of over three years. Only health plans that submit data for all members are included in the database, ensuring complete data capture and representative samples. Data contributions are also subjected to a series of quality checks to ensure a standardized format and to minimize error rates.

A contributing plan's data undergoes rigorous data quality review prior to its addition to LifeLink ®Health Plan Claims Database. Reviews are conducted at the initial submission of data, and at each subsequent transmission of data. Each incoming dataset must conform to basic data validity norms, and it is thoroughly analyzed to ensure its inclusion into the database is appropriate. Quality reviews are also performed after records are grouped to ensure end-to-end data integrity.

Our data philosophy is not to change data that is potentially erroneous, but rather to fully understand variations and ensure anomalies are minimal. Each study, norms product, or extract either includes or excludes data based upon the specific requirements as stated by the end user, or equips the user to make any inclusion or exclusion decisions. It is believed this approach allows for the most robust application of that data, and best allows the end user to adjust the dataset to their needs.



Because the data comes from a number of different sources, we focus on standardizing the variables so that comparable analyses can be performed easily and confidently. Significant effort is expended in cross-walking key variables to standard definitions, and in continuous data maintenance to ensure these mappings remain current and accurate.

A standard extract from the LifeLink® Health Plan Claims Database consists of the following files:

- Claims Detail - The claim specific elements are from the detail record lines for all claims submitted to the health plan for payment, whether on a HCFA-1500, UB04, via a pharmacy claims system or PBM, or any other source. The file is sorted by patient id and from date.
- Enrollment - demographic and eligibility information for the selected population. Consists of one record per member and is sorted by patient id
- Enrollment2 - product and paytype information by month, three records per enrollee
- NDC Reference - contains reference information for all the NDC codes found in the extract.
- Diagnosis Reference - contains reference information for all the ICD-9-CM diagnosis codes found in the extract.
- Procedure Code Reference - contains reference information for all the CPT and HCPCS codes found in the extract.
- NDC Codes Used for Patient Selection - contains all the NDC codes used in selecting patients for the extract.

In collecting enrollment data from so many sources covering different time periods, adjustments have to be made to allow for an unbiased comparative analysis. The methods utilized to accomplish this end are thoroughly described in Section III. It is highly recommended that section be studied prior to any analysis relying on enrollment data.

NOTE: The claims and enrollment files can be linked through the PAT_ID field.



2. File Details

The tables below provide information on the variables present in a standard extract data file.

| CLAIMS | |
|--|---|
| Variable | Description |
| ALLOWED | <p>Allowed Amount: The amount the plan allows for a particular service. It is typically the paid amount plus any member liability (e.g., co-pay, deductible, coinsurance).</p> <p><i>This is the cost variable that most closely reflects the burden of illness. If any non-imputed cost variable is used, confinements should be kept whole. Refer to Section IV.</i></p> |
| CHARGE | <p>Charge Amount: This is the amount charged for a particular service by the servicing provider.</p> |
| CLUSPVID | <p>Provider ID for Each Cluster: The cluster provider ID identifies the provider who directed the services within a cluster.</p> |
| CONF_NUM | <p>Confinement Number: Assigned number for unique inpatient stays. The confinement is constructed by using the 'frm_dt' on the first room and board record and the 'thru_dt' on the last facility record in a series of facility records that have the same provider id and overlapping or contiguous dates. The confinement number is then assigned to all records that fall within that time frame.</p> <p>Confinement number is not unique within an extract. You must use pat_id and conf_num for uniqueness.</p> |
| DAYSSUP | <p>Days Supplied: Where available, the days supplied value of a prescription is reported.</p> |
| DIAG1 DIAG2 DIAG3 DIAG4 | <p>Diagnosis 1-4: Each record may include up to four ICD-9 (International Classification of Diseases, 9th revision). These codes describe the patient's condition or diagnosis.</p> |
| FROM_DT | <p>Service Begin Date: The initial date of service delivery.</p> |



| CLAIMS | |
|-----------------|---|
| Variable | Description |
| NDC | National Drug Code: The NDC identifies the manufacturer, product name, and package size of each approved or repackaged prescription drug. These codes are assigned by the FDA. |
| PAID | Paid Amount: The dollar amount actually paid by the plan for a service. |
| PAT_ID | Patient ID: An encrypted unique ID for each patient. |
| POS | Place of Service: A HCFA standard variable identifying the location, or place, where the services were rendered |
| PROC_CDE | <p>Procedure Code: A unique code identifying each procedure. These are either:</p> <ul style="list-style-type: none"> ❑ <u>CPT-4 Codes</u>: AMA developed codes that describe medical, surgical, and diagnostic services performed by clinicians. ❑ <u>HCPCS Codes</u>: Codes developed by HCFA to describe the provision of supplies, injections, materials, services, durable medical equipment, and non-physician procedures. ❑ <u>Revenue Center Codes</u>: These codes are used to associate certain costs with specific revenue centers in a facility. |
| PROV_ID | Provider ID: An encrypted unique ID for each service provider. |
| PTYPEFLG | <p>Provider Type:</p> <ul style="list-style-type: none"> ❑ 0 - Clinician ❑ 1 - Facility ❑ 2 - Other <p><i>We have a high degree of confidence that when the PTYPE = 1, the billing represents charges presented on a UB-92.</i></p> |
| QUAN | Quantity Dispensed: Where available, the quantity dispensed value of a prescription is reported. |
| RECTYPE | <p>Record Type: Each record is classified as one of the following five record types:</p> <ul style="list-style-type: none"> ❑ (M)anagement ❑ (S)urgical ❑ (F)acility ❑ (A)ncillary ❑ (P)harmaceutical |



| CLAIMS | |
|----------|---|
| Variable | Description |
| REC_SPEC | Specialty: The specialty of the record level or service provider. |
| TO_DT | To Date: The final date of service delivery. |



| ENROLLMENT | |
|-------------------|---|
| Variable | Description |
| CLM_FRST | First Claim Date - The date of the first claim in the claim file. Used as a proxy for enroll first in the event true enrollment doesn't exist for the patient. |
| CLM_LAST | Last Claim Date - The date of the last claim in the claim file. Used as a proxy for enroll last in the event true enrollment doesn't exist for the patient. |
| DER_SEX | Derived Sex - If there is a sex listed on the enrollment record it is used. Otherwise, the sex is derived from the claims file and populated into the field. |
| DER_YOB | Derived Year of Birth - If there is a YOB listed on the enrollment record, it is used. Otherwise, the YOB is derived from the claims file and populated into the field. |
| ENR_FRST | First Enrollment Date - The date of the first enrollment record where true enrollment exists. Modified to correspond with the date of complete claims data for the plan. |
| ENR_LAST | Last Enrollment Date - The date of the last enrollment record where true enrollment exists. Modified to correspond with the date of complete claims data for the plan. |
| ESTRING | Enrollment String - A 252-character string of "X"s and "-"s. Each position in the ESTRING represents a single month. It begins as of January 1995. If a member is enrolled in a particular month, an "X" is placed in the appropriate position. If the member was not enrolled in that month, a "-" is in the position. The Estring is described in depth in Section III, and a Date-Position Reference Table is supplied in Section V. |
| MXCE_FST | First enrollment date where it exists, else, the first claim date |
| MXCE_LST | Last enrollment date where it exists, else, the last claim date |
| PAT_ID | Same as PAT_ID in the claims file. |
| PAT_REGION | Region (Patient): Each record is assigned to one of the following four regions, as defined by the US Census Bureau: <ul style="list-style-type: none">❑ Northeast❑ South❑ MidWest❑ West |



| ENROLLMENT2 | |
|--------------------|---|
| Variable | Description |
| PAT_ID | Same as PAT_ID in the claims and enrollment file. |
| STRING | <p>Enrollment2 String – Same construct as the enrollment esttring, but instead of 'X'S and '-s, values reflective of product type, payer type or pharmacy benefit flag are present.</p> <p>Each enrollee will have either no enrollment2 records (in the event of proxy enrollment). Or, they will have three enrollment2 records.</p> <p>Values present when the string type is 'prd_type' (product type) are:</p> <ul style="list-style-type: none">❑ D- Consumer Directed Health Care❑ H - Health Maintenance Organization❑ I - Indemnity❑ S - Point of Service❑ P - Preferred Provided Organization❑ U – Missing/Unknown <p>Values present when the string type is 'pay_type' (payer type) are:</p> <ul style="list-style-type: none">❑ C - Commercial❑ K – State Children’s Health Insurance Program (SCHIP)❑ M - Medicaid❑ R – Medicare Risk❑ S – Self-Insured❑ T – Medicare Cost❑ U – Missing/Unknown <p>Values present when the string type is 'ben_rx' (pharmacy benefit) are:</p> <ul style="list-style-type: none">❑ Y – The person has a pharmacy benefit❑ N – The person does not have a pharmacy benefit❑ - - Not enrolled, or missing or unknown pharmacy benefit |
| STRING_TYPE | Defines the type of values that will appear in the string – pay_type, prd_type or ben_rx. |



