COMMERCIAL CLAIMS AND ENCOUNTERS
MEDICARE SUPPLEMENTAL
DATA YEAR 2016 EDITION

The authorized recipient of these materials shall treat the information contained therein as confidential proprietary information owned by Truven Health Analytics Inc. The recipient shall not disclose or permit to be disclosed, in full or in part, to third parties any information contained therein. No part of these materials may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from Truven Health.

Requests for permission to make copies of any part of this report should be mailed to:

Truven Health Analytics
100 Phoenix Drive
Ann Arbor, Michigan 48108

The file should be cited as follows:
SOURCE: Truven Health MarketScan® Research Databases.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>OVERVIEW OF TABLES</td>
<td>4</td>
</tr>
<tr>
<td>OVERVIEW OF ENCOUNTER RECORDS</td>
<td>8</td>
</tr>
<tr>
<td>FINANCIAL VARIABLES</td>
<td>9</td>
</tr>
<tr>
<td>PERSON-LEVEL IDENTIFIERS</td>
<td>15</td>
</tr>
<tr>
<td>CLINICAL VARIABLES</td>
<td>17</td>
</tr>
<tr>
<td>MARKETSCAN DATABASE CONSTRUCTION</td>
<td>19</td>
</tr>
<tr>
<td>PLAN TYPE DEFINITIONS</td>
<td>21</td>
</tr>
<tr>
<td>KEY TABLE AND FIELD RELATIONSHIPS</td>
<td>23</td>
</tr>
<tr>
<td>GLOSSARY OF ACRONYMS, ABBREVIATIONS, AND TERMS</td>
<td>24</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>31</td>
</tr>
<tr>
<td>FREQUENTLY ASKED QUESTIONS</td>
<td>61</td>
</tr>
<tr>
<td>APPENDIX A. NEW IN 2016</td>
<td>66</td>
</tr>
<tr>
<td>APPENDIX B. HISTORICAL DATA RELEASES</td>
<td>67</td>
</tr>
</tbody>
</table>
INTRODUCTION

TRUVEN HEALTH MARKETSCAN DATABASE OVERVIEW

The Truven Health MarketScan® Research Databases capture person-specific clinical utilization, expenditures, and enrollment across inpatient, outpatient, prescription drug, and carve-out services. The data come from a selection of large employers, health plans, and government and public organizations. The MarketScan Research Databases link paid claims and encounter data to detailed patient information across sites and types of providers and over time. The annual medical databases include private-sector health data from approximately 350 payers. Historically, more than 20 billion service records are available in the MarketScan databases. These data represent the medical experience of insured employees and their dependents for active employees, early retirees, Consolidated Omnibus Budget Reconciliation Act (COBRA) continuees, and Medicare-eligible retirees with employer-provided Medicare Supplemental plans.

The Truven Health MarketScan Research Databases are composed of the following individual databases (also see Exhibit 1):

Commercial Claims and Encounters Database

The MarketScan Commercial Claims and Encounters Database contains data from active employees, early retirees, COBRA continuees, and dependents insured by employer-sponsored plans (i.e., persons not eligible for Medicare).

The database has the following table structure:
- Inpatient Admissions Table (I)
- Facility Header Table (F)
- Inpatient Services Table (S)
- Outpatient Services Table (O)
- Outpatient Pharmaceutical Claims Table (D)
- Annual Enrollment Summary Table (A)
- Enrollment Detail Table (T)

Medicare Supplemental

The MarketScan Medicare Supplemental and Coordination of Benefits (COB) Database is created for Medicare-eligible retirees with employer-sponsored Medicare Supplemental plans. This database contains predominantly fee-for-service plan data.

The Medicare Supplemental Database table structure is identical to the Commercial Claims and Encounters table structure.

Both the Medicare-paid amounts and the employer-paid supplemental insurance amounts are included in this database. Only plans where both the Medicare-paid amounts and the employer-paid amounts were available and evident on the claims were selected for this database.

Health and Productivity Management Database

The MarketScan Health and Productivity Management (HPM) Database is an integrated database that contains absence, short-term disability, long-term disability, and worker’s compensation experience. This information is linkable to the medical, pharmacy, and enrollment data in the MarketScan Commercial Claims and Encounters Database for these employees, making the resulting database a unique and valuable resource for examining health and productivity issues for an employed, privately insured population.

A separate User Guide is provided to customers licensing the HPM database.

Benefit Plan Design Database

The MarketScan Benefit Plan Design (BPD) Database consists of data for selected benefit plans represented in the MarketScan Research Databases from 1995 forward. A separate User Guide is provided to customers licensing the BPD Database. Benefit plan design information is available for Commercial Claims and Encounters and Medicare Supplemental Databases.
**Medicaid Database**

The MarketScan Medicaid Database contains the pooled healthcare experience of approximately seven million Medicaid enrollees from multiple States. It includes inpatient services and prescription drug claims, as well as information on enrollment, long-term care, and other medical care. In addition to standard demographic variables such as age and sex, the database includes variables of particular value to researchers investigating Medicaid populations (e.g., race/ethnicity, maintenance assistance status, Medicare eligibility).

**MarketScan Lab**

The MarketScan Lab Database contains the pooled healthcare experience of over 1 million covered lives, gleaned from sources that include both commercial and Medicare Supplemental coverage. It captures laboratory tests for a subset of the covered lives and mainly represents lab tests ordered in office-based practice. Linkage of lab results to claims supports analyses that are not feasible with claims alone, such as determining effectiveness of treatment, measuring severity of illness, identifying patients for whom treatment may be indicated, and verifying diagnoses recorded on claims.

Note: This User Guide is intended to cover the Commercial Claims and Encounters Database and the Medicare Supplemental Database. The data you receive may contain some or all of the MarketScan data described herein.

**Exhibit 1. Overview of the MarketScan Research Databases.**

<table>
<thead>
<tr>
<th>Database</th>
<th>Content</th>
<th>Covered Lives</th>
<th>Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial Claims and Encounters</td>
<td>Healthcare coverage eligibility and service use of individuals in plans or product lines with fee-for-service plans and fully capitated or partially capitated plans</td>
<td>Active employees and dependents, early (non-Medicare) retirees and dependents, COBRA continues</td>
<td>Medical/Surgical: Inpatient Admissions (I) Facility Header (F) Inpatient Services (S) Outpatient Services (O) Prescription Drug (D) Enrollment (A,T)</td>
</tr>
<tr>
<td>Medicare Supplemental and Coordination of Benefits (COB)</td>
<td>Healthcare coverage eligibility and service use of individuals in plans or product lines with fee-for-service plans and fully capitated or partially capitated plans</td>
<td>Medicare-eligible active and retired employees and their Medicare-eligible dependents from employer-sponsored supplemental plans</td>
<td>Medical/Surgical: Inpatient Admissions (I) Facility Header (F) Inpatient Services (S) Outpatient Services (O) Prescription Drug (D) Enrollment (A,T)</td>
</tr>
<tr>
<td>Benefit Plan Design (BPD)</td>
<td>Plan characteristics derived from the medical claims submitted by each plan. Additional information specific to each plan is available in the BPD User Guide.</td>
<td>Not Applicable</td>
<td>Links to Commercial Claims and Encounters and Medicare Supplemental and COB Databases for a subset of plans included in those databases</td>
</tr>
<tr>
<td>Database</td>
<td>Content</td>
<td>Covered Lives</td>
<td>Tables</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Health and Productivity Management</td>
<td>Absence, short-term disability, long term disability and worker’s compensation experience for a subset of the covered lives represented in the Commercial Claims and Encounters Database</td>
<td>Active employees</td>
<td>Absenteeism (ABS)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Short Term Disability (STD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Long Term Disability (LTD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Workers’ Compensation (WC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Eligibility (E)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Linkable to the medical and prescription drug claims information appearing in the Commercial Claims and Encounters Database</td>
</tr>
<tr>
<td>Medicaid</td>
<td>Healthcare coverage eligibility and service use of individuals enrolled in State Medicaid programs for several States and/or Medicaid Managed Care programs</td>
<td>Medicaid recipients for several States</td>
<td>Medical/Surgical:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inpatient Admissions (I)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Facility Header (F)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inpatient Services (S)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Outpatient Services (O)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Long Term Care (L)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Prescription Drug (D)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Enrollment (A,T)</td>
</tr>
<tr>
<td>Lab</td>
<td>Healthcare service use and eligibility for individuals enrolled in commercial and Medicare Supplemental programs, along with laboratory test records and results</td>
<td>Individuals enrolled in commercial and Medicare Supplemental programs</td>
<td>Medical/Surgical:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inpatient Admissions (I)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Facility Header (F)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inpatient Services (S)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Outpatient Services (O)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Prescription Drug (D)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Enrollment (A,T)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lab Test Results (R)</td>
</tr>
</tbody>
</table>
OVERVIEW OF TABLES

Note: All of the tables and databases described below are available in both the Commercial Claims and Encounters Database and the Medicare Supplemental and Coordination of Benefits Database. Exhibit 2 contains the data flow diagram.

Medical/Surgical Tables

The MarketScan databases contain inpatient and outpatient medical/surgical data stored in four tables: Inpatient Admissions, Inpatient Services, Facility Header, and Outpatient Services.

Inpatient Admissions Table (I)

The Inpatient Admissions Table contains records that summarize information about a hospital admission. Truven Health constructs this table after identifying all of the encounters or claims (service records) associated with an admission (e.g., hospital claims, physician claims, surgeon claims, and claims from independent laboratories). Facility and professional payment information is then summarized for all services. The summarized information is stored in an admission record in the Inpatient Admissions Table. Please refer to Section 3: Financial Variables for definitions of key financial variables.

The admission record also includes data that can only be identified after all claims for an admission have been identified. These additional data include the principal procedure, principal diagnosis, major diagnostic category (MDC), and diagnosis-related group (DRG). Truven Health uses the Centers for Medicare & Medicaid Services (CMS) DRG Grouper to assign an MDC and DRG to the admission record.

In addition to the principal procedure and diagnosis codes, the admission record includes all diagnoses and procedures (up to 14 each) found on the service records that make up the admission. These additional codes (Diagnosis 2 through Diagnosis 15 and Procedure 2 through Procedure 15) are assigned chronologically based on service dates and do not duplicate the principal code.

To be considered an admission, the grouping of these service records must meet certain criteria (e.g., a room and board claim must be present). If these criteria are not met, the records are stored in the Outpatient Services Table (O) and no admission record is created.

Facility Header Table (F)

The Facility Header Table contains complete header information from facility claims. A Facility Header Record identifier (FACHDID) exists on both the Facility Header Table and the Inpatient Services and Outpatient Claims Tables to identify the individual service records that each header record comprises.

Facility inpatient service records are derived from the Uniform Billing (UB04) form. This form does not link financial information to specific procedures or diagnoses.

Inpatient Services Table (S)

The Inpatient Services Table contains the individual facility and professional encounters and services that the inpatient admission record comprises. A Cases and Services Link (CASEID) identifier exists on both the Inpatient Admissions and the Inpatient Services Tables to identify the individual service records that each admission record comprises.

Facility inpatient service records are derived from the Uniform Billing (UB04) form. This form does not link financial information to specific procedures or diagnoses. Physician services are derived from the Centers for Medicare & Medicaid Services (CMS) 1500 form.

Note: The Inpatient Services Table contains both facility and physician services associated with an inpatient admission. The Inpatient Admissions Table differs from UB04 discharge data in that Truven Health combines the facility charges with the physician services associated with an inpatient admission. UB04 revenue codes are retained in the MarketScan data when available; however, not all
data contributors provide the codes on adjudicated claims.