| NDC Reference File | |
|------------------------|--|
| Variable | Description |
| Ahfsc | The American Hospital Formulary Service (AHFS) publishes the AHFSCC Pharmacologic-Therapeutic Classification Compilation codes which are used throughout the health care industry as a means for drug classification. It is less specific than both the GPI therapeutic classification and the USC therapeutic classification system. |
| Brand_indicator | The brand indicator identifies whether a product is trademark, generic or branded generic (gen_brand). Branded generics are: <ul style="list-style-type: none"> • generally a generic drug product • more than one company manufactures this drug product • the manufacturer can be a research and development company or a generic company • sold under a branded generic or a generic name |
| Brand_name | The Brand or Trademark name |
| Dosage_form | The Dosage Form indicates the form (solid, liquid, or gas) in which the drug product is dispensed. The "C6" refers to the sixth subset of the GPI which generally represents the Dosage Form (for example, the GPI for Furosemide Tab 20 mg, 37 20 00 30 00 03 05, has an "03" in the C6 position designating the Dosage Form "TAB"). |
| Drug_class | The first four digits (second subset) of the GPI code represent the Drug Class which identifies specific therapeutic drug classes designed to accommodate more detailed market research. The Drug Class also serves as the structural base for most therapeutic drug monitoring applications (such as dosage screening and disease contraindication monitoring). For example, Drug Group 37, DIURETICS, includes the following Drug Classes: <p>37-00-00-00-00 *DIURETICS*</p> <p>37-10-00-00-00 *Carbonic Anhydrase Inhibitors**</p> <p>37-20-00-00-00 *Loop Diuretics**</p> <p>37-30-00-00-00 *Mercurial Diuretics**</p> <p>37-40-00-00-00 *Osmotic Diuretics**</p> <p>37-50-00-00-00 *Potassium Sparing Diuretics**</p> <p>37-60-00-00-00 *Thiazides**</p> <p>37-90-00-00-00 *Miscellaneous Diuretics**</p> <p>37-99-00-00-00 *Combination Diuretics**</p> <p>The number "99" found in the second or third subset in some GPIs is used in Facts and Comparisons' TCS to identify combination drug products.</p> |



| NDC Reference File | |
|---------------------------|---|
| Variable | Description |
| Fd_drug_name | GPI Generic Name. Identifies the product represented by the GPI which includes generic name, strength and dosage form. |
| Gen_nm | Generic Name/Active Ingredient |
| Gpi | <p>The 14-character GPI (generic product indicator) consists of a hierarchy of seven subsets, each providing increasingly more specific information about drug products. These subsets are structured and identified below:</p> <p>GPI Record</p> <p>12-xx-xx-xx-xx-xx-xx Drug Group *MISC. ENDOCRINE*</p> <p>12-34-xx-xx-xx-xx-xx Drug Class *Posterior Pituitary**</p> <p>12-34-56-xx-xx-xx-xx Drug Subclass *Vasopressin***</p> <p>12-34-56-78-xx-xx-xx Drug Name Desmopressin</p> <p>12-34-56-78-90-xx-xx Drug Name Acetate</p> <p>12-34-56-78-90-12-xx Dosage Form Tablet</p> <p>12-34-56-78-90-12-34 Strength 0.1MG</p> |
| Last_chg_dt | The Last Change Date identifies the most recent date any field in the drug file was changed. |
| Manufacturer_name | The Manufacturer's (Labeler) Name indicates the manufacturer, distributor, and/or division whose name is included on the label. |
| Ndc | <p>Drug products are identified and reported using an 11-digit, 3-segment number called the National Drug Code (NDC), which is a universal product identifier for human drugs. The NDC code identifies the labeler, product, and trade package size. The first segment (5-digits), the labeler code, is assigned by the FDA. A labeler is any firm that manufactures (including repackers or relabelers), or distributes (under its own name) the drug. The second segment (4-digits), the product code, identifies a specific strength, dosage form, and formulation for a particular firm. The third segment (2-digits), the package code, identifies package sizes and types. Both the product and package codes are assigned by the firm. The NDC code is reported in the following configuration: 5-4-2.</p> |
| Route_admin | <p>The Route of Administration indicates how the medication's dosage form is administered to the patient. Examples are:</p> <p>ORAL (OR) Taken by mouth</p> <p>RECTAL (RE) Administered into the rectum</p> <p>SUBCUTANEOUS (SC) Injection through the skin into the loose subcutaneous tissue under the skin</p> <p>INHALATION (IN) Drug administration into the lungs (either during a drawn or forced breath)</p> <p>INJECTION (IJ) A set of one or more injectable routes or the route of injection is not specified</p> <p>INTRAMUSCULAR (IM) Injection into a muscle group</p> |



| NDC Reference File | |
|---------------------------|--|
| Variable | Description |
| Rx_group | <p>Drug Group</p> <p>The two-character Drug Group (first two digits and first subset of the GPI code) classifies general drug products. Examples are:</p> <p>01-00-00-00-00 *PENICILLINS*</p> <p>25-00-00-00-00 *CONTRACEPTIVES*</p> <p>27-00-00-00-00 *ANTIDIABETIC*</p> <p>36-00-00-00-00 *ANTIHYPERTENSIVE*</p> <p>86-00-00-00-00 *OPHTHALMIC*</p> |
| Rx_otc | Indicates whether the product is available only by prescription (Rx) or over-the-counter (OTC). |
| Strength | <p>The Strength is present for products with a single active ingredient. For combination products, the strength is designated as zero. The "C7" refers to the seventh subset of the GPI which generally represents the product Strength (for example, the GPI for Furosemide Tab 20 mg, 37 20 00 30 00 03 05, has an "05" in the C7 position designating the product Strength 20mg)</p> |
| Tee_code | A two-character code indicating the FDA rating of the therapeutic equivalence of a drug product with other pharmaceutically equivalent drug products, as published in the Orange Book. Use of the TEE Code with the GPI enables the end-user to have a list of bioequivalent, substitutable drug products sharing the same active ingredients including the same route, strength, and dosage form. |
| Unit | <p>The Strength Unit of Measure, when combined with the Strength represents the dosage strength as provided by the manufacturer. For example:</p> <p>Drug Product: Indocin</p> <p>Strength: 25</p> <p>Strength Unit of Measure: MG</p> <p>Interpreted Strength: Indocin 25 mg</p> |
| Usc_code | The Uniform System of Classification (USC) is a therapeutic classification system created by IMS America and the Pharmaceutical Marketing Research Group. This system is market-oriented and is unique to the United States. The USC Classification System is more specific than the American Hospital Formulary Service (AHFS) Therapeutic Classification System and less specific than Medi-Span's Therapeutic Classification System. One USC Code can be applied to multiple GPI's while a GPI will have only one USC code. Not all GPIs have a USC Code. |
| Usc_description | Text description of the USC code. |



| Diagnosis Reference File | |
|--------------------------|--|
| Variable | Description |
| DX | Alpha-numeric code for the ICD-9-CM Diagnosis Code |
| Diag_description | Short version of the text description for the code |

| Procedure Code Reference File | |
|-------------------------------|---|
| Variable | Description |
| Proc_cde | Alpha-numeric code for the CPT-4, HCPCS or Revenue Code |
| Proc_description | Short version of the text description for the code |



| NDC Codes Used for Patient Selection | |
|---|--|
| Variable | Description |
| Brand_nm | The Brand or Trademark name |
| Gen_nam | Generic Name/Active Ingredient |
| Ndc | Drug products are identified and reported using an 11-digit, 3-segment number called the National Drug Code (NDC), which is a universal product identifier for human drugs. The NDC code identifies the labeler, product, and trade package size. The first segment (5-digits), the labeler code, is assigned by the FDA. A labeler is any firm that manufactures (including repackers or relabelers), or distributes (under its own name) the drug. The second segment (4-digits), the product code, identifies a specific strength, dosage form, and formulation for a particular firm. The third segment (2-digits), the package code, identifies package sizes and types. Both the product and package codes are assigned by the firm. The NDC code is reported in the following configuration: 5-4-2. |
| Rx_group | Drug Group The two-character Drug Group (first two digits and first subset of the GPI code) classifies general drug products. Examples are: 01-00-00-00-00 *PENICILLINS* 25-00-00-00-00 *CONTRACEPTIVES* 27-00-00-00-00 *ANTIDIABETIC* 36-00-00-00-00 *ANTIHYPERTENSIVE* 86-00-00-00-00 *OPHTHALMIC* |
| Strength | The Strength is present for products with a single active ingredient. For combination products, the strength is designated as zero. The "C7" refers to the seventh subset of the GPI which generally represents the product Strength (for example, the GPI for Furosemide Tab 20 mg, 37 20 00 30 00 03 05 , has an "05" in the C7 position designating the product Strength 20mg) |
| unit | The Strength Unit of Measure, when combined with the Strength represents the dosage strength as provided by the manufacturer. For example: Drug Product: Indocin Strength: 25 Strength Unit of Measure: MG Interpreted Strength: Indocin 25 mg |





3. Detailed Variable Descriptions

Financial Variables:

From the preceding tables it is understood that generally, the amount billed to a health insurance plan is the CHARGE amount, the amount the health plan considers for payment (including any patient liability) is the ALLOWED amount, and the amount for which payment was actually sent is the PAID amount.

Given irregularities in the capture of facility claims, to ensure valid cost analysis, inpatient facility records must be kept whole and the allowed amounts summed. Please refer to Section IV.

Eligibility:

Individual contributors pass enrollment information in a myriad of different ways. Some contributors will create and send a new record for an enrollee if a dependent is added to the enrollee's record. Some will send an enrollment record for each calendar year, even if there is no change to the enrollee's benefits or plan. Some send a new record if an enrollee's classification with their employer changes, even if that has no impact on the enrollee's health benefits. In summary, enrollment records come from contributors in a vast number of variations.

To simplify and standardize the results so they are appropriate for analysis, IMS works with each contributor to determine the most appropriate way to concatenate the numerous changes into one record. The result of those is the creation of one record per person enrolled in the plan with their initial enrollment beginning and the final enrolled dates (as discussed further on in this section, these dates are later adjusted for the valid date range of the contributor's plan). And while the enrollment dates (ENR_FRST & ENR_LAST) will provide this summary interval, the Estring maintains any breaks in enrollment during this period. (The Estring is also discussed in more depth later in this section).

True Enrollment:

A contributor's data is said to have "True Enrollment" if the contributor provided IMS with a file of all the enrolled members. The vast majority of the plans in the LifeLink® Health Plan Claims Database do provide this information. However, on the few plans that do not, actions are taken to provide end-users with appropriate proxies. Of course, extracts can be requested to include exclusively data from plans for which true enrollment information exists. However, even when the plan does provide an enrollment file, it's not unusual for a very small percent of people in the claim file to not have a corresponding enrollment record. In those cases, the individual patient will have a proxy enrollment file.

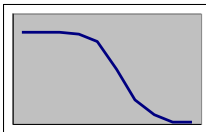


Proxy enrollment records are easy to distinguish – they have blank ENR_FIRST, ENR_LAST and estrng.

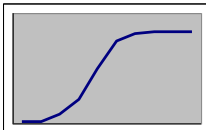
By default all extracts are limited to plans that contribute true enrollment. Unless it was specifically requested otherwise, all patients are from plans that provide this information. Consequently, any patients without enrollment dates are the result of claims existing for individuals for which there was no corresponding enrollment record. For those clients creating their own queries and extracts, it is strongly recommended that in the database selection section, patient level enrollment be set to YES.

Date Adjustments:

When data is produced by a contributor for IMS, it is extracted from their systems using the date of service of a claim, or the paid date of the claim. The existence of both methodologies requires the global addressing of the generation of “tails” in the quantity of claims by month of service:



The graph above illustrates the generally observed trend in the quantity of claims by month of service at the end of a submission period when a plan submits their data by paid date. The far right edge represents the current date. The fall off in claims present is primarily caused by the lag of claims being delivered to the health plan and the time needed for the health plan to process the claims.



The second graph directly above illustrates the generally observed trend in the quantity of claims by month of service at the beginning of a submission period also when a plan submits their data by paid date. The far right edge represents the index date being used in the extraction of the data by the plan contributor. The “run-in” or “run-up” period is caused by the inverse of the phenomenon that created the first graph above. Specifically, if a request were made for all claims paid on or after a certain date in time, the dates of service of those claims paid on that index date would be concentrated immediately before the index date, decreasing in quantity as one looks back further in time.

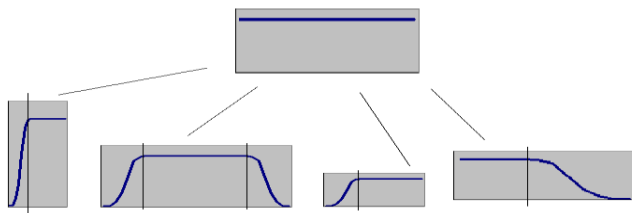
When data is produced by a contributor based on the date of service, the volume of claims with a date of service in month one could be expected to be generally and relatively identical any to any other month in the data. The exception is if a contributor extracts their data



through the current date, from which a trend identical to the one in the first graph above will again be observed.

In reviewing one dataset, these trends can be compensated for rather easily. However, in that data provided arrives at different frequencies, extracted under different parameters, and for different date ranges, each set must be individually studied prior to its being incorporated into the greater database. This is further complicated by the prohibition which prevents IMS from providing contributor specific information.

To resolve these complications, we provide on the eligibility file adjusted, conservative dates based on proprietary algorithms and contributor-specific plan level knowledge. Individual enrollment periods within the dataset exceeding the valid period will be truncated (as illustrated below). In doing so, the set itself can now be integrated into the larger database, and users can now rest confidently assured they have the most complete, valid data available for their study.



During the processing of the data, it is reviewed in total to determine for what valid date range claims data is present:

The dates in the CLM_FRST & CLM_LAST, and the ENR_FRST & ENR_LAST fields are altered to mirror the period for which full claims information is present. (As the MXCE_FRST & MXCE_LST are products of these other fields, they will also be adjusted)

Specifically, the enrollment date fields will mirror the period itself, unless the individual him or herself has an actual enrollment date beginning or ending within the valid date range. The claim first and last date fields will also be revised, to indicate the date of first claim within the valid date range.

When working with these date ranges, it is important to realize that while the ranges themselves are truncated, the individuals' claims are not removed from the claim file. They are not removed as their inclusion allows for the maximum application of the data and the provision of the modified dates allows users to exclude those claims should they wish to. Users should therefore be aware that claims may be observed with dates before or after the



ENR_FRST & ENR_LAST dates on the patient file, and that the CLM_FRST & CLM_LAST dates may not correspond to the earliest and latest dates of service on the claim file.

Estring:

Estring is a 252 character string of "X" and "-" characters. Each position in the string equates to a month-year date, in which the presence of "X" indicates the person was enrolled that month. The Estring is created only on plans for which true enrollment is present, and then only for individuals within those plan for which enrollment date was provided (A very small subset of participants are sometimes noted to not be included on a plan's enrollment file, thought to be due to data generation issues at the contributor level).

There are a significant number of rules applied in the generation of the Estring. If a participant is enrolled on any one day of a month, an "X" is populated for that month. If there is a one-month gap in a series of "X"s, an "X" is placed in the gap month under the assumption the gap was caused by complex eligibility shifts at the plan level. The Estring's initial beginning and final end dates are adjusted to the contributing plan's period of complete data, as described in the preceding section. Finally, the Estring captures breaks in coverage that would otherwise be lost in working solely from the beginning and ending dates of coverage (as described in the following topic).

Continuous Enrollment:

Many studies performed by our customers require periods of what is commonly referred to as "continuous enrollment", or periods of time for which a participant must have a documented period of enrollment. Continuous enrollment can be calculated over:

- A set, universal period of time (... "continuously enrolled for all of 2003"...),
- for a participant-variable period of time (... "enrolled for at least 12 months"...), or
- based upon a period around a participant-specific index date (... "all individuals with a diagnosis of XYZ.XX, and 12 months of continuous enrollment prior to and after the initial date of the first date of service with that diagnosis"...).

Other Enrollment Elements:

In situations where true enrollment is present with a contributor's data, the DER_SEX and DER_YOB are populated with the values passed by the contributor. In situations where true enrollment is not present, IMS will place in these fields the first valid sex and year of birth values on the first claim encountered for the individual.

The MXCE_FRST & MXCE_LST fields on the enrollment file carry the true enrollment begin and end dates, if a plan is providing an enrollment file. If a contributor does not provide an eligibility file, these fields are populated with the contents of the CLM_FRST & CLM_LAST



fields. Please recall from information earlier in this section that all of these values are subject to adjustment, should they fall outside the individual contributor plan's valid period.

Please also note that while true enrollment may have been present for a particular plan, there are situations where a minimal number of individuals within a given plan may not be listed on the eligibility file.