**Outpatient Services Table (O)**

The Outpatient Services Table contains encounters and claims for services that were rendered in a doctor’s office, hospital outpatient facility, emergency room, or other outpatient facility. A small percentage of claims in this table may represent inpatient services, because the claim was not incorporated into an inpatient admission (e.g., no room and board charge was found). These claims generally have an “inpatient” Place of Service (STDPLAC) code.

**Outpatient Pharmaceutical Claims Table (D)**

Outpatient pharmaceutical claims data are available for a large portion of the individuals represented in the medical/surgical and populations tables. The outpatient pharmaceutical data are linked by ENROLID to the medical/surgical data. Each record represents either a mail-order or card program prescription drug claim.

Note: Before you begin your analysis, carefully determine which data sources (e.g., medical/surgical, outpatient pharmaceutical, enrollment) will be necessary to support your analytic plan. If you require more than one of these data sources, it first may be necessary to utilize the various cohort flags to determine which data contributors or plans have the required data (via RX, MHSACOVG, and/or EIDFLAG/ENRFLAG variables).

**Enrollment Tables (A, T)**

The Enrollment Tables contain person-level enrollment records with demographic and plan information on users and non-users of services contained in the MarketScan Commercial Claims and Encounters and Medicare Supplemental Databases.

The Annual Enrollment Summary Table contains a single record per-person, per-year. The annual summary contains monthly arrays of certain variables such as indicators of enrollment (yes/no), days enrolled, data type, and plan type in each month during the year. There are also variables indicating the number of months during the year with enrollment and the total annual enrollment days.

The Enrollment Detail Table contains one record per person per month of enrollment for an individual enrollee regardless of whether any demographic values have changed from the previous month.

If you need to track changes in variables such as Cohort Drug Indicator (RX) or Geographic Location of Employee (EGEOLOC), use the Enrollment Detail Table.

Beginning with the 2001 data, all data contributors submit person-level enrollment information. When using MarketScan database releases prior to 2001, the Enrollment Flag (ENRFLAG) variable allows the user to select only claims supported by person-level enrollment. When ENRFLAG=1, it indicates that person-level enrollment information is available for that data contributor.

**Records Where ENROLID Is Missing**

There may be records where ENRFLAG=1 but the Enrollee ID (ENROLID) is missing. This occurs in less than 1% of records. Individual claim records from a data contributor may not have these identifiers assigned if certain key variables are missing (see Section 4: Person-Level Identifiers). These records may be excluded from analysis, depending upon the needs of your study.

**Member Days (MEMDAYS)**

When obtaining an underlying population or covered life count, evaluate the Date Enrollment Start (DTSTART) and Date Enrollment End (DTEND) data before summing Member Days (MEMDAYS). If a time-based subset or study period is required, the DTSTART and DTEND may be outside the beginning and ending dates of the subset criteria. If so, adjust the DTSTART and DTEND to match the study period and recalculate
the member days before calculating an enrollee count.

For example, a record may have DTSTART and DTEND of 1/1/2013 and 1/31/2013, respectively. The MEMDAYS variable on this record is 31 days.

If the study period of data begins on 1/15/2013, the DTSTART should be reset to reflect the 1/15/2013 beginning date and MEMDAYS should be recalculated to 16 days (MEMDAYS = DTEND – DTSTART + 1).
Exhibit 2. Data flow diagram.
OVERVIEW OF ENCOUNTER RECORDS

Encounter records represent the service use and cost of individuals in partially and fully capitated plans and allow for the empirical investigation of healthcare under a variety of managed care arrangements.

Historically, not all fully or partially capitated health plans have maintained rigorous cost and utilization data collection systems. Many managed care services are pre-paid in fixed sums for each member, which minimizes the need for administrative systems to collect financial encounter information at the time of service delivery. Therefore, unlike indemnity plans (which adjudicate claims for reimbursement), certain types of managed care plans do not process claims for the purpose of financial reporting. For these plans, service delivery information is disconnected from charge and payment information. Instead of generating a claim for reimbursement of pre-paid capitated services, a managed care plan generates an encounter record.

An encounter record provides demographic information about the patient, provider characteristics, and diagnosis and procedure codes, but in many instances it provides only limited financial information. This presents a certain challenge when analyzing healthcare costs using encounter records.

The challenge involves the correct measurement of reimbursement for capitated managed care plans. Many encounter records contain a payment (PAY) amount of $1 or $0 for capitated services. The pre-paid capitation amounts, whether in the form of per member per month fees or bulk capitation payments, were not contributed by the managed care plans represented in this database. However, managed care plans are beginning to enhance encounter records with fee-for-service-equivalent financials. Fee-for-service-equivalent financial amounts are intended to be approximate reasonable and customary charges or payments for medical services or procedures. See the Financial Variables section of this User Guide for other important information.

The implementation of fee-for-service-equivalent financial amounts is in its early stages; as a result, financial variables are potentially understated. Financial measures derived from encounter records should be interpreted with caution, with the exception of Copayment (COPAY), Deductible (DEDUCT), and Coordination of Benefits and Other Savings (COB) amounts—all of which are recorded with reasonable accuracy.

In constructing the MarketScan Research Databases, encounter records are rigorously tested by overall plan-by-plan utilization rates to ensure that plans appearing to submit incomplete data are excluded.
FINANCIAL VARIABLES

Truven Health receives paid claims from approximately 350 data sources. Financial variables are consistently defined across all data contributors. Exhibit 3 contains an example of a financial variable calculation.

The definitions in Exhibit 4 apply to all MarketScan Research Databases. The definitions apply to the capitated encounter data, even though some of the financial variables are set to zero (0) or one (1) because encounter records may not contain fee-for-service charge and payment equivalents.

In order to protect business-confidential discount arrangements between our data contributors and their providers, information on submitted charges and allowed amounts are never licensed simultaneously on the same MarketScan dataset.

Exhibit 3. Truven Health Financial Variable Calculation: Example

<table>
<thead>
<tr>
<th>Charge Type</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submitted Charges</td>
<td>$1,200.00</td>
</tr>
<tr>
<td>Not Covered Charges</td>
<td>-$100.00</td>
</tr>
<tr>
<td>Eligible Charges</td>
<td>$1,100.00</td>
</tr>
<tr>
<td>Pricing Reductions</td>
<td>-$100.00</td>
</tr>
</tbody>
</table>

(Amounts above are not standard MarketScan variables)

<table>
<thead>
<tr>
<th>Description</th>
<th>Data Element</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross Covered Payments</td>
<td>Gross Covered Payments (PAY)</td>
<td>$1,000.00</td>
</tr>
<tr>
<td>Remaining Deductible</td>
<td>Deductible (DEDUCT)</td>
<td>-$100.00</td>
</tr>
<tr>
<td>Coinsurance at 20%</td>
<td>Coinsurance (COINS)</td>
<td>-$180.00</td>
</tr>
<tr>
<td>Penalty for no Pre-cert</td>
<td>Coordination of Benefits and Other Savings (COB)</td>
<td>-$270.00</td>
</tr>
<tr>
<td>Net Payments</td>
<td>Net Payments (NETPAY)</td>
<td>$450.00</td>
</tr>
</tbody>
</table>
Medical/Surgical Financials

The following abbreviations indicate the tables on which the variable resides:

- I – Inpatient Admissions
- F – Facility Header
- S – Inpatient Services
- – Outpatient Services
- D – Outpatient Pharmaceutical Claims

Exhibit 4. Definitions of Variables

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>MarketScan Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Payment</td>
<td>Total gross payment to all providers associated with the admission</td>
<td>TOTPAY (Payments Total Case) (I)</td>
</tr>
<tr>
<td></td>
<td>Format: Dollars and cents</td>
<td></td>
</tr>
<tr>
<td>Payment</td>
<td>Total gross payment to a provider for a specific service; i.e., the amount eligible for payment after applying pricing guidelines such as fee schedules and discounts, and before applying deductibles, copayments, and COB.</td>
<td>PAY (Payment) (S,O,D)</td>
</tr>
<tr>
<td></td>
<td>Format: Dollars and cents</td>
<td></td>
</tr>
<tr>
<td>Deductible</td>
<td>Amount of Gross Covered Payments applied toward the deductible.</td>
<td>TOTDED (Deductible Total Case) (I)</td>
</tr>
<tr>
<td></td>
<td>Format: Dollars and cents</td>
<td>DEDUCT (Deductible) (F,S,O,D)</td>
</tr>
<tr>
<td>Coinsurance/Copayment</td>
<td>Amount of Coinsurance applied toward the stop loss and/or amount of copayment.</td>
<td>TOTCOPAY (Copayment Total Case) (I)</td>
</tr>
<tr>
<td></td>
<td>Format: Dollars and cents</td>
<td>TOTCOINS (Coinsurance Total Case) (I)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COPAY (Copayment) (F,S,O,D)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COINS (Coinsurance) (F,S,O,D)</td>
</tr>
<tr>
<td>Net Payment</td>
<td>Payment received by the provider excluding patient out-of-pocket and coordination of benefits; i.e., employer or plan liability.</td>
<td>NETPAY (Payments Net) (F,S,O,D)</td>
</tr>
<tr>
<td></td>
<td>Format: Dollars and cents</td>
<td></td>
</tr>
<tr>
<td>Total Net Payment</td>
<td>Total net payment to all providers associated with the admission; i.e., sum of service-level net payments.</td>
<td>TOTNET (Payments Net Case) (I)</td>
</tr>
<tr>
<td></td>
<td>Format: Dollars and cents</td>
<td></td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
<td>MarketScan Variable</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Hospital Payments</td>
<td>Total gross payments to the hospital for an admission. Format: Dollars and cents</td>
<td>HOSPPAY (Payments Hospital) (I)</td>
</tr>
<tr>
<td>Physician Payment</td>
<td>Total gross payments to the principal physician (the professional who charges the most during the admission). Format: Dollars and cents NOTE: Payments to physicians other than the principal physician are included in TOTPAY (Payments Total Case).</td>
<td>PHYSPAY (Payments Physician) (I)</td>
</tr>
<tr>
<td>Hospital Net Payment</td>
<td>Payment received by the hospital for an admission excluding patient out-of-pocket and coordination of benefits; i.e., employer or plan liability. Format: Dollars and cents</td>
<td>HOSPNET (Net Payment Hospital) (I)</td>
</tr>
<tr>
<td>Physician Net Payment</td>
<td>Payment received by the principal physician (the professional who charges the most during the admission) excluding patient out-of-pocket and coordination of benefits; i.e., employer or plan liability. Format: Dollars and cents</td>
<td>PHYSNET (Net Payment Physician) (I)</td>
</tr>
<tr>
<td>Third Party Payment</td>
<td>Payment received by the provider from a source other than the patient or the submitting plan. Format: Dollars and cents</td>
<td>TOTCOB (COB and Other Savings Total Case) (I) COB (COB and Other Savings) (F,S,O,D)</td>
</tr>
</tbody>
</table>

**Prescription Drug Financial Variables**

The Outpatient Pharmaceutical Claims Table contains the Payment (PAY), Copayment (COPAY), Coinsurance (COINS), Deductible (DEDUCT), and Coordination of Benefits and Other Savings (COB) variables, as previously described. Financial variables specific to prescription drug claims are provided in Exhibit 5.
### Exhibit 5. Key Data Elements: Outpatient Pharmaceutical Financials

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>MarketScan Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Wholesale Price*</td>
<td>The average wholesale price charged by wholesalers for the specific drug. Format: Dollars and cents</td>
<td>AWP (Average Wholesale Price) (D)</td>
</tr>
<tr>
<td>Administrative Dispensing Fee</td>
<td>Administrative fee charged by the pharmacy for dispensing the prescription. Format: Dollars and cents</td>
<td>DISPFEE (Dispensing Fee) (D)</td>
</tr>
<tr>
<td>Ingredient Cost</td>
<td>The cost or charge associated with the pharmaceutical product. The Ingredient Cost plus the Dispensing Fee and Sales Tax, if applicable, usually represents the entire cost of a prescription. Format: Dollars and cents</td>
<td>INGCOST (Ingredient Cost) (D)</td>
</tr>
<tr>
<td>Sales Tax</td>
<td>The amount of sales tax applied to the cost of the prescription. (The sales tax, if applicable, is usually calculated based on the Ingredient Cost plus the Dispensing Fee.) Format: Dollars and cents</td>
<td>SALETAX (Sales Tax) (D)</td>
</tr>
</tbody>
</table>

*The Red Book® Systems Licensed Content may only be used as a referential look-up tool, and not for an automated claims processing system. For Red Book® System Licensed Content only. The prices contained in Truven Health RED BOOK™ are based on data reported by manufacturers. Truven Health Micromedex® has not performed an independent analysis of the actual prices paid by wholesalers and providers in the marketplace. Thus, actual prices paid by wholesalers and providers may vary from the prices contained in this database, and all prices are subject to change without notice. Further, Truven Health Micromedex does not warrant the accuracy of the database contents or the pricing information. Please refer to the "Average Wholesale Price Policy" in the RED BOOK product for more information.
**Encounter Record Financial Variables**

Financial information is captured in a variety of ways for encounter claims. There may be times when a capitated claim may have financial variables with amounts of zero because there is no associated paid claim. At other times, the copayment amount may be the only financial information on the claim. If a capitated claim does not include financial information, the financial variables are set to “0” or “1.”

**Medicare Supplemental and COB Financial Variables**

Medicare Supplemental claim records include paid claims for fee-for-service plans and contain all of the Payment (PAY), Deductible (DEDUCT), Copayment (COPAY), Coinsurance (COINS), Coordination of Benefits and Other Savings (COB) and NETPAY (Payments Net) variables, as previously described. In 1998, Medicare Supplemental encounter records were added to the Medicare Supplemental Database (please refer to the Encounter Record Financial Variables section). The Medicare paid amount is reflected in the COB variable, so the majority of the breakdown of PAY will be captured in COB for the medical claims. The Medicare supplemental payments made by the employer will be captured in the NETPAY variable.

Because outpatient prescription drugs are generally covered by the employer rather than Medicare, the majority of PAY will be captured in the NETPAY variable for outpatient pharmaceutical claims in the Medicare and COB Database.

**Adjustment Records**

Some claims have negative amounts in payment or other financial variables. These are adjustment records entered by claims processors to correct a payment error or any type of coding error.

Resolution of adjustments combines the financials on the original record with the financials on the adjustment. No information is lost when resolving adjustment records. The sum total of financials remains the same. However, instead of reading across multiple records to understand the services rendered, resolution of adjustments creates a single service-level record. Adjustment records are resolved on both the Outpatient Services Table and the Outpatient Pharmaceutical Claims Table. Adjustment records are not resolved on the Inpatient Services Table.

There are two methods claims processors typically use for entering adjustment records: the adjustment method and the void and replace method.

The Adjustment Method allows the entry of a new claim that exactly duplicates all of the correct variables on the erroneous claim, including the date of service. If the financial information is incorrect, an adjusted dollar amount is entered in the appropriate financial variable(s) (e.g., PAY), and all of the other financial variables are $0. If a non-financial variable is incorrect, the data in the appropriate variable (e.g., DX1) are corrected, and all of the financial variables are $0 on the adjustment record. This way, the sum of the financial variables of the erroneous claim and the adjustment claim equals the correct financial amounts. Under this method, therefore, two records represent a single transaction. An example is provided in Exhibit 6.

In order to resolve the adjustment, the MarketScan Database build process matches the adjustment with the original record, with the requirement that certain non-financial variables are exactly the same on both records. The financial information on the two records is summed, creating one resulting record.
The Void and Replace Method allows entry of a new claim that exactly duplicates all of the variables from the erroneous claim, except that the financial variables are entered as negative numbers. In this way, the original erroneous claim is fully voided, and the claim is re-entered with complete correct data in each variable. Under this method, three records are present to represent a single transaction. An example is provided in Exhibit 7.

In order to resolve the adjustment, the MarketScan Database build process matches the void record with the original record; provided certain non-financial information is exactly the same on both records and the financial information of the void record is the exact negative of the original record. The void and original records are dropped from the database, because all financial information on the combined record is zero. Only the replacement record remains.

Unresolved Adjustments

Because strict matching criteria are required to resolve adjustments, some adjustment records remain unresolved; these account for less than one percent of the records in the MarketScan Outpatient Services Table. These records generally contain changes to a variable that would normally be used to match the original and adjustment records. For example, if the original Provider ID (PROVID) was incorrect and the adjustment record adjusted for that ID, the two records would not match because PROVID is a key variable. Both records would remain. When performing person-level analysis or broader levels of analysis such as geographic region, all claims should be included.

Exhibit 6. The Adjustment Method: Example

<table>
<thead>
<tr>
<th>Record Type</th>
<th>ENROLID</th>
<th>SVCDATE</th>
<th>DX1</th>
<th>PAY</th>
<th>NETPAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>9876501</td>
<td>19970630</td>
<td>12345</td>
<td>100</td>
<td>70</td>
</tr>
<tr>
<td>Adjustment</td>
<td>9876501</td>
<td>19970630</td>
<td>12345</td>
<td>-20</td>
<td>0</td>
</tr>
<tr>
<td>Resulting</td>
<td>9876501</td>
<td>19970630</td>
<td>12345</td>
<td>80</td>
<td>70</td>
</tr>
</tbody>
</table>

Exhibit 7. The Void and Replace Method: Example

<table>
<thead>
<tr>
<th>Record Type</th>
<th>ENROLID</th>
<th>SVCDATE</th>
<th>DX1</th>
<th>PAY</th>
<th>NETPAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>9876501</td>
<td>19970630</td>
<td>12345</td>
<td>100</td>
<td>70</td>
</tr>
<tr>
<td>Void</td>
<td>9876501</td>
<td>19970630</td>
<td>12345</td>
<td>-100</td>
<td>-70</td>
</tr>
<tr>
<td>Replacement</td>
<td>9876501</td>
<td>19970630</td>
<td>12345</td>
<td>80</td>
<td>70</td>
</tr>
<tr>
<td>Resulting</td>
<td>9876501</td>
<td>19970630</td>
<td>12345</td>
<td>80</td>
<td>70</td>
</tr>
</tbody>
</table>
PERSON-LEVEL IDENTIFIERS

Enrollee Identifiers

One of the major strengths of the MarketScan Databases is the ability to track patients and families longitudinally. The unique person-level identifier is consistent across an individual’s enrollment, medical, and drug records, even as the individual moves from the Commercial Claims and Encounters Database to the Medicare Supplemental and COB database.

The enrollee identifier (ENROLID) is assigned based upon the data contributor, an encrypted employee identifier (usually an encrypted contract identifier), the relationship of the enrollee to the contract holder, the sex of the enrollee, and the enrollee date of birth or birth year and month.

Enrollee Identifiers Prior to 2001

Beginning in 2001, all MarketScan contributors submitted person-level enrollment information. In data prior to 2001, enrollee identifiers are derived from all data contributors and are not limited to those submitting person-level enrollment data. The methodology used to assign ENROLID differs, depending on the level of information available from a particular data contributor.

MarketScan data contributors fall into three categories with respect to the level of information available on claims data for assigning ENROLID:

1. Contributors submitting person-level enrollment data and also reporting patient date of birth
2. Contributors not submitting person-level enrollment data but reporting patient date of birth
3. Contributors not submitting person-level enrollment data or patient date of birth but reporting patient age.

Type 1 data contributors submit sufficient information on enrollment records to differentiate individuals and accurately assign enrollee identifiers. For Type 2 and Type 3 data contributors, enrollee identifiers cannot be assigned using enrollment data; therefore, elements found in the claims data become the basis for assigning enrollee identifiers.

For Type 2 data contributors, ENROLID assignment is derived from claims data using the same set of variables as Type 1 data contributors, but the data source is the claim rather than a person-level eligibility record.

For Type 3 data contributors, ENROLID is assigned by using the patient age provided on the claim to derive the year of birth. Because the date of birth is an approximation for Type 3 contributors, it is impossible to distinguish between same-sex siblings born within a year of each other.

The Enrollee ID Derivation Flag (EIDFLAG) describes which of these three methodologies was used to assign the enrollee identifier. See Exhibit 8 for a summary of the flag contributors.

- EIDFLAG=1 indicates that the data contributor supplied person-level enrollment data (ENRFLAG=1) and that an individual’s enrollment record was used to assign ENROLID.
- EIDFLAG=2 indicates that the data contributor supplied enrollment data (ENRFLAG=1) but the variables used to assign ENROLID on a claim did not link to a single person record in the Enrollment data. Claim information was used to assign ENROLID.
- EIDFLAG=3 indicates that the data contributor supplied enrollment data (ENRFLAG=1) but one or more of the variables needed to identify an individual was missing from the claims record (i.e., the claim was missing enrollee relationship to contract holder, sex, or patient date of birth). ENROLID is set to missing.
- EIDFLAG=4 indicates that the data contributor did not supply person-level enrollment data (ENRFLAG=0) and enrollee identifiers were assigned using claim information.
- EIDFLAG=5 indicates that data contributor did not supply person-level enrollment data (ENRFLAG=0) and one or more of the variables needed to identify an individual was missing (i.e., the claim was missing enrollee relationship to contract holder, sex, or patient date of birth). ENROLID is set to missing.
- EIDFLAG=6 indicates that the data contributor did not supply person-level enrollment data (ENRFLAG=0) and did not supply patient date of birth on the claim. A "pseudo" ENROLID was assigned based on information derived from the medical claim.
## Exhibit 8. Enrollee ID Derivation Flag (EIDFLAG)

<table>
<thead>
<tr>
<th>Enrollment Data Contributors (ENRFLAG = 1)</th>
<th>Non-Enrollment Data Contributors (ENRFLAG = 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EIDFLAG = 1</td>
<td>EIDFLAG = 4</td>
</tr>
<tr>
<td>EIDFLAG = 2</td>
<td>EIDFLAG = 5</td>
</tr>
<tr>
<td>EIDFLAG = 3</td>
<td>EIDFLAG = 6</td>
</tr>
<tr>
<td>Enrollment</td>
<td>Claim</td>
</tr>
<tr>
<td>ENROLID Present</td>
<td>ENROLID Present</td>
</tr>
<tr>
<td>ENROLID Present</td>
<td>ENROLID Present</td>
</tr>
<tr>
<td>ENROLID Present</td>
<td>ENROLID Present</td>
</tr>
<tr>
<td>ENROLID Present</td>
<td>ENROLID Present</td>
</tr>
<tr>
<td>ENROLID Missing</td>
<td>ENROLID Present*</td>
</tr>
</tbody>
</table>

* "pseudo" ENROLID assigned. May be indistinct.
Diagnosis codes in MarketScan data use the International Classification of Disease, Ninth Revision, Clinical Modifications (ICD-9-CM) classification system for service dates September 30, 2015 and prior. For service dates starting October 1, 2015, the International Classification of Disease, Tenth Revision, Clinical Modifications (ICD-10-CM) classification system is used. A Diagnosis Version field (DXVER) is included in the data to indicate which coding system is in use. Note that it is possible for one string to be valid in both systems.