HIPAA Compliance:

IMS has conducted numerous investigations and taken significant actions to ensure full compliance with HIPAA. All appropriate elements are removed or encrypted, with the PAT_ID and DER_YOB (year of birth) remaining often as the only directly unique individual identifiers. Statistical reviews are also conducted to ensure rare treatment patterns that could lead to de-identification are also noted. Should such an incident be observed in a requested extract, prior to delivery the matter will be discussed with the data client requesting the data. Additionally, the year of birth for elderly patients (those over 85) is modified so that they all appear to be 85 years old in order to ensure de-identifiable patients.

Specialty

The specialty provided on the claims file is the record level specialty for the provider of service.

In most instances, the assigned specialty is consistent with what was provided by the plan. However, there are situations where the assigned specialty may be over-written with and inferred specialty. Inference uses the data and clinical algorithms to make an assignment based upon the evidence of who the provider is treating, for what disease and the services they are providing. Specifically the three instances in which a specialty would be inferred are when a specialty is not provided, when a provider is coded within the general "internal medicine" category but the diagnosis and procedural codes present would not support that, and when a Radiation Oncologist is potentially miscoded as a Radiologist.

Provider level analysis is not possible with data from an IMS extract. No efforts are undertaken to standardize provider identifications across the individual contributors' plans, and as such, any provider noted in an extract could exist in various other forms within the data (i.e. simply listed under a different code number, under a medical group, as a component of a surgical or facility service, etc...).



4. Setting Level Segmentation Approach

This section covers methods that have been developed by IMS for counting health care services and determining costs for those services that goes beyond simple claim detail line counting. It's important to make the distinction that this section covers methods rather than specific data elements.

For the purposes of segmenting and reporting cost and utilization by treatment setting or event, it is recommended that the claims detail lines be aggregated so as to combine claim detail lines into meaningful units of analysis. Terminology used through out this explanation includes:

Event/Visit – a grouping of claim detail lines into a set that basically reflects a certain type of healthcare setting and/or utilization. An inpatient hospitalization is an event, as is an outpatient surgery or a doctor's office visit. Event is the generic, high level term used to denote all of those things. The term Visit is used mostly interchangeably, with the finer by generally immaterial point of difference being that an event could in actuality be comprised of visits to multiple geographic locations.

Inpatient Hospitalization – also referred to as inpatient stay and/or confinement, is an aggregated set of health care claims reflecting a room and board charge and all associated records/charges for that same event.

ORDER OF OPERATIONS

To correctly process the data into the reporting categories, the order in which the data is processed within each of the three categories (inpatient, outpatient, retail drug) is also key. This document will define a variety of methods of aggregating and counting claims based upon the nature of how the individual types of services are delivered and how they typically appear in claims data.

In order to analyze treatment setting or event based utilization, records should first be divided into three major categories:

- Retail drug (where ndccode <> null)
- Inpatient (where conf_num <> null)
- Outpatient (where conf_num = null)

When data are aggregated in this way, the allowed amount is the recommended cost measure.



RETAIL DRUG CLAIMS

Retail drug claims need not be aggregated; each record stands alone. But, a single record may represent one month's supply of drug or three months supply. Therefore, normalization should occur for counting purposes. A prescription is defined as one 30-Day Supply.

To count a prescription - If Days Supplied is less than or equal to 30, then the number of prescriptions = 1. If Days Supplied is greater than 30, then the value is divided by 30 and the quotient is rounded to the nearest whole number. For example, 90 Days Supplied would be defined as 3 prescriptions ($90 \text{ Days Supplied} / 30 = 3$), 33 Days Supplied would be interpreted as 1 ($33/30$ - rounded to 1) and 58 Days Supplied would be interpreted as 2 ($58/30$ - rounded to 2).

Use allowed amount for the cost of the prescription.

To report an average cost per prescription divide the total allowed amount by the number of prescriptions.

Drug co-pay can be derived by subtracting the paid amount from the allowed amount.

INPATIENT HOSPITALIZATIONS

The unique combination of patient id + conf_num is a single inpatient admission. To count admissions, count each patid+conf_num.

To count days, use all 'F' rectypes within the admission, take the latest to_dt and subtract the earliest from_date. It is possible for there to be multiple records for the same stay with the same to_dt. For counting days, this is irrelevant, but it may matter when assigning a primary diagnosis to the stay.

To assign a diagnosis to the admission use the diag1 on the last F record. The last F record is the one with the latest to_dt. This is the presumed discharge diagnosis. As mentioned above, it is possible for there to be multiple records with the same to_dt, and, for the diag1 to be different on those records. This is a very small occurrence (less than 2%), so you may or may not elect to develop rules to 'pick' one.

To assign an admission as 'Surgical', look for the presence of a record type of 'S'.

To assign the admission to a specific surgery, (e.g., GI operation, you must define the qualifying codes, then look through all records for that admission, if any record with the code of interest is present in the admission, assign the admission to the surgery).

To determine total cost for the admission, sum all 'allowed' amounts.



OUTPATIENT CLAIMS

Emergency Room

Because we first segment inpatient records from outpatient records, emergency room visits that result in an inpatient stay are retained with the admission. If you want to segment them as emergency room visits, use this logic to pull out ER claims before segmenting inpatient from outpatient.

If any record meets the following criteria, we count all records for the same patient on the same day as the emergency room visit (event). Criteria are:

(proccode = 450,452, 459) or (proccode = 99281 - 99288) OR (pos 23 AND proccode = 10040-69979)

To count an emergency room visit, count as one each patient/day in which any record exists qualifying for emergency room.

The total costs are the sum of all allowed amounts for the patient/day records.

Laboratory/Pathology

Define as: (pos 81) or (proccode 80002-90042) or (proccode 300-319, 921,971)

Count each record as one

Use allowed amount on each record

Radiology

Radiology is more complicated because it is comprised of two components - professional and technical. In claims, sometimes the two components are on two separate records, sometimes they are combined into one. When on two separate records, they may both have exactly the same procedure code, or they may not. The approach defined here is our best effort at dealing with all the variants of billing style and data capture across health plans:

Using the CPT code range 70010-79999;

Count any occurrence of 'patid + from_date+proccode', as one and sum the allowed amount.



Outpatient Surgical or Diagnostic Procedures

This method is intended to be used for any treatment/procedure that may include a number of records all related. For example, any of the 'scope' procedures, neurological tests such as EMG's or electrophysiological procedures/studies. First you must define the codes that will comprise the procedure. Then, categorize all records on that date as associated with that event.

Count each patid+date on the date on which the patient has any record that is on the qualifying procedure list.

Sum the allowed amount for all records on the day.

The logic is that any number of different procedure codes may appear on the day of the procedure. There can be so many and with such variety that it can be impossible to define them all. But, one can be reasonably certain that if the patient is having a scope on a day, in all likelihood all claims are related to that event.

Office Visits

You may elect to define the range(s) of Evaluation and Management codes, or you may specify record type = M and pos = 11 to determine professional office visits. The latter will definitely give you office visits, and exclude inpatient, home health and ambulatory surgery. But, please note that it may also exclude valid office visits with a 'bad' place of service.

Count all records with the same patid + provid+date as one visit, sum the allowed amount on all records in the visit.



5. Frequently Asked Questions

This section provides brief responses to many of the more common questions with references back to any more detailed sections throughout the document.

1. How do I count inpatient admissions?

A unique inpatient admission can be determined by using the 'pat_id' and the 'conf_num.' Each combination is a unique inpatient admission.

2. How do I determine the 'primary' diagnosis for an inpatient admission?

The 'diag1' on the last 'F' record within a confinement contains the discharge, or primary diagnosis for that admission.

3. How do I calculate Length of Stay (LOS) for inpatient admissions?

To determine length of stay for a particular confinement for a particular patient from the detail file, scan all records with a rectype=F with the particular confinement number for the patient and determine the earliest FROM_DT and the latest TO_DT. Length of stay is then equal to the latest TO_DT.

First output all 'F' recs with a confinement number to a dataset, which in this example is called 'confinements.' This dataset needs to contain pat_id, conf_num, from_dt, and to_dt. This bit of SAS code will produce a dataset called 'get_los,' which is a confinement-level dataset containing the variables pat_id, conf_num, and los.

```
proc sort data = confinements ;  
    by pat_id conf_num from_dt ;  
run ;  
  
data get_los (keep= pat_id conf_num los) ;  
    set confinements ;  
    by pat_id conf_num ;  
  
    retain first_day last_day ;  
  
    if first.conf_num then do ;  
        first_day = from_dt ;  
        last_day = to_dt ;  
    end ;  
  
    if to_dt gt last_day then last_day = to_dt ;  
  
    if last.conf_num then do ;
```



```
    los = last_day - first_day ;  
    output ;  
end ;  
run ;
```

Why are some charges with an inpatient place of service not assigned a confinement number?