ICD-9-CM diagnosis codes are three to five digits in length. The first character can be alphanumeric (0–9, E or V); characters two through five are numeric or blank. There are approximately 15,800 valid ICD-9-CM codes. In MarketScan data, the decimal point is implied between the third and fourth digit of the code. The data are left justified. Examples are provided in Exhibit 9a.

### Exhibit 9a. ICD-9-CM Diagnosis Codes: Example

<table>
<thead>
<tr>
<th>ICD-9-CM</th>
<th>MarketScan data value</th>
</tr>
</thead>
<tbody>
<tr>
<td>390</td>
<td>390 (followed by 2 spaces)</td>
</tr>
<tr>
<td>012.1</td>
<td>0121 (followed by 1 space)</td>
</tr>
<tr>
<td>223.89</td>
<td>22389</td>
</tr>
</tbody>
</table>

ICD-10-CM diagnosis codes are three to seven digits in length. The first character can be alphanumeric; the second character is numeric, the third character is alphanumeric, the fourth through seventh characters are alphanumeric or blank. There are approximately 70,000 valid ICD-10-CM codes. In MarketScan data, the decimal point is implied between the third and fourth digit of the code. The data are left justified. Examples are provided in Exhibit 9b.

### Exhibit 9b. ICD-10-CM Diagnosis Codes: Example

<table>
<thead>
<tr>
<th>ICD-10-CM</th>
<th>MarketScan data value</th>
</tr>
</thead>
<tbody>
<tr>
<td>E02</td>
<td>E02 (followed by 4 spaces)</td>
</tr>
<tr>
<td>M86.9</td>
<td>M869 (followed by 3 spaces)</td>
</tr>
<tr>
<td>C72.20</td>
<td>C7220 (followed by 2 spaces)</td>
</tr>
<tr>
<td>B08.010</td>
<td>B08010 (followed by 1 space)</td>
</tr>
<tr>
<td>W00.9XXA</td>
<td>W009XXA</td>
</tr>
</tbody>
</table>

Up to four diagnosis codes (DX1, DX2, DX3, DX4) are recorded on every Inpatient Service record. The principal diagnosis on the Inpatient Admissions Table is generally identified as the discharge diagnosis on a hospital claim. Up to 14 secondary diagnosis codes (DX2 through DX15) from individual Inpatient Service records are included on the corresponding Inpatient Admission record. Up to four diagnosis codes (DX1, DX2, DX3, DX4) are recorded on each Outpatient Service record. Up to nine diagnosis codes (DX1 through DX9) are recorded on each Facility Header record.

Procedure Codes in MarketScan data are three to seven positions in length, depending on the classification system used. The Current Procedural Terminology, 4th Edition\(^1\),\(^2\) (CPT\(^\circ\)-4) coding system is most prevalent. CPT-4 procedure codes appear on physician claims and many outpatient facility claims. CPT-4 codes are five-digit numeric codes.

ICD-9-CM or ICD-10-PCS procedure codes are found on facility claims. These codes are three to four digits in length and are all numeric. There is an implied decimal between the second and third digits for ICD-9-CM procedure codes; there is no decimal

---

\(^1\) CPT copyright 2015 American Medical Association. All rights reserved. Applicable FARS/DFARS Restrictions Apply to Government Use.

\(^2\) Fee schedules, relative value units, conversion factors and/or related components are not assigned by the American Medical Association (AMA), are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein.
point in ICD-10-PCS procedure codes. Examples are provided in Exhibit 10.

Exhibit 10. Example: ICD-9-CM, ICD-10-PCS Procedure Codes

<table>
<thead>
<tr>
<th>ICD-9-CM, ICD-10-PCS</th>
<th>MarketScan data value</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.9</td>
<td>139 (followed by 4 spaces)</td>
</tr>
<tr>
<td>13.19</td>
<td>1319 (followed by 3 spaces)</td>
</tr>
<tr>
<td>001U3J7</td>
<td>001U3J7</td>
</tr>
</tbody>
</table>

Note: Effective with the 2000 data year, the MarketScan databases contain CPT-4\(^{1,2}\) procedure code modifiers for some data contributors.

The Centers for Medicare & Medicaid Studies (CMS) Healthcare Common Procedural Coding System (HCPCS) procedure codes are found less often than CPT and ICD procedure codes in MarketScan data. These codes are five digits in length. The first character is alpha; all other characters are numeric. HCPCS codes beginning with “J” are included in the MarketScan databases and represent injectable drugs.

One procedure code (PROC1) is stored on each Inpatient Service record. From the individual Inpatient Services comprising one Inpatient Admission record, one procedure code is identified and assigned as the principal procedure (PPROC). Up to 14 secondary procedure codes (PROC2 through PROC15) from individual Inpatient Service records are included on the corresponding Inpatient Admission record. One procedure code (PROC1) is included on each Outpatient Service record. Up to six procedure codes (PROC1 through PROC6) are included on each Facility Header record. Most procedure codes on the Facility Header Table use the ICD-9-CM or ICD-10-PCS procedure coding systems.

The variable Procedure Code Type (PROCTYP) identifies the type of procedure code (e.g., HCPCS, CPT-4). Use this variable in conjunction with the Procedure Code 1 (PROC1) variables on the Inpatient Service and Outpatient Service records to designate the coding system of interest.

The quality of diagnosis and procedure coding varies among the approximately 350 payers or administrators represented in the MarketScan databases. Every effort is made to select the data contributors with the best coding. The diagnosis and procedure codes are validated and edited, if necessary. (See Frequently Asked Questions, Q12 for a detailed description of validation and editing.)

If data contributors submit old codes, these codes are retained in the MarketScan data and reflect their original definition.

Note: When defining a diagnosis or procedure of interest, first run a frequency distribution in the range of interest (e.g., analyze the frequency of 53x.xx (ICD-9-CM), K25.xxxx (ICD-10-CM) diagnosis codes for patients with stomach ulcers), analyze the coding practices, and then create the criteria for diagnosis and procedure selection.
MARKETSCAN DATABASE CONSTRUCTION

The MarketScan Research Databases are constructed from privately insured, paid medical and prescription drug claims. The data contributors are generally self-insured. Collectively, the databases incorporate data from almost 350 payers, including commercial insurance companies, Blue Cross and Blue Shield plans, and third-party administrators (TPAs).

Each contributor database is constructed by collecting raw data from the appropriate payer(s). These raw data are service-level adjudicated paid claims and capitated encounters containing both inpatient and outpatient services. Financial, clinical, and demographic variables are standardized to common definitions, and variables specific to employers are also added. Clinical detail is added to the Outpatient Pharmaceutical Claims Table. Examples of the detail include therapeutic class, therapeutic group, manufacturer’s average wholesale price, and generic product identifier.

Truven Health then applies an admission construction methodology to assemble the inpatient paid services into one record per inpatient admission. During the admission creation process, variables such as Primary Diagnosis (PDX) are created and included on both the inpatient admission record and the inpatient service record.

Data Quality

Edits on the reasonableness of data check the distribution of categorical fields to ensure that they are reasonable against norms. Validity checks are conducted for selected fields, including diagnosis codes, procedure codes, date(s) of service, sex, and age, to compare recorded values to lists of possible valid values for those fields. Improper coding is flagged to recommend data quality improvement actions to the carrier or data processor.

The MarketScan Databases are created by combining the standard variables of the individual databases (data contributors) and also creating links between years of data and across all data types. The MarketScan Databases are created as a snapshot in time and are based on a calendar-year incurred period. The MarketScan data flow is depicted in Exhibit 11.

Claims lag periods (the amount of time between the date of service on the claim and the date payment is made) vary considerably across the insurance carriers in the MarketScan Databases. Because of this, the data are collected when close to 100% of claims have been paid, which takes about 6 months after year end.

Additional enhancements to the data during the MarketScan Database creation process include:

- Comparing and validating diagnosis and procedure codes to codes that were in effect at that time.
- Adding the Metropolitan Statistical Area (MSA) of the primary beneficiary to claims.
- Integrating benefit plan characteristics, enrollment, outpatient pharmaceutical claims, and medical/surgical data.
- Adding major diagnostic categories (MDCs) and diagnosis-related groups (DRGs) to claims.
- Creating a common synthetic patient identifier, which enables a patient to be tracked over years across medical/surgical, outpatient pharmaceutical, enrollment, and benefit plan files and across databases (e.g., Commercial Claims and Encounters Database and Medicare Supplemental and COB Database) while ensuring patient confidentiality.
- Identifying the type of plan for the patient, such as preferred provider organization (PPO), point of service (POS) plan, or comprehensive plan.
- Verifying that both the experience and the denominator populations exist for all subsets of the data.
- Standardizing place, service type, and provider type values and industry classifications.

Note: Data are not edited for concordance between diagnosis and procedure codes or demographic variables such as sex.
Exhibit 11. MarketScan data flow chart

MarketScan® Data Flow

Employer Databases
- Cases
- Inpatient Services
- Outpatient Services
- Enrollment Counts
- Outpatient Prescription Drug Claims
- Capitated Encounter Records

Data Quality Review
- “Good Data Subset”
- Year to Year Comparisons of Utilization and Payments Rates
- Links made across years and across data types

Standardization
- Financial Fields
- Variable Mapping
- Variable Selection
- DX/Proc Code Validation

Aggregation
- Variable Creation
- Variance across Individual Databases (e.g., Pay/Capita)

MarketScan Databases
- Medical/Surgical Database
  - Inpatient Cases
  - Inpatient Services
  - Outpatient Services
  - Capitated Encounter Records
  - Populations
  - Facility Header

- Outpatient Prescription Drug Table
- Enrollment Table

Summary Plan Description Booklets

Abstract Information

Double Entry/Edit

Benefit Plan Design Database
PLAN TYPE DEFINITIONS

The plan types in the MarketScan databases are based on the definitions provided in Exhibit 12. The summary grid identifies the basic differences between plan types.

Exhibit 12. Type of Plan (PLANTYP)

<table>
<thead>
<tr>
<th>Plan Type</th>
<th>Patient incentive to use certain providers?</th>
<th>PCP assigned?</th>
<th>Referrals from PCP to specialists required?</th>
<th>Out of network services covered?</th>
<th>Partially or fully capitated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 B/MM</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>n/a</td>
<td>No</td>
</tr>
<tr>
<td>2 COMP</td>
<td>No</td>
<td>No</td>
<td>n/a</td>
<td>n/a</td>
<td>No</td>
</tr>
<tr>
<td>3 EPO</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4 HMO</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>5 Non-Cap POS</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>6 PPO</td>
<td>Yes</td>
<td>No</td>
<td>n/a</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>7 Cap or Part Cap POS</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>8 CDHP</td>
<td>Varies</td>
<td>No</td>
<td>n/a</td>
<td>Varies</td>
<td>No</td>
</tr>
<tr>
<td>9 HDHP</td>
<td>Varies</td>
<td>No</td>
<td>n/a</td>
<td>Varies</td>
<td>No</td>
</tr>
</tbody>
</table>

PCP = primary care physician; other abbreviations are defined below.