4. How can I easily determine the date for a position within the Estring?

Variables adhering to this format are 252-character strings. Each position represents a single month, beginning in January 1995. As shown in this schematic, position 1-12 apply to 1995, positions 13-24 to 1996, etc.

String position for given month

| Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1997 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
| 1998 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 |
| 1999 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| 2000 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 |
| 2001 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 |
| 2002 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 |
| 2003 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 |
| 2004 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 |
| 2005 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 |
| 2006 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 |
| 2007 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 |
| 2008 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 |
| 2009 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 |
| ... | | | | | | | | | | | | |
| 2015 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 |



5. What methods can you recommend to review periods of enrollment and to determine if a participant was enrolled during a particular time period?

There are two primary methods recommended:

The Estring is ideally suited to test for continuous, uninterrupted enrollment over an extended time period. For example:

| Requirement | Implementation |
|---|--|
| "Patients with two years continuous enrollment will be selected." | Search estring for 24 consecutive X's at any point: has_enr = (index(estring, 'XXXXXXXXXXXXXXXXXXXXXXXXX') > 0) ; |
| "All patients must be continuously enrolled between July 2002 and June 2003." | Search that specific section of estring for 12 X's: has_enr = (substr(estring, 91, 12) = 'XXXXXXXXXXXX') ; A small fraction of patients passing the estring test may have incomplete enrollment in the first and/or last months of the study period. The most rigorous technique for applying a continuous enrollment criteria is to test both estring and mixed enrollment dates: has_enr = %(substr(estring, 91, 12) = 'XXXXXXXXXXXX') and (mxce_fst <= '1-JUL-2002'd) and (mxce_lst >= '30-JUN-2003'd) ; |

Using the Estring is not recommended for testing short periods of enrollment, such as one or two months. For example:

| Requirement | Incorrect |
|--|--|
| "Patients must have at least one month of eligibility for healthcare benefits in 2002" | Search that specific section of estring for one X: has_enr = (index(substr(estring, 85, 12), 'X')) > 0) ; Patients with as little as one day of eligibility pass. Better to test mixed enrollment date variables (see below). |

The second tactic or the so-called "mixed" method utilizes the mxce_fst and mxce_lst enrollment date variables. For true-enrollment plans, these dates are provided by the data contributor. For non-true enrollment plans, these dates are inferred from the earliest and latest claim service dates. Checking patient enrollment using the mixed dates is less rigorous than testing the estring. This is because some patients have gaps in their eligibility.



Whether the estring or mixed-dates method is the more appropriate tool for patient selection depends on the stated requirement and also study objectives:

- For short required enrollment (1 day to 3 months), use mixed dates.
- To maximize the number of eligible patients, use mixed dates.
- For multidimensional studies, the more rigorous estring method assures uninterrupted enrollment and therefore full claims sets for the study population.

Examples:

| Requirement | Implementation |
|---|---|
| "Patients must have at least one month of eligibility for healthcare benefits in 2002" | <pre>if missing(mxce_fst) or missing(mxce_lst) then has_enr = 0 ; else has_enroll = (mxce_fst <= '31-DEC-2002'd) and (mxce_lst >= '01-JAN-2002'd) and (mxce_lst - mxce_fst + 1 >= 30) ;</pre> <p>A minority of patients have missing mixed dates, so that should always be part of the test.</p> |
| "Patients who do not have continuous health plan enrollment for at least 12 months after the first observed prescription and 3 months previous to the first prescription will be excluded." | <pre>if mxce_fst = . or mxce_lst = . then delete ; else if (mxce_fst > index_dt - 90) or (mxce_lst < index_dt + 360) then delete ;</pre> |

6. How does IMS count patients?

Over the course of a requested time period, we count the unique Pat_Ids over that period. It should be noted that if counts were requested over multiple intervals, a single unique patient would be counted once in each of the intervals if they met the criteria for inclusion in each interval. (i.e. A count of patients with a particular condition for 2000 through 2001 will likely not equal the sum of the counts for 2000 and 2001 individually. A single patient could be counted once in each of the 2000 and 2001 individual year counts, but would only be counted once in the two year 2000-2001 period.).

7. What does "-no cpi-" in the SPEC field (Specialty) on my extract represent?

The grouping software makes a determination on the provider presumed to be ordering a set of services (drug, ancillary) for a given event. This is called the cluster provider. If the grouper can not make the determination, the drug or ancillary record has no cluster provider identified (no cpi).

8. What does the trailing off in my data in the last months of my extract represent?



Two factors cause this phenomenon in any extract cut through the current date. Our dataset is least complete in the three to four months just prior to the current date, because of the lag associated with claims data (As previously described in Section III). As we simply do not have a comparable number of individuals as we do in the periods preceding that time, it is expected the number of individuals meeting the criteria of the extract would be smaller. The second factor is that complete claims information for an individual may not be present in the weeks closest to current date. We believe this to be relatively minimal so long as only the time period in which the patient is eligible is being considered. Because we adjust the enrollment period to the valid period of the underlying plan minimizes the possibility of patients within a sample with incomplete.

9. Can you briefly explain the record type assignment process?

- Management Services (M) are provider charges related to the direct evaluation or management of a patient (i.e. an office visit).
- Surgical Services (S) are surgical charges (i.e. A surgeon's charge to perform an appendectomy).
- Facility Services (F) are the room and board charges of an institutional provider.
- Pharmacy Services (P) are the drug charges identified by an NDC code or a HCPCS code
- Ancillary Services (A) are charges incidental to the direct care of a patient (i.e. X-Ray, Labs). This is the broadest based category, with many of the more vague charges being assigned to this group.



6. Standard Field Values

This section contains the values for several variables that are standardized by IMS (e.g. specialty, provider type, etc.).

Plan Product Type Table (PRODTYPE)

| Plan Product Type Values | Description | Definition |
|---------------------------------|---------------------------------|---|
| D | Consumer Directed Health Care | A self-identified plan option, thought to be similar to a commercial plan, but with underlying benefit plan design changes. These options can show a high degree of variation, but are thought to be high-deductible/high co-insurance plans. |
| H | Health Maintenance Organization | Historically, HMO coverage is associated with lower premiums, lower patient contributions (through relatively modest co-payments and deductibles), tightest provider network management controls, most restrictions on choice of providers and the most restrictions on the ability to self-refer. Over time, however, many HMO plans have changed to offer more choice and greater access to providers (and higher patient contributions to pay for it). |
| I | Indemnity Plan | Indemnity insurance is traditional health insurance. Typically, indemnity insurance has the highest premiums and greatest patient choice of providers and very few restrictions on self-referral. Patients typically pay 20% of fees paid to providers while the plan typically pays the remaining 80%. Patients also pay deductibles, typically higher than HMO plans, on selected services such as Emergency Room visits and Inpatient Admissions. |
| S | Point Of Service | POS plans are a hybrid of HMO and Indemnity plans. If a patient stays within the provider network and does not self-refer, patient contributions are structured like an HMO (i.e. relatively modest co-payments and deductibles) but the patient may elect to self refer and pay standard 20% coinsurance rates and typically higher deductibles. |
| P | Preferred Provider Organization | PPO plans are a hybrid of managed-care-like plans and indemnity plans. Patient contributions are structured like an indemnity plan (i.e. patient typically pay 20% coinsurance rates and deductibles on selected services), but if the patient stays within the network, the fees paid to providers are discounted. Like traditional indemnity plans, there are very few restrictions on self-referral. |
| U | Other or Unknown | Product type is unknown. |

**Payer Type Table (PAYTYPE)**

| Payer Type Value | Description | Definition |
|-------------------------|---|---|
| C | Commercial Plan | Commercial plans are primarily employer based. The health plan assumes the risk of insuring the enrolled members. |
| K | State Children's Health Insurance Program (SCHIP) | A Medicaid variant, primarily earmarked for children. It has been segment as often this population exhibits less enrollment variation, but generally can be considered identical to the Medicaid category. |
| M | Medicaid | Medicaid, a state and federal health insurance program for qualifying low income individuals, contracts in some cases with private health insurers to manage the health care for Medicaid enrollees. The health plan assumes the financial risk of insuring the enrollees and typically manages the plan like an HMO. |
| R | Medicare Risk | Medicare, a federal health plan for senior individuals and individuals with selected disabilities, contracts in some cases with private health insurers to manage the health care for Medicare enrollees. The health plan assumes the financial risk of insuring the enrollees and typically manages the plan like an HMO. Medicare Risk plans typically cover more services, including drugs, than traditional Medicare insurance, although the choice of providers and access to providers is more limited than traditional Medicare insurance. |
| T | Medicare Cost | Medicare Cost (also known as Medicare Gap or Medicare Supplemental) insurance is insurance purchased by individuals to cover services not covered by traditional Medicare insurance. Patients submit claims to either Medicare or the Medicare Cost insurer depending on the services consumed. Only the data submitted to the Medicare Cost insurer is in the health plan database. Medicare Cost plans are structured similarly to indemnity plans. |
| S | Self-Insured | Self-insured plans are a subset of Commercial plans where the employer assumes the risk of insuring the population. Large employers typically self-insure while small employers typically do not. With the exception of who is at risk (i.e. the employer), self-insured plans are typically run like HMO, PPO, or POS plans. |
| U | Other or Unknown | Other or Unknown |



Caveats and Observations Regarding Product and Payer Types

While we expect, on average, the plans to adhere to the industry standard definitions, it should be noted that the product type and payer type definitions are self-reported by the plan contributors. It should also be noted that over time plan characteristics evolve. A good example of this is that HMO products now often allow greater choice of providers and greater access to providers (along with the higher co-pays and deductibles associated with paying for that choice) while non-HMO products are adopting many of the network management techniques once used mainly by HMO's (e.g. using provider contracts that specify incentives and/or penalties for meeting clinical goals and utilization & cost containment goals). In the health plan database we have typically found that HMO, PPO, and POS product types have similar cost and utilization trends as do the payer types of Commercial and Self-Insured.



Place of Service

| CODE | DESCRIPTION |
|-------------|--|
| 01 | Unspecified Inpatient code |
| 02 | Unspecified Outpatient Code |
| 11 | Office |
| 21 | Inpatient Hospital |
| 22 | Outpatient Hospital |
| 23 | Emergency Room-Hospital |
| 24 | Ambulatory Surgical Center |
| 25 | Birthing Center |
| 26 | Military Treatment Facility |
| 34 | Hospice |
| 41 | Ambulance-Land |
| 42 | Ambulance-Air or Water |
| 50 | Federally Qualified Health Center |
| 51 | Inpatient Psychiatric Hospital |
| 52 | Psych. Facility Partial Hospital |
| 53 | Community Mental Health Center |
| 62 | Comprehensive OP Rehab Facility |
| 61 | Comprehensive IP Rehab Facility |
| 65 | End Stage Renal Disease Treatment Facility |
| 71 | State or Local Pub Health Clinic |
| 72 | Rural Health Clinic |
| 81 | Independent Lab |
| 99 | Other Unlisted Facility |
| 00 | Pharmacy |
| XX | Other |
| ZZ | Unknown |



Provider Specialty Table

| Specialty Type Value | Description |
|-----------------------------|--|
| ALLERGY | Allergy and Immunology |
| ANESTH | Anesthesiology |
| ASC | Ambulatory Surgery Center |
| CARDIOL | Cardiology |
| CHIRO | Chiropractic |
| COLON_SR | Colon-Rectal Surgery |
| CT_SRG | Cardio-Thoracic Surgery |
| DENTIST | Dentist |
| DERMATOL | Dermatology |
| DME_HH | Durable Medical Equipment/Home Health |
| ENDOCRIN | Endocrinology |
| ENT | ENT |
| ER_PHYS | Emergency Medicine Physician |
| GASTRO | Gastroenterology |
| GENETICS | Medical Genetics |
| GERIATRC | Geriatrics |
| GP_FP | General Practice/Family Practice |
| HAND_SRG | Hand Surgery |
| HEMATOL | Hematology |
| HEM_ONC | Hematology/Oncology |
| HOSPITAL | Hospital |
| INF_DIS | Infectious Disease |
| INTERN | Internal Medicine |
| MHSA_FAC | Mental Health/Substance Abuse Facility |
| MIDWIFE | Midwife |
| NEONAT | Neonatology |
| NEPHROL | Nephrology |
| NEUR_SRG | Neurosurgery |
| NEUROL | Neurology |
| --no cpi-- | The record doesn't group to a cluster, therefore no specialty is assigned. |
| N/A | Not Available |
| NRS_ANES | Nurse Anesthetist |
| NRS_PRCT | Nurse Practitioner |
| OB_GYN | Obstetrics and Gynecology |
| OCC_THER | Occupational Therapy |
| OPHTHAL | Ophthalmology |
| OPTOMTRY | Optometry |
| ORTH_SRG | Orthopedic Surgery |
| ORTHO | Orthopedics |
| OSTOPATH | Osteopath |
| OTHER | Other |
| OTHR_FAC | Other Facility |
| OTHR_SPC | Other Specialty |
| OTHR_SRG | Other Surgeon |



| Specialty Type Value | Description |
|----------------------|---|
| PATHOL | Pathology |
| PED | Pediatrics |
| PHYS_AST | Physician Assistant |
| PHYS_MED | Physical Medicine and Rehabilitation |
| PLST_SRG | Plastic Surgery |
| PODIATRY | Podiatry |
| PSYCHIAT | Psychiatry |
| PSYCHOL | Psychology |
| PT | Physical Therapy |
| PULMONAR | Pulmonology |
| RADIOL | Radiology |
| RAD_ONC | Radiation Oncology |
| RHEUM | Rheumatology |
| RN | Registered Nurse |
| SOC_WORK | Social Work |
| SNF | Skilled Nursing Facility/Long Term Care |
| SURGERY | General Surgery |
| URG_FAC | Urgent Care Facility |
| URGENT | Urgent Care Medicine |
| UROLOGY | Urology |
| VAS_SRG | Vascular Surgery |



7. U.S. Health Claims Coding

CPT Procedure Codes

CPT codes (Current Procedural Terminology) are a proprietary code set developed and maintained by the American Medical Association to enumerate procedures and services performed by physicians. However, the use of CPT codes is not limited to physician services. CPT codes are also found on claims submitted by other health care providers.

Each CPT code designates a service provided. One visit to a physician may result in numerous services, with each of these services associated and billed under a 5-digit, numeric CPT code. The codes describe various services, including but not limited to Evaluation and Management ("E/M" codes or physician encounters), Anesthesia, Surgery, Radiology, Pathology, and Laboratory services. Format Example: ##### (i.e. 99350).

CPT Modifiers further describe the service performed, providing a means to report greater granularity or details surrounding the service. Some examples include if an assistant surgeon was utilized, if one or both knees were treated in a replacement surgery, or if a bunion was removed from the left or right foot.

HCPCS Codes

HCPCS (Healthcare Procedural Coding System) is an amalgam of three levels of codes that permit identification of the procedures, services, and supplies necessary to meet the operational needs of Medicare and Medicaid. HCPCS Level I codes are for services typically provided by physicians. HCPCS Level II codes are a collection of approximately 3,800 and growing additional codes that identify medical and surgical supplies, certain drugs, certain durable medical equipment, and some professional services generally provided by non-physicians. HCPCS Level III codes are "local" codes reserved for use by the individual state Medicaid carriers.

The format is always an alpha-character, followed by 4-digits. Also of note is that the alpha character often provides category information about the code (For example "D" indicates a dental service, "J" indicates an injectable drug, while "W"-"Z" are local codes.

Format Example: Alpha-#### (i.e. J3490).



ICD-9 Diagnosis Codes

Formally, ICD-9 CM Volume III, this is arguably the primary and most important of all the code sets within the health care arena. The set includes both diagnosis and procedure codes:

- The ICD-9 procedure codes are a tabular and alphabetical list of procedures and conditions that bears little resemblance to either HCPCS or CPT codes. The final code set rules under HIPAA required that organizations and individuals use different collections of code sets for procedures depending on the location where a procedure is performed. Generally, if a procedure is performed in the inpatient setting, the charges are billed on a UB-92 billing form and ICD-9 procedure codes are used. If the procedure is performed in any outpatient setting, a HCFA 1500 and CPT or HCPCS codes are used.
 - Generally Health Plans do not retain ICD-9 procedure codes in their data warehouses; consequently, this data element is generally not present in the submission data.
 - Virtually every ICD-9 Procedure code has a corresponding CPT code; therefore, the absence of the ICD-9 procedure code has no effect.
- ICD-9 Diagnosis codes are the dominant coding system for identification of a particular medical condition. With some exceptions, the format is at least three numeric characters, often followed by a decimal point and two additional numeric characters. The more characters, the more specific the particular diagnosis. Also, the conditions are hierarchical, as shown in the example below:
 - 300 Anxiety, Dissociative And Somatoform Disorders
 - **300.0** Anxiety States
 - **300.01** Panic Disorder Without Agoraphobia
 - **300.1** Dissociative, Conversion, and Factitious Disorders
 - **300.11** Conversion Disorder
 - **300.12** Dissociative Amnesia

The exceptions to the 5-character numeric format are the “V” and “E” codes.