1. Basic/Major Medical (B/MM)

There is no incentive for the patient to use a particular list of providers. Coverage is handled in two phases: a basic policy handles the first set of charges—usually a hospital admission—with no out-of-pocket charge. After the basic policy will no longer pay, a major medical policy assumes coverage, usually with a deductible and coinsurance.

2. Comprehensive (COMP)

There is no incentive for the patient to use a particular list of providers. Coverage is handled by only one policy, with a deductible and coinsurance.

3. Exclusive Provider Organization (EPO)

Patients must choose from a particular list of providers for all non-emergency care. Each patient chooses a primary care physician (PCP) to manage all of his or her care. Referral from the PCP is required for treatment by specialists. Services are not paid by the plan on a capitated basis.

4. Health Maintenance Organization (HMO)

Patients must choose from a particular list of providers for all non-emergency care. Each patient chooses a PCP to manage all of his or her care. Referral from the PCP is required for treatment by specialists. All or some services are paid by the plan on a capitated basis.
5. Non-Capitated Point-of-Service (Non-Cap POS)

Patients are offered financial incentives through a lower copay or deductible to use a particular list of providers. Each patient chooses a PCP to manage all of his or her care. Referral from the PCP is required for treatment by specialists. No services are capitated and patients may seek treatment outside the network, usually with a financial penalty.

6. Preferred Provider Organization (PPO)

Patients have financial incentives, such as a lower copay or deductible, to use a particular list of providers. A PCP is not required and specialist referrals are not necessary. No services are capitated. Patients may seek treatment outside the network, usually with a financial penalty. The financial incentives may be offered only through discounted rates within the network.

7. Capitated or Partially-Capitated Point-of-Service (Cap or Part Cap POS)

Patients are offered financial incentives to use a particular list of providers through a lower copay or deductible. Each patient chooses a PCP to manage all of his or her care. Referral from the PCP is required for treatment by specialists. All or some services are paid on a capitated basis. Patients may seek treatment outside the network, usually with a financial penalty.

8. Consumer-Driven Health Plan (CDHP)

A Consumer Driven Health Plan (CDHP) is a PPO plan coupled with a Health Reimbursement Arrangement (HRA). The PPO plan typically has a relatively high deductible but may carve drugs in or out of the HRA and plan deductible. The HRA is a notional account that is 100% paid from employer funds; an HRA is not pre-funded with actual employer monies.

9. High Deductible Health Plan (HDHP)

A High Deductible Health Plan is a statutory HDHP (as defined in the Medicare Modernization Act of 2003) that is coupled with a Health Savings Account (HSA). An employee is 100% vesting in HSA funds, and either the employer or employee can contribute to the HSA. The HSA is a tax-advantaged, portable savings account owned by the employee. HDHP plan design features such as deductibles and contribution limits are indexed each year by the Treasury Department. An HDHP must conform to the statutory plan design requirements in order to use an HSA to defray HDHP costs.
KEY TABLE AND FIELD RELATIONSHIPS

Although the databases in their native format are not truly normalized, there are several key fields that are used to relate tables to each other. These relationships are described below.

1. ENROLID

Related Tables: Inpatient Admissions (I), Inpatient Services (S), Outpatient Services (O), Prescription Drugs (D), Facility Header (F), Annual Enrollment Summary (A), Enrollment Detail (T)

Relationship: Unique on A; not unique on I, S, O, D, F, T.

Function: This is the unique enrollee identifier across all MarketScan data products. The Annual Enrollment Summary (A) table provides one record per enrollee for the entire year, so ENROLID will be unique on this table. The Enrollment Detail table (T) provides one record per enrollee per enrolled month, so one ENROLID can appear on as many records in the T table as months an individual was enrolled. ENROLID can appear multiple times (or not at all, if a person did not receive any services) in the medical and pharmacy claims files.

2. CASEID

Related Tables: Inpatient Admissions (I), Inpatient Services (S), Facility Header (F)

Relationship: Unique on I; not unique on S, F

Function: This field is a unique identifier for each inpatient admission in the data. The Inpatient Admissions (I) table is structured as one record per inpatient admission, so CASEID values will be unique on the I table. The individual detail service records that comprise all services that make up an admission are stored in the Inpatient Services (S) table, and all of these individual services will have the corresponding CASEID value.

CASEID also appears on the Facility Header (F) table, where applicable.

The CASEID value for a particular admission will not necessarily remain the same between different versions of the same database. Blending database versions is not recommended.

3. FACHDID

Related Tables: Facility Header (F), Inpatient Services (S), Outpatient Claims (O)

Relationship: Unique on F; not unique on S or O.

Function: This field is a unique identifier for a Facility Header claim. It is the header information from one UB04 Facility claim form. The related detail information from each facility claim form is found in either the Inpatient Services (S) or Outpatient Claims (O) table, depending on the site of service (inpatient or outpatient). FACHDID is unique on the F table. It links to the many detail line services found in either the S or the O table.

Note: some of our data suppliers create an artificial, universal one-to-one relationship between header and detail (i.e., every facility header record from those data suppliers has exactly one associated detail row).

The FACHDID value for a particular claim header will not necessarily remain the same between different versions of the same database. Blending database versions is not recommended.

4. NDCNUM

Related Tables: Prescription Drug (D), RED BOOK (R)

Relationship: Unique on R; not unique on D

Function: The RED BOOK table is a supplemental table to provide additional information about prescription drugs (e.g., generic name, manufacturer, therapeutic class). Drugs are listed in this file by National Drug Code. The code is linkable to the Prescription Drug Claims table (D), so that selection of drug claims may be made by the categorical fields included in the RED BOOK.
GLOSSARY OF ACRONYMS, ABBREVIATIONS, AND TERMS

Acute care—Services within a hospital setting intended to provide patients with medical and surgical care over a relatively short period of time.

Adjudication—The process of claims review by the carrier to determine whether or not the claims should be paid and how much money should be paid for each claim.

Adjustment records—Claims in some databases that represent financial adjustments to original claims. The dollar amounts of these adjustments may be negative, or the record may include an “adjustment indicator” that indicates whether the adjustment is positive or negative. There are also specific terms that refer to adjustments as we receive them from carriers. A “bulk adjustment” is a single quarterly or annual adjustment for a hospital discount (not typically loaded on the database). A “void adjustment” is a record that simply cancels an earlier claim record. A replacement claim record usually follows it. A “void and replace” adjustment is a single record that stores both the cancellations of the earlier claim and the new claim. An “adjustment to net pay” just shows the difference between the original net pay amount and what the carrier actually paid.

Administrator—Person or firm who pays claims under an Administrative Services Only (ASO) contract. Also known as a third-party administrator (TPA).

Admission—An acute inpatient hospital stay covered by the patient’s benefit plan. To the extent that such care is covered, admissions may include hospital stays, psychiatric stays, psychiatric night care, and stays for alcoholism, substance abuse, and rehabilitative care. Also called a “case” or a “stay”.

Admission date—The date a patient begins a stay in a hospital or other overnight healthcare facility.

Ambulatory care—Medical services provided on an outpatient (non-hospitalized) basis. Services may include diagnosis, treatment, surgery, and rehabilitation.

Ambulatory surgery—Surgery for which there is no overnight stay in a hospital. The patient walks in and walks out on the same day.

Annualization—A statistical technique for estimating a yearly rate using data collected over a shorter time frame (e.g., a quarter or month) or over a longer time frame (e.g., 30 months).

Average length of stay (ALOS)—The average number of days per hospital admission for a group of admissions. Usually examined by looking at the ALOS for a single MDC or DRG at a given employee location or other variable and comparing that to a norm, another location, etc. See also length of stay.

Benefit—Conventionally defined as the amount payable for a loss under a specific insurance coverage (indemnity benefits) or as the guarantee that certain services will be paid.

Business coalitions—Groups of employers, which may or may not include health plans, that seek to control healthcare costs and ensure quality by aggressively regulating prices, assuming administrative tasks related to healthcare, and/or asking health plans to develop and provide data on measures of quality and outcomes.

Capitation—(1) A predetermined amount prepaid to a provider for a specific group of services that are defined in the contract, usually in a health maintenance organization (HMO) arrangement. The provider is paid based upon the number of members who have selected him/her as their primary care physician (PCP). (2) A fixed, predetermined amount paid to a provider for each member who has elected to seek care from that provider. Total payment to the provider (sum of per person enrolled payment amount) is based on the number of people who enroll without regard to the actual number or nature of services provided to members. This is the characteristic payment method for primary care in HMOs.
Carrier—The party to the group contract who agrees to underwrite and provide certain types of coverage and service. Examples are commercial insurers (e.g., Aetna, Metropolitan, Prudential) and Blue Cross and Blue Shield.

Carve-out—A program separate from the primary group health plan that is designed to provide a specialized type of care, such as a mental health carve-out. Also, a method of integrating Medicare with an employer’s retiree health plan (making the employer plan excess or secondary), which tends to produce the lowest employer cost.

Case level—A variable that is found in the Inpatient Admissions Table. These tend to be demographic variables that are the same for the entire case (e.g., patient age and sex, employee ID number), clinical variables that refer to the case as a whole (e.g., MDC, DRG), or financial variables that summarize all services for a case (e.g., total payments). See service level for comparison.

Center for Medicare and Medicaid Services (CMS)—(1) A division within the U.S. Department of Health and Human Services (DHHS). This division oversees all of the regulatory and financing activities for Medicare and Medicaid. (2) The portion of the Federal Government responsible for payment of Medicare. CMS was formerly named the Health Care Financing Administration (HCFA).

Charges—The amount patients or third-party payers are billed for care.

Claims data—Information that comes from provider claims to third-party payers. Claims data usually include personal patient identifying information, the services performed, and the amount paid by the patient. Claim forms are generally used by enrollees of standard indemnity plans (i.e., fee-for-service plans).

Claims lag—(1) Generally refers to the period between the date a healthcare service is incurred and the date the claim for that service is submitted to the administrator for payment. (2) The Truven Health definition is the period between the service date and the paid date on a claim. See also run-off.

Coding—The carrier’s claims-data handling process. “Coding problems” means that the carrier has entered inaccurate or imprecise data into the claims record, has failed to fill in one or more data variables, or has failed to include one or more variables in the record extract.

Coinsurance—(1) The percentage of a covered medical expense that a health plan or beneficiary must pay (after a deductible is met). (2) A policy provision by which both the insured and the insurer share hospital and medical expenses in a specified ratio (commonly 20%:80%), after the deductible is met. Coinsurance amounts are stored in the Truven Health variable COINS.

Completion factors—(1) Factors that allow a quantitative measure of data completeness. These factors range in value between 0 (no data) and 100 (a full month of data) for services in any month. Completion factors are used to derive the number of months of data and an annualization factor for rate calculations. They are also used to derive weighted population averages. (2) A percentage that estimates how many of the cases that occurred in a given month are online in a client database. Completion factors of less than 100% are due to run-off or run-up. The percentage of data missing for each month is used to annualize the cost and use rates for that month on clinical reports.

Comprehensive Omnibus Budget Reconciliation Act (COBRA)—(1) A congressional act passed in 1985 that requires continuation of benefits to plan participants who previously would have been ineligible because of a qualifying event. (2) A program that gives employees who leave a firm the option of continuing their health coverage with that firm for a period of time. The employee pays the premium.

Coordination of benefits/maintenance of benefits (COB/MOB)—(1) After one insurance carrier has paid a claim, the second carrier pays an amount that covers the patient up to the benefit level of the second policy only. (2) COB coverage between carriers so that the insured does not receive double payment for services when a subscriber has coverage from two or more sources. An example is a husband and wife who work at different companies and choose to be covered by both employers’ insurance. COB policies also establish primary and secondary payment responsibilities. (In the Truven Health system for older databases, the
COB variable may represent dollars saved for reasons other than COB, such as penalties for non-compliance.)

Copay or copayments—(1) Copayments are generally a preset amount per covered visit or service (e.g., $10) paid by the patient. (2) A fixed payment, paid by the patient, for a given service or procedure. This payment is customarily made at the time of service. Copayment amounts are stored in the Truven Health variable COPAY.

Cost sharing—Arrangements whereby consumers pay a portion of the cost of the health services, sharing costs with employers. Deductibles, copayments, coinsurance, and payroll deductions (premium contributions) are forms of cost sharing.

Cost shifting—Occurs when a provider inflates charges for a given procedure or patient in order to cover losses associated with charges (payments received) for other patients or procedures.

CPT or CPT-4 codes—Physicians’ Current Procedural Terminology codes—(1) Physicians’ most commonly used coding scheme (five-digit codes) used to identify the medical or surgical procedure that occurred for a patient; most frequently used for billing by professionals. (It is often referred to as CPT-4, with 4 representing the 4th revision). (2) A system developed by the American Medical Association used to classify procedures and services rendered by physicians. It is used by physicians on the CMS 1500 form to describe services rendered to a patient and to request payment for those services. See also ICD-9-CM, HCPCS.

Deductible—The portion of a subscriber’s healthcare expenses that must be paid out of pocket before any insurance coverage applies. Commonly $100 to $300. Not allowed in federally qualified HMOs. The deductible usually must be met again each benefit year before the insurer will begin paying for benefits. The deductible amount is stored in the Truven Health variable DEDUCT.

Dependent—An insured’s spouse and unmarried children who meet certain eligibility requirements and who are not otherwise insured under the same group policy. The precise definition of a dependent varies by insurer or employer.

Diagnosis (Dx)—The determination of the nature of a disease based on the medical symptoms of a patient; a concise technical classification of a health situation. The diagnosis helps to determine necessary procedures.

Discount—Arrangement whereby a payer has negotiated a reduced payment with a provider in return for a patient incentive.

Eligible—A contract holder and his or her spouse and dependents who are enrolled in a benefit plan.

Encounter—(1) A unit of measure denoting one patient-provider contact or appointment. Multiple services may be delivered during one encounter. Encounters can take place on an inpatient or outpatient basis. (2) A patient visit to a capitated provider; no fee-for-service payment.

Encounter record—A record of a patient encounter reflecting who visited what provider and what services were provided. The form used to capture encounter data applies to non-fee-for-service arrangements (capitated).

Enrollees—Employees, contract holders, spouses, and dependents who are enrolled in a benefit plan (also known as “covered lives”).

Exclusions—Services or procedures that are not covered according to the plan provisions.

Exclusive provider organization (EPO)—A preferred provider organization (PPO) in which patients are required to use the PPO network providers.

Fee-for-service (FFS)—A method of payment based on reimbursing providers for each unit of service or treatment provided.

Fee-for-service equivalent (FFSE)—An amount sent on claims records representing what would have been charged for a service if the service were not covered by a capitation arrangement.

Gatekeeper—(1) The PCP responsible for managing medical treatment rendered to an enrollee of a health plan. (2) A designated healthcare practitioner who provides primary care services and coordinates specialist and other care for health plan members. Members typically are charged extra costs for care that is not provided or coordinated by the gatekeeper.
Grouper—Software that assigns claims to a common clinical grouping. In MarketScan, groupers are used to assign a DRG and MDC to each inpatient admission based on diagnosis and procedure coding received from the carrier (provided the diagnosis and procedure coding from the carrier is adequate).

HCFA Common Procedure Coding System (HCPCS)—(1) A procedure coding system that includes all CPT-4 codes plus supplemental codes not included in CPT-4’s (e.g., ambulance, chiropractic services). (2). One of several schemes used to classify healthcare activity. HCPCS was based on CPT-4 coding and expanded to include non-physician provider procedures. The acronym is pronounced “hick-picks.” See also CPT-4, ICD-9-CM.

Health maintenance organization (HMO)—(1) An entity that accepts responsibility and financial risk for providing specified healthcare services to a defined population during a defined period of time at a fixed price. There is generally no coverage for non-ER care panels of practitioners and providers. (2) The Health Maintenance Act of 1973 (PL93-222) defines an HMO as a legal entity or organized system of healthcare that provides an agreed-upon set of comprehensive health services to a voluntary enrolled population in exchange for a predetermined, fixed, and periodic payment. See also open-ended HMO.

Hospital payments—Facility payments only.

Incurred but not reported (IBNR)—Claims for services that have been incurred but not yet paid by the carrier. See also claims lag.

International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)—A nationally uniform system for coding clinical conditions (diagnoses), used prior to October 1, 2015 by nearly all providers and claims payers. It also includes procedures coding used by hospitals. ICD-9-CM includes both diagnostic and procedure coding required by the Grouper to assign DRGs and MDCs. It is also known as I9. See also CPT-4, HCPCS.

International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System (ICD-10-CM/PCS)—A nationally uniform system for coding clinical conditions (diagnoses), used effective October 1, 2015 by nearly all providers and claims payers. It also includes procedures coding used by hospitals. ICD-10-CM/PCS includes both diagnostic and procedure coding required by the Grouper to assign DRGs and MDCs. It is also known as I10. See also CPT-4, HCPCS.

Incurred date—The date on which the activity or service took place. See also paid date, claims lag, IBNR.

Indemnity (traditional) insurance—(1) A healthcare insurance plan designed to reimburse patients for losses due to healthcare costs; typically used to characterize fee-for-service payment plans. (2) The most common form of health insurance coverage in recent decades. The indemnity insurer usually administers claims and does not provide healthcare services. A typical coverage arrangement is 80% of a claim covered by the insurer and 20% covered by the patient or enrollee (also referred to as coinsurance). Indemnity plans typically also require that the covered person meet an annual deductible (e.g., $200) before the insurer will begin to pay a percentage of claims incurred.

Individual practice association (IPA)—A type of HMO. A group of physicians who practice independently, but also provide services for an HMO under a contract agreement. An IPA physician can and does provide “traditional” fee-for-service healthcare to non-HMO patients.

Inpatient—(1) Pertaining to the medical care of an individual admitted to the hospital for at least one night. (2) That portion of the base relating to hospital admissions. Length of stay (DAYS) will be at least one day.

Inpatient payments—All facility, professional and other payments related to a hospital admission.

Length of stay (LOS)—The number of days (DAYS) the patient was confined (spent in the hospitals) during the inpatient admission. See also ALOS.

Long-term disability (LTD)—(1) A significant period of disability generally ranging from 6 months to life. (2) Wage replacement insurance for...
individuals who are (partially or totally) permanently disabled.

**Mail-order pharmacy**—A company that receives prescriptions from physicians or patients via fax or mail and then mails the medication to patients. Meanwhile, the physician provides the patient with enough of the medication to last until the prescription arrives. Generally, the cost per prescription from mail-order pharmacies is lower than the cost at other pharmacies because of higher volume and lower overhead.

**Major Diagnostic Category (MDC)**—(1) A classification system for grouping medical conditions into one of 25 categories. The first 16 categories refer to major body systems; the remaining categories encompass more than one body system. (2) A widely recognized classification system that groups medical conditions into broad classifications, mostly by body system. Each DRG is assigned to one MDC.

**Managed care**—(1) Employing incentives at both the provider and patient level that encourage the efficient provision of healthcare services. Common elements of managed care include: capitation, a primary physician acting as a gatekeeper, and patient copayments. (2) An organized system of healthcare services in contrast to the fee-for-service system.

**Medical**—Clinical in nature, as opposed to surgical.

**Medicare**—(1) A system of medical insurance provided by the Federal Government for all Americans aged 65 years and older and for Americans who are permanently disabled or have renal failure. (2) A Federal program under Title XIX of the Social Security Act providing health insurance for persons aged 65 and older and for other specified groups. Part A of Medicare covers hospitalization and is compulsory (automatically provided to any beneficiary who has qualified for participation in Social Security). Part B of the program covers outpatient services and is voluntary.

**National Drug Code (NDC)**—A standard 12-digit coding system used to identify drugs on drug claims.

**Net Pay**—The portion of the charge for a healthcare service that the carrier paid to the employee or assigned provider. NETPAY is calculated as: PAY minus DEDUCT minus COPAY minus COINS minus COB.

**Network providers**—Providers who have contracted to be part of a plan’s network; they may be capitated or on a discounted fee-for-service arrangement. Patients who visit out-of-network providers generally pay greater out-of-pocket amounts.

**Open-ended HMO**—An HMO that allows the patient to receive services from a non-network provider. Although such services will be covered, the patient must pay higher-than-normal copayments and deductibles.

**Out-of-pocket costs (OOP)**—The portion of the claim that the patient or enrollee is obligated to pay (copayments, coinsurance, deductible). Typically there is an annual OOP maximum. If the maximum is met, the insurer pays 100% of the costs incurred by the enrollee for the remainder of the plan year.

**Paid date**—The date on which a claim is paid (PDDATE). Claims data is usually received from carriers on the basis of paid date. For example, a submitted data file may contain all claims that were paid during the fourth quarter of 2013, regardless of when the claims were incurred. See also incurred date, claims lag.

**Point of service (POS) plan**—Replacement of an indemnity plan. (1) A managed care plan that pays (reduced) benefits when patients receive healthcare services either from non-managed care network providers or without proper referral by their primary care physicians. (2) A benefit plan design where enrollees must access the healthcare system through a gatekeeper. In addition to differential co-insurance/copayment levels described under PPO, POS plans may include a differential deductible for in- and out-of-network services used (e.g., in-network deductible may be $250 and out-of-network deductible may be $500).

**Precertification or preauthorization**—Permission from the administrator for the hospital admission to occur or the services to be performed. This is a form
of utilization review based on the patient’s health status and treatment needs.

Preferred provider arrangement or prudent purchaser arrangement (PPA)—Same as a PPO.

Preferred provider organization (PPO)—(1) A health plan that gives patients lower rates if they use the physicians in the preferred group of providers. Patients may still use doctors outside that list, but usually pay more to do so. Participating physicians are normally under a contract and keep an independent practice in the community. Typically, they also enroll in other preferred provider programs. Physicians receive reduced rates in return for a larger patient flow—lower price for the promise of higher volume. (2) Providers (e.g., hospitals, physicians) offering discounts or other reduced rates to a healthcare purchaser. Patients are usually “channeled” by receiving improved benefits (lower/no deductibles or copayments). See also EPO, point-of-service PPO.

Premium—An amount paid periodically to purchase health benefits; for self-insured groups that do not purchase insurance, the term may refer to the per employee or per family cost of health benefits and may be used for planning and analysis purposes, even when no contribution to coverage is collected from the employee.

Primary care physician (PCP)—The physician that a patient in a managed care plan must see first for any health problem; the PCP acts as a gatekeeper and determines if and when the patient needs to see a specialist. PCPs are generally internists, pediatricians, family physicians, general practitioners, and occasionally obstetricians/gynecologists.

Procedure Group—Outpatient procedure groupings based on CPT-4 and HCPCS procedure code values.

Provider—A person or organization that provides healthcare services, such as a physician or hospital.

Referral—(1) Written authorization from a patient’s PCP for the patient to see a specialist. (2) An arrangement for a patient to be evaluated or treated by another provider.

Reimbursement—The dollar cost of covered products and services for which insurers pay.