- “V” codes are a compilation of codes loosely defined as non-sickness/illness conditions or medical history factors. Examples include:
 - V21.31 Low Birth Weight Status, Less Than 500 Grams
 - V31 Twin Birth, Mate Liveborn
 - V10.04 Personal History of Malignant Neoplasm of Stomach



- “E” codes are very specific codes that were designed to provide very specific circumstances surrounding an illness event. It should be noted that they are not often or reliably used in practice. Examples include:
 - E814 Motor Vehicle Traffic Accident Involving Collision with Pedestrian
 - E842.8 Accident to Unpowered Aircraft Injuring Ground Crew
 - E953.1 Suicide and Self-inflicted Injury by Suffocation By Plastic Bag

The ICD-9 system was originally called the Bertillon Classification System, and dates back to 1893. The acronym “ICD-9-CM” actually represents the 9th iteration of the International Classification of Diseases, Clinical Modification, and the system is maintained and updated in the United States by the Department of Health and Human Services and four contributing parties [American Hospital Association, the American Health Information Management Association, the Health Care Financing Authority (now CMS), and the National Center of Health Statistics.].

ICD-10 is the most current version of the ICD code set and was released in 1992. The World Health Organization maintains that (and future) versions for international statistical reporting of morbidity and mortality. We will continually monitor the Medicare rules regarding the adaptation to this code set for purposes of reimbursement.

Format Examples:

- ICD-9 Procedure code: ##.## (i.e. 57.34 Bladder Biopsy)
- ICD-09-CM diagnosis codes: ### through ###.##, or V##.## or E###.##
(See previous page for examples)



NDC Drug Identifier Codes

The NDC, or National Drug Classification, code was created and is maintained by the U.S. Food and Drug Administration, but a variant exists in common practice. Specifically, the official regulation NDCs are 10-character numeric codes, which are a compellation of three sub-codes. The three sub-codes identify the labeler, the product, and the package. However, under the 10-digit method the length of the sub-codes is variable (4-4-2 or 5-4-1 or 5-3-2). This makes for extremely changing analysis, and hence private industry has adopted a standard 11-digit format based on the 10-digit framework. In the 11-digit format, the structure is always 5-4-2, with a leading zero added to the appropriate sub-code bring the total number of digits to 11. IMS uses the 11-digit format, along with almost every U.S. pharmaceutical manufacturer, healthcare, pharmacy, and pharmacy benefit manager firm.

Format Example:

- Often ##### (i.e. 16806583490; occasionally 16806-5834-90)



8. Data Quality Overview

Data Processing and Quality Review

The IMS Integrated Outcomes Database consists of claims data from over 80 different managed care organizations and other like sources. It consists of more than 75 million patients and over seven billion claims lines. Each data set undergoes rigorous data quality review before it is included in the production database. The robustness of the production database enables us to establish data quality 'norms' that can then be applied to each health care claims data set.

Approach

Our goal is not to change data that is erroneous, but rather to fully understand and document variations. To that end, IMS retains all incoming data and appropriately 'flags' each record if it matches a pre-defined quality problem. Each study or norms product either includes or excludes data based upon the specific requirements of the end product.

Each data contributor is provided a 'Data Submission Requirements' documentation that outlines the required data variables for submission, the file layout and the standard variable definitions. Each contributor is asked to provide descriptions and definitions of the variables within their organizational context.

Because the data comes from a number of different sources, we focus on standardizing the variables so that comparable analyses can be performed. This is accomplished by cross-walking key variables to standard definitions, including provider type, provider specialty, type of service, product type, reimbursement type and payer type. The standard definitions are provided to each contributor, so that they can map their variables to the standards. Alternatively, we map the client variables to the list, with client input.

Intake and Grouping Quality Review

Data is submitted to IMS in a variety of media along with documentation defining the file layout and mappings of the contributor's data to the standard definitions. Upon receipt of the data, a thorough comparison of the actual data with the documentation is performed to ensure that the data is as described. This may result in requests for clarification or confirmation from the client, to ensure that the contributor's data is fully understood.

Before the data are grouped, evaluation of the distribution of values of all variables is performed. The results are compared to expected distributions based upon our production database norms. When variances are identified, further analysis is performed to determine the reason for the differences. In the event the variances cannot be resolved, the data set,



in its entirety, would not be included in the production database, and a resubmission may be requested.

Assuming the data set as a whole meets the criteria for inclusion in the production database, each record is evaluated for a number of processing “flags” that may result in individual record level flagging of data quality issues. The flags are classified as “Warnings” in which case the issue is notable, but not considered significant, or “Errors”, in which case the record is excluded from the grouping process. This method allows for flexibility around the research approach.

Post-Grouping Quality Review

The post grouping data quality review process employs a system of SAS programs that generate a significant volume of statistical and descriptive information about the contents of the health plan claims database, its standard variables, selected data relationships and the quality of observations.

As with the incoming raw data, the grouped data is compared to expected distributions based upon our production database norms. Unexplained variances or identified problems with the grouped data will result in the exclusion of the client’s data from the production database, while we work with the client to resolve issues, or generate a new data extract.



Database Metrics

In most cases, variance from a single metric norm will not result in a plan's data being rejected. Exceptions are:

- Too few patients with a pharmacy claim
- Too few pharmacy claims
- Too few facility records
- Length of stay for facility records all zero days

Otherwise, the decision regarding the inclusion of the data in production is based upon the over-all 'picture' of the data. It is acceptable if the data are within the norms for most the metrics.

Metrics

Normal Distribution

Unique Patient Counts

| | |
|--|----------------------|
| Member Sex | 40% male, 60% female |
| Patients with at least 1 Ancillary Record | 80-87% |
| Patients with at least 1 Management Record | 89-94% |
| Patients with at least 1 Pharmacy Record | >69% |
| Patients with at least 1 Surgical Record | 31-39% |
| Patients with at least 1 Facility Record | 9-13% |

Details

| | |
|--|-------|
| Length of Stay for Facility Records >1 day | >80% |
| Non Rx claims with blank procedure code | <5% |
| Non Rx claims with blank diagnosis code | <1% |
| Services: Patient | 38-56 |
| Invalid Patient Age (<0, >130) | <.01% |



Diagnosis, Procedure, NDC Codes

| | |
|------------------------|------|
| Diagnosis Code Invalid | <.5% |
| NDC Code Invalid | <.8% |

Grouping Metrics

| | |
|---------------|----------|
| Record Types | |
| Ancillary | 36-44% |
| Management | 21-26% |
| Pharmacy | 29-36% |
| Facility | .3-.5% |
| Surgical | 2.3-2.8% |
| Un-assignable | <2.2% |

Cost Metrics

| | |
|----------------|-------|
| Charge Amounts | |
| % Negative | <.01% |
| % Zero | <3% |

| | |
|------------------------------|-------|
| % Zero Charge by Record Type | |
| Ancillary | <2.7% |
| Facility | <1.4% |
| Management | <1.2% |
| Pharmacy | <.8% |
| Surgical | <1.4% |

| | |
|--------------|------|
| Paid Amounts | |
| % Negative | <.1% |
| % Zero | <25% |

| | |
|----------------------------|------|
| % Zero Paid by Record Type | |
| Ancillary | <32% |
| Facility | <25% |
| Management | <30% |
| Pharmacy | <21% |
| Surgical | <18% |



| EVALUATION OF PLAN'S INBOUND DATA, BEFORE STANDARDIZATION AND CONSOLIDATION | |
|--|---|
| METRIC | CRITERIA |
| Claim Distributions | Claims are evenly distributed across all months? If incoming data has multiple file types (e.g. medical, pharmacy, facility) all types have claims evenly distributed across all months? |
| Connecting Current Patient IDs With Previous Patient IDs | High/Low Thresholds |
| Connecting Current Provider data with previous Provider IDs | High/Low Thresholds |
| Provider ID Frequencies | High/Low Thresholds |
| Member ID Frequencies | High/Low Thresholds |
| Merger of Patient IDs with IDs in enrollment | High/Low Thresholds |
| Product Type Distribution | Are the product type percentages similar to last iteration? |
| Pay Type Distribution | Are the pay type percentages similar to last iteration? |
| Gender Distribution | Are the gender percentages similar to the last iteration? |
| Enrollment Begin Date | Are the enrollment begin dates similar to last iteration? |
| Enrollment End Date | Are the enrollment end dates similar to last iteration? |
| RX Benefit Distribution | Is this field populated with Y and N with similar percentages to last iteration? |
| New ID Growth | Is the growth in new ids similar to past iterations for this plan? |
| True Enrollment Begin / End Dates | Do enrollment dates match what is expected? |
| Total Number of Months Enrolled | Does the number of months enrolled make sense? |
| Basic Demographics - Year of Birth | Do the years of birth listed make sense? |
| Basic Demographics - Gender Distribution | Are the majority of values M or F? |
| Basic Demographics - State | Are the majority of values valid state abbreviations? |