Risk sharing—An agreement whereby the risks of providing care under a capitated arrangement are shared by multiple parties. For example, a pharmaceutical manufacturer assumes a portion of the financial risk for the use of a product with the provider. A risk-sharing arrangement may include a capitated payment for the unlimited use of a product, promotion of appropriate usage by the manufacturer, or performance guarantees based on predetermined outcomes.

Run-off period—The period of time representing the number of months between a claim’s service date and paid date. If the runoff month’s variable is equal to 6, it indicates that most claims are paid within 6 months of their service date.

Self-insurance—Funding of medical care expenses in whole or part through internal resources rather than through transfer of risk to an insurer.

Service date—The date that a medical care service is provided (SVCDATE).

Service level—A variable that is found in the Inpatient Services Table. These variables can be different for each service within an admission. Examples are service date, provider ID, diagnosis and procedure codes, and financial variables that contain only the amount for that service (e.g., charge, payment). See case level for comparison.

Short-term disability (STD)—(1) Wage replacement insurance for individuals temporarily disabled because of non-occupational injury or illness. (2) Often considered to be a disability lasting not longer than 6 months.

Stop-loss (out-of-pocket max)—(1) Usually, the maximum out-of-pocket amount that an individual or family could pay in a single plan year, including deductibles and copayment amounts. Alternatively, it may refer to the total dollar value of covered services after which the plan pays 100%. (2) The maximum out-of-pocket liability for a patient each year for deductibles, copayment, and coinsurance.

Subrogation—The assumption by a third-party (such as an insurance company) of another’s legal right to collect a debt or damages. It is related to COB (e.g., recoveries from auto insurance may reduce an insurer’s health benefit liability).
Summary Plan Description (SPD)—A legally required document that summarizes a company’s healthcare benefit plan.

Surgical—Pertaining to a service performed by a surgeon or involving surgery.

Third-party administration or administrator (TPA)—Administration of a group insurance plan by some person or firm other than the insurer or the policyholder. TPAs may also pay claims. TPA is also defined as the administrator or claims administrator.

Total charges—Total eligible charges, prior to reductions for reasonable and customary limits and PPO discounts.

Total payments—Total eligible charges less any reasonable and customary amounts and discounts for PPO services, but prior to reductions for deductibles, copayments, and other savings.

Uniform Billing (UB)—A standardized billing format for hospitals to use when submitting data to third-party payers. It is usually followed by a year that indicates when it was last revised (e.g., UB04).

Unbundling—Creative or fraudulent billing practices used by providers to increase payment by charging item-by-item for components of a medical procedure.

Usual, customary, and reasonable (UCR)—A method of payment to physicians based on the usual (U) charge of a particular physician for the procedure, the customary (C) charge for the procedure among physicians in the community, and a determination of what a reasonable (R) payment should be by the payer. This system is highly inflationary, because physicians typically increase their charges substantially to ensure that they attain a certain income. Plans often pay a percentage of UCR or a percentage of R and C. The patient is liable for the remainder, unless the physician is contractually obligated to accept it in full. (Balance billing is the practice of billing the patient for the remainder.)

Utilization review (UR)—(1) A generic term referring to any of a number of programs to control hospital admissions and/or lengths of stay. Examples are second surgical opinion programs, length-of-stay certification, concurrent review, and pre-admission certification. (2) A managed care process focused on the point at which care is (to be) provided, typically for expensive events, e.g., in the case of hospital admission or outpatient surgery, the necessity and appropriateness of the procedure are reviewed against medical criteria by a third party.

Wellness benefits—A broad range of employer or union-sponsored facilities and activities designed to promote safety and good health among employees. The purpose is to increase worker morale and reduce the costs of accidents and ill health such as absenteeism, lower productivity, and healthcare costs. It may include physical fitness programs, smoking cessation, health risk appraisals, diet information and weight loss, stress management, and blood pressure screening.

Withhold amount/pool process—The dollar amount retained or withheld from the servicing provider and placed in a risk-sharing pool for future distribution.
In preparing an analytic plan, it may be useful to refer to studies that have utilized the MarketScan Research Databases. It may also be helpful to examine other references regarding analysis of administrative databases. Since 1988, healthcare researchers have used MarketScan data to understand disease progression, treatment patterns, health outcomes, and their associated costs to patients, employers, health plans, and the government. Fully HIPAA compliant, the MarketScan databases from Truven Health are the gold standard in proprietary databases used for healthcare research in the United States. They are the basis of over 300 peer-reviewed articles published in leading journals since the first article by J.B. Hillman et al. appeared in the New England Journal of Medicine in 1990. MarketScan-based research has made a substantial contribution to the body of literature used to formulate policy decisions and improve healthcare for Americans.

**MarketScan Studies: Abbreviated Bibliography**

**2017**


Calip GS, Adimadhyam S, Xing S, Rincon JC, Lee W-J, Anguiano RH. Medication adherence and persistence over time with self-administered TNF-alpha inhibitors among young adult, middle-aged, and older patients with rheumatologic conditions. Semin Arthritis Rheum 2017; March 21; [Epub ahead of print].


Carney PL, Yao J, Lin J, Law A. Comparison of healthcare costs among commercially insured women in the united states who underwent hysteroscopic sterilization versus laparoscopic bilateral tubal ligation sterilization. J Womens Health (Larchmt) 2017; February 3; [Epub ahead of print].


Idowu OA, Boyajian HH, Ramos E, Shi LL, Lee MJ. Trend of spine surgeries in the outpatient hospital setting vs. ambulatory surgical center. Spine 2017; March 31; [Epub ahead of print].


Khan T, Tsipas S, Wozniak G. Medical care expenditures for individuals with prediabetes: The potential cost savings in reducing the risk of developing diabetes. Popul Health Manag 2017; February 13; [Epub ahead of print].


Lagman C, Ugiliweneza B, Boayke M, Drazin D. Spine surgery outcomes in the elderly versus the general adult patient population in the United States: A MarketScan analysis. World Neurosurg 2017; April 10; [Epub ahead of print].


Liu G, Sterling NW, Kong L, Lewis MM, Mailman RB, Chen H, et al. Statins may facilitate Parkinson’s disease: Insight gained from a large, national claims database. Mov Disord 2017; April 3; [Epub ahead of print].


Lua LL, Vicente ED, Pathak P, Lybbert D, Dandolu V. Comparative analysis of overall cost and rate of healthcare utilization among apical prolapse procedures. Int Urogynecol J 2017; March 31; [Epub ahead of print].


Nichols CI, Vose JG. Incidence of bleeding-related complications during primary implantation and replacement of cardiac implantable electronic devices. J Am Heart Assoc 2017; January 22; [Epub ahead of print].


Preussler JM, Meyer CL, Mau L-W, Majhail NS, Denzen E, Edsall KC, et al. Healthcare costs and utilization for patients age 50 to 64 years with acute myeloid leukemia treated with chemotherapy or with chemotherapy and allogeneic hematopoietic cell transplantation. Biol Blood Marrow Transplant 2017; March 2; [Epub ahead of print].


Schenfeld J, Illes J, Trivedi M, Accortt NA. Dose relationship between oral glucocorticoids and tumor necrosis factor inhibitors and the risk of hospitalized infectious events among patients with rheumatoid arthritis. Rheumatol Int 2017; March 2; [Epub ahead of print].

Sears ED, Momoh AO, Chung KC, Lu Y-T, Zhong L, Waljee JF. A National study of the impact of delayed flap timing for treatment of patients with deep sternal wound infection. Plast Reconstr Surg 2017; April 1; [Epub ahead of print].


Ullal AJ, Kaiser DW, Fan J, Schmitz SK, Than CT, Winkelmayer WC, Heidenreich PA, Piccini JP,


2016


Baggs J, Fridkin SK, Pollack LA, Srinivasan A, Jernigan JA. Estimating national trends in inpatient
antibiotic use among US hospitals from 2006 to 2012. JAMA Internal Medicine 2016; September 19; [Epub ahead of print].


Broder MS, Chang E, Cherepanov D, Neary MP, Ludlam WH. Incidence and prevalence of acromegaly in the United States: A claims-based analysis. Endocrine Practice 2016; August 19; [Epub ahead of print].


Broder MS, Sarsour K, Chang E, Collinson N, Tuckwell K, Napalkov P, Klearman M. Corticosteroid-related adverse events in patients


Canner JK, Giuliano K, Gani F, Schneider EB. Thirty-day re-admission after traumatic brain injury: Results from MarketScan®. Brain Injury 2016; September 2; [Epub ahead of print].

Carney PI, Lin J, Xia F, Law A. Temporal trend in the use of hysteroscopic vs laparoscopic sterilization and the characteristics of commercially insured and Medicaid-insured females in the US who have had the procedures. International Journal of Women’s Health 2016; 8: 137–44.


Clemow DB, Nyhuis AW, Robinson RL. Clinical impact of not achieving recommended dose on duration of atomoxetine treatment in adults with attention-deficit/hyperactivity disorder. CNS Neuroscience & Therapeutics 2016; July 31; [Epub ahead of print].


Encinosa W, Davidoff AJ. Changes in antiemetic overuse in response to choosing wisely recommendations. JAMA Oncology 2016; September 15; [Epub ahead of print].


of Radiation Oncology, Biology, Physics 2016; 94(4): 709–18.


Johnson SP, Chung KC, Zhong L, Sears ED, Waljee JF. Utilization of postoperative radiographs following operative fixation of distal radius fractures. Plastic and Reconstructive Surgery 2016; August 15; [Epub ahead of print].


Kehl KL, Shen C, Litton JK, Arun B, Giordano SH. Rates of BRCA1/2 mutation testing among young...


Li AY, Azad TD, Veeravagu A, Bhatti I, Li A, Cole T, Desai A, Ratliff JK. Impact of inpatient venous thromboembolism continues after discharge: Retrospective propensity scored analysis in a
longitudinal database. Clinical Spine Surgery 2016; October 12; [Epub ahead of print].


Lip GY, Keshishian A, Kamble S, Pan X, Mardekian J, Horblyuk R, Hamilton M. Real-world comparison of major bleeding risk among non-valvular atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban, or warfarin. A propensity score matched analysis. Thrombosis and Haemostasis 2016; August 19; [Epub ahead of print].


Macht R, George J, Ameli O, Hess D, Cabral H, Kazis L. Factors associated with bariatric postoperative emergency department visits. Surgery for Obesity and Related Diseases 2016; March 2; [Epub ahead of print].


Margolis JM, Deitelzweig S, Kline J, Tran O, Smith DM, Crivera C, Bookhart B, Schein J. Pulmonary embolism inpatients treated with rivaroxaban had shorter hospital stays and lower costs compared with warfarin. Clinical Therapeutics 2016; October 14; [Epub ahead of print].


Nichols CI, Vose JG. Clinical outcomes and costs within 90 days of primary or revision total joint arthroplasty. Journal of Arthroplasty 2016; 31(7): 1400–1406.e3.


Obama B. United States Health care reform: Progress to date and next steps. JAMA 2016; 316(5): 525–32.


Painter JE, Collier SA, Gargano JW. Association between Giardia and arthritis or joint pain in a large health insurance cohort: Could it be reactive arthritis? Epidemiology and Infection 2016; September 19; [Epub ahead of print].

Palmer JB, Li Y, Herrera V, Liao M, Tran M, Ozturk ZE. Treatment patterns and costs for anti-


Rose S. A machine learning framework for plan payment risk adjustment. Health Services Research 2016; February 19; [Epub ahead of print].