| EVALUATION OF PLAN'S ENROLLMENT DATA AFTER STANDARDIZED AND CONSOLIDATION | |
|--|--|
| METRIC | CRITERIA |
| Comparison of Patient Claim Dates to Enrollment Dates | Majority of claims fall within enrollment begin and end dates |
| Comparison of Patient Enroll Dates | Enrollment end date is after enrollment begin date |
| Enrollee By Month Counts | Are the enrollees evenly distributed across all months? |
| Total Number of Months Enrolled | Number of patients enrolled by month with peaks at 12 month intervals and a bolus in the highest number of months, otherwise evenly distributed? |
| Enrollment Type Crosstabulation | High/Low Thresholds |
| Mean Age at Last Enrollment | High/Low Thresholds |
| Enrollment Gap Analysis | High/Low Thresholds |
| Enrollment Begin Date Pre-dates First Claim Date | High/Low Thresholds |
| Enrollment End Date Post-dates Last Claim Date | High/Low Thresholds |
| % of Patients with First Claim Within 30 Days of Enrollment | High/Low Thresholds |
| % of Patients with First Claim Within 90 Days of Enrollment | High/Low Thresholds |
| % Monthly Pay Type Missing | High/Low Thresholds |
| Missing Payer Type (months) | High/Low Thresholds |
| % Unknown Payer Type (months) | High/Low Thresholds |
| Payer Type Changes | High/Low Thresholds |
| Monthly Prod Type Missing | High/Low Thresholds |
| Missing Prod Type (months) | High/Low Thresholds |
| % Unknown Prod Type (months) | High/Low Thresholds |
| Prod Type Changes | High/Low Thresholds |
| % Ben_Rx Flag Missing | High/Low Thresholds |
| Missing Ben_Rx Type (months) | High/Low Thresholds |
| Gender distribution | Male and Female comprise 50% +/- 2% |
| Year of Birth | Enrollees evenly distributed between 1951- |



| EVALUATION OF PLAN'S ENROLLMENT DATA AFTER STANDARDIZED AND CONSOLIDATION | |
|--|---------------------|
| METRIC | CRITERIA |
| Distribution | 1991? |
| Gender | High/Low Thresholds |
| Paid Amount \$0 | High/Low Thresholds |
| Paid Amount Negative | High/Low Thresholds |
| Paid Amount Missing | High/Low Thresholds |
| Zero Paid by Record Type - Facility | High/Low Thresholds |
| Zero Paid by Record Type - Management | High/Low Thresholds |
| Zero Paid by Record Type - Ancillary | High/Low Thresholds |
| Zero Paid by Record Type - Surgery | High/Low Thresholds |
| Zero Paid by Record Type - Pharmacy | High/Low Thresholds |
| Allowed Amount \$0 | High/Low Thresholds |
| Allowed Amount Negative | High/Low Thresholds |
| Allowed Amount Missing | High/Low Thresholds |
| Zero Allowed by Record Type - Facility | High/Low Thresholds |
| Zero Allowed by Record Type - Management | High/Low Thresholds |
| Zero Allowed by Record Type - Ancillary | High/Low Thresholds |
| Zero Allowed by Record Type - Surgery | High/Low Thresholds |
| Zero Allowed by Record Type - Pharmacy | High/Low Thresholds |
| Charge Amount \$0 | High/Low Thresholds |
| Charge Amount Negative | High/Low Thresholds |
| Charge Amount Missing | High/Low Thresholds |
| Zero Charge by Record Type - Facility | High/Low Thresholds |
| Zero Charge by Record Type - Management | High/Low Thresholds |
| Zero Charge by Record Type - Ancillary | High/Low Thresholds |
| Zero Charge by Record Type - Surgery | High/Low Thresholds |
| Zero Charge by Record Type - Pharmacy | High/Low Thresholds |
| Length of Stay - % F | High/Low Thresholds |



| EVALUATION OF PLAN'S ENROLLMENT DATA AFTER STANDARDIZED AND CONSOLIDATION | |
|--|---------------------|
| METRIC | CRITERIA |
| Records > 1day | |
| Drug Quantity Dispensed - 30 days | High/Low Thresholds |
| Drug Quantity Dispensed - 60 Days | High/Low Thresholds |
| Drug Quantity Dispensed - 90 Days | High/Low Thresholds |
| Drug Quantity Dispensed - 120 days | High/Low Thresholds |
| Drug Days Supplied - 7 days | High/Low Thresholds |
| Drug Days Supplied - 30 Days | High/Low Thresholds |
| Drug Days Supplied - 60 Days | High/Low Thresholds |
| Drug Days Supplied - 90 days | High/Low Thresholds |
| Place of Service - 00 | High/Low Thresholds |
| Place of Service - 11 | High/Low Thresholds |
| Place of Service 12 | High/Low Thresholds |
| Place of Service - 21 | High/Low Thresholds |
| Place of Service - 22 | High/Low Thresholds |
| Place of Service - 23 | High/Low Thresholds |
| Place of Service - 81 | High/Low Thresholds |
| Place of Service - ZZ | High/Low Thresholds |
| Procedure Code - CPT4 | High/Low Thresholds |
| Procedure Code - HCPCS | High/Low Thresholds |
| Procedure Code - Revenue | High/Low Thresholds |
| Procedure Code - Invld or Blank | High/Low Thresholds |
| Diagnosis Code 1, Blank (non-pharmacy) | High/Low Thresholds |
| Diagnosis Code 2 | High/Low Thresholds |
| Diagnosis Code 3 | High/Low Thresholds |
| NDC, found | High/Low Thresholds |
| NDC, Invalid | High/Low Thresholds |
| Provider Type - Clinician | High/Low Thresholds |
| Provider Type - Other | High/Low Thresholds |
| Provider Type - Facility | High/Low Thresholds |
| Plan Type, Blank/Invalid | High/Low Thresholds |
| Payer Type, Blank/Invalid | High/Low Thresholds |



| EVALUATION OF PLAN'S ENROLLMENT DATA AFTER STANDARDIZED AND CONSOLIDATION | |
|--|---------------------|
| METRIC | CRITERIA |
| Record Type, Ancillary | High/Low Thresholds |
| Record Type, Pharmacy | High/Low Thresholds |
| Record Type, Management | High/Low Thresholds |
| Record Type, Surgery | High/Low Thresholds |
| Record Type, Facility | High/Low Thresholds |
| Record Type, Ungroupable | High/Low Thresholds |
| Unique Patient by Record Type Count, Ancillary | High/Low Thresholds |
| Unique Patient by Record Type Count, Pharmacy | High/Low Thresholds |
| Unique Patient by Record Type Count, Management | High/Low Thresholds |
| Unique Patient by Record Type Count, Surgery | High/Low Thresholds |
| Unique Patient by Record Type Count, Facility | High/Low Thresholds |
| Unique Patient by Record Type Count, Ungroupable | High/Low Thresholds |
| Member Age - Blank/Invalid | High/Low Thresholds |
| Specialty, unknown, invalid | High/Low Thresholds |
| Specialty, Internal Medicine | High/Low Thresholds |
| Specialty, General Pracice | High/Low Thresholds |
| Specialty, Pediatrics | High/Low Thresholds |
| Specialty, Ob/Gyn | High/Low Thresholds |
| Specialty, Cardiology | High/Low Thresholds |
| Specialty, Hematology/Oncology | High/Low Thresholds |
| Distribution of All Record Types by Month | High/Low Thresholds |
| Distribution of Ancillary Records by Day | High/Low Thresholds |
| Distribution of Facility Records by Day | High/Low Thresholds |
| Distribution of Management Records | High/Low Thresholds |



| EVALUATION OF PLAN'S ENROLLMENT DATA AFTER STANDARDIZED AND CONSOLIDATION | |
|--|---------------------|
| METRIC | CRITERIA |
| by Day | |
| Distribution of Pharmacy Records by Day | High/Low Thresholds |
| Distribution of Surgical Records by Day | High/Low Thresholds |
| Relationship of Management Record Charge/Allowed/Paid amounts is valid | High/Low Thresholds |
| Relationship of Pharmacy Record Charge/Allowed/Paid amounts is valid | High/Low Thresholds |
| Relationship of Surgical Record Charge/Allowed/Paid amounts is valid | High/Low Thresholds |
| Relationship of Confinement Charge/Allowed/Paid amounts is valid | High/Low Thresholds |
| Ratio of Enrollees Under Age 65 with Any Claim by Month | High/Low Thresholds |
| Ratio of Enrollees Under Age 65 with Pharmacy Claim by Month | High/Low Thresholds |
| Ratio of Enrollees Over Age 65 with Any Claim by Month | High/Low Thresholds |
| Ratio of Enrollees Over Age 65 with Pharmacy Claim by Month | High/Low Thresholds |