Saeed MJ, Olsen MA, Powderly WG, Presti RM. Diabetes mellitus is associated with higher risk of developing decompensated cirrhosis in chronic hepatitis C patients. Journal of Clinical Gastroenterology 2016; June 15; [Epub ahead of print].

Sajisevi M, Schulz K, Cyr DD, Woydyla D, Rosenfeld RM, Tucci D, Witsell DL. Nonadherence to guideline recommendations for tympanostomy tube insertion in children based on mega-database claims analysis. Otolaryngology—Head and Neck Surgery 2016; September 13; [Epub ahead of print].
Schreiner A, Simpson K. Primary care and chronic disease: the intersection of comfort and specialty involvement - a cross-sectional study. Journal of Evaluation in Clinical Practice 2016; August 15; [Epub ahead of print].


2015

Accortt NA, Bonafede MM, Collier DH, Iles J, Curtis JR. Risk of subsequent infection among patients using TNF-inhibitors and other DMARDs. Arthritis & Rheumatology 2015; September 11; [Epub ahead of print].


Bonafede M, Johnson BH, Tang DH, Shah N, Harrison DJ, Collier DH. Etanercept-methotrexate combination therapy initiators have greater adherence and persistence than triple therapy initiators with rheumatoid arthritis. Arthritis Care & Research 2015; June 19; [Epub ahead of print].


Bonafede MM, Johnson BH, Richhariya A, Gandra SR. Medical costs associated with cardiovascular events among high-risk patients with


Broder MS, Neary MP, Chang E, Ludlam WH. Incremental healthcare resource utilization and costs in US patients with Cushing's disease compared with diabetes mellitus and population controls. Pituitary 2015; April 5; [Epub ahead of print].


Coleman CI, Kohn CG, Bunz TJ. Derivation and validation of the In-hospital Mortality for


Hess LM, Michael D, Mytelka DS, Beyrer J, Liepa AM, Nicol S. Chemotherapy treatment patterns, costs, and outcomes of patients with gastric cancer in the United States: A retrospective analysis of electronic medical record (EMR) and administrative claims data. Gastric Cancer 2015; March 20; [Epub ahead of print].

Higgins PD, Skup M, Mulani PM, Lin J, Chao J. Increased risk of venous thromboembolic events with corticosteroid vs biologic therapy for...


Kabul S, Alatorre C, Montejano LB, Farr AM, Clemow DB. Real-world dosing patterns of atomoxetine in adults with attention-deficit/hyperactivity disorder. CNS Neuroscience & Therapeutics 2015; August 30; [Epub ahead of print].


Kulaylat AN, Schubart JR, Schaefer EW, Hollenbeak CS, Cooper AB, Gusani NJ. Costs of hepato-pancreato-biliary surgery and readmissions in privately insured US patients. Journal of Surgical Research 2015; May 7; [Epub ahead of print].


Machnicki G, Ong SH, Chen W, Wei ZJ, Kahler KH. Comparison of amlodipine/ valsartan/ hydrochlorothiazide single pill combination and free combination: Adherence, persistence, health care utilization and costs. Current Medical Research and Opinion 2015; September 23; [Epub ahead of print].


cancer diagnosis. Urologic Oncology 2015; August 13; [Epub ahead of print].


Marcus SC, Zummo J, Pettit AR, Stoddard J, Doshi JA. Antipsychotic adherence and rehospitalization in schizophrenia patients receiving oral versus long-acting injectable antipsychotics following hospital discharge. Journal of Managed Care Specialty Pharmacy 2015; 21(9): 754–69.


Nouriaie M, Gordeuk VR. Blood transfusion and 30-day readmission rate in adult patients hospitalized with sickle cell disease crisis. Transfusion 2015; June 30; [Epub ahead of print].


Ovcinnikova O, Panca M, Guest JF. Cost-effectiveness of using an extensively hydrolyzed casein formula plus the probiotic Lactobacillus rhamnosus GG compared to an extensively
hydrolyzed formula alone or an amino acid formula as first-line dietary management for cow's milk allergy in the US. ClinicoEconomics and Outcomes Research 2015; 7: 145–52.


Patel CG, Tao G. The significant impact of different insurance enrollment criteria on the HEDIS chlamydia screening measure for young women enrolled in Medicaid and commercial insurance plans. Sexually Transmitted Diseases 2015; 42(10): 575–9.

Pearson WS, Davis AD, Hoover KW, Gift TL, Owusu-Edusei K, Tao G. Demographic and health services characteristics associated with testing for sexually transmitted infections among a commercially insured population of HIV-positive patients. Journal of Acquired Immune Deficiency Syndromes 2015; June 1; [Epub ahead of print].


Sullivan PW, Campbell JD, Ghushchyan VH, Globe G. Outcomes before and after treatment escalation to Global Initiative for Asthma steps 4 and 5 in severe asthma. Annals of Allergy, Asthma, & Immunology 2015; 114(6): 462–9.

Sun E, Baker LC. Concentration in orthopedic markets was associated with a 7 percent increase in physician fees for total knee replacements. Health Affairs 2015; 34(6): 916–21.


FREQUENTLY ASKED QUESTIONS

Q1. How do individuals track longitudinally across years, plans, and employers?

Truven Health maintains a unique person-level identifier consisting of a family and member identifier. The person-level identifier is consistent across all tables, plans, databases, and years. However, if an employee changes employer and both the previous and new employers are contained in the MarketScan Databases, the family and person-identifiers will change. The family identifiers we receive are encrypted in a different manner for each employer.

Please refer to Section 4, Person-Level Identifiers, for additional information.

Q2. Why do I have a claim where the enrollment flag (ENRFLAG) is set to 1 but the claim does not have an ENROLID?

This typically happens when a piece of information on the claim such as sex, relationship to employee, or date of birth is missing. This usually occurs for less than 1% of claims (EIDFLAG=3).

Q3. How do I identify continuously enrolled covered lives?

To determine if an individual was enrolled for an entire calendar year, MEMDAYS should equal 365. To identify the period of continuous enrollment, use the ENRIND1 to ENRIND12 flags. Each flag corresponds to 1 month (e.g., ENRIND1 = January enrollment, ENRIND2 = February enrollment). The start of continuous enrollment is the first ENRINDx flag that is equal to 1. The end of continuous enrollment is the last ENRINDx flag that is equal to 1.

Q4. How do I select the subset of individuals with outpatient pharmaceutical data?

Analysts may wish to construct a subset of individuals in plans with drug information in each year. These individuals can be identified by the RX flag in the medical/surgical claims, enrollment, and populations tables. The RX flag variable (“1” or “0”) that indicates drug data is available (for the data contributor) in the Outpatient Pharmaceutical Claims Table during that month/year. To select the medical plans with accompanying drug information during a specific month/year, subset to claims where RX=“1.” This flag does not identify individual patients who submitted a drug claim; it is intended to identify records that came from contributing plans that also contribute a drug feed to the MarketScan databases.

Employer contributors (HLTHPLAN=0) will have the same value of RX for each patient for the entire year; Health Plan contributors (HLTHPLAN=1) may have their enrollees’ RX values change from month to month.

Q5. How do I select patients that had both medical and prescription drug claims in the most current year or in the most current 2 years?

Drug data are available for a significant portion of the total medical-eligible population and for a portion of the medical-eligible population with enrollment data. Therefore, you will need to construct a subset of individuals with drug information in each month/year.

The Cohort Drug Indicator (RX) variable indicates whether an individual is covered by a drug plan in the Outpatient Pharmaceutical Claims Table during that month/year. Use this flag (RX=1) to select the medical plans with accompanying drug information. Employer contributors (HLTHPLAN=0) will have the same value of RX for each patient for the entire year; Health Plan contributors (HLTHPLAN=1) may have their enrollees’ RX values change from month to month.

Q6. How do I know whether a patient’s lack of utilization data represents a lack of healthcare use or disenrollment from a plan?

You can match the patient’s utilization to enrollment information in the following way:

Create a subset of Medical and/or outpatient pharmaceutical claims where EIDFLAG=1.

Use ENROLID from the claims utilization as the subset of criteria for the Enrollment data. The resulting subset contains the Enrollment records for
the patients in the corresponding claims.

**Q7. How do I establish a fixed window of continuous enrollment?**

Use the Annual Enrollment tables and subset to records with enrolled months that are within the time window of interest (e.g., all ENRINDx's = 1).

Subset the utilization information (e.g., claims) to SVCDATE within the time window of interest. Sort the utilization information (e.g., claims) by ENROLID. Merge restricted and sorted Enrollment data with sorted utilization information by ENROLID where records appear in both sets.

**Q8. How do I establish a sliding window of continuous enrollment?**

For the sliding window continuous enrollment method, only those persons who actually utilized healthcare can be considered. Therefore, determination of sliding window enrollment status begins with the claims information (medical/surgical or pharmaceutical claims) for identification of the event of interest, and then the enrollment information is considered.

Determine the month and year of the utilization claim of interest. Utilization dates may be a Date Service Incurred (SVCDATE), Date of Admission (ADMDATE), Date Service Ending (TSVCDATE), the beginning of an episode of care, or the end of an episode of care.

Using the enrollment flags (ENRIND1-ENRIND12) in the Annual Enrollment table, determine the earliest and latest dates of continuous enrollment. Create variables to identify these dates. It may be necessary to concatenate multiple years of Annual Enrollment tables. An individual may have multiple continuous enrollment periods.

Merge the utilization data with the enrollment data. Select the time period that includes the utilization date of interest.

If the user is interested in enrollment prior to the utilization date of interest or an ending utilization, then define those dates, and determine if the continuous enrollment period selected includes them.

**Q9. What is the source of the data?**

The MarketScan Databases are constructed from privately insured paid medical and prescription drug claims contributed by employers and health plans who have business relationships with Truven Health. The employers are generally self-insured. Collectively, the databases incorporate data from approximately 350 payers, including commercial insurance companies, Blue Cross and Blue Shield plans, and third-party administrators (TPAs).

Each contributor’s database is constructed by collecting raw data from the appropriate payer(s). These raw data are service-level adjudicated paid claims and capitated encounters containing both inpatient and outpatient services. Financial, clinical, and demographic variables standardized to common definitions and variables that are specific to employers are also added. Clinical detail is added to the Outpatient Pharmaceutical Claims Table (e.g., therapeutic class, therapeutic group, manufacturer’s average wholesale price, and generic product identifier).

**Q10. How are geographic location of the employee (EGEOLOC) and Metropolitan Statistical Area (MSA) determined?**

Geographic Location Employee (EGEOLOC) is mapped from the postal ZIP Code of the primary beneficiary’s residence. Because EGOLOC is often used for rate-based analysis, EGOLOC must reside on both the claims and the enrollment files. If there is some uncertainty in the coding of either source, the EGOLOC values are made more generic than the State level and are set to categories such as Division, Region, or Total U.S.

Metropolitan Statistical Area (MSA) is mapped to Enrollment Detail and Summary, Inpatient Admissions, Inpatient Services, Outpatient Services, and Outpatient Pharmaceutical Claims Tables.

**Q11. Do you ensure that diagnoses, procedures, and demographic information are in concordance with each other?**

Diagnosis and Procedure codes are edited for validity. If they are invalid, they are set to missing.
**Q12. What variables can I use to calculate a rate (e.g., per capita, per employee)?**

Metrics that require a population-based denominator (such as procedures per 1000 covered lives) can only be calculated by selecting on demographic variables that are contained in the Enrollment Table. Typical subsets for such counts include the geographic location of the employee (EGEOLOC), the type of plan (PLANTYP) or the sex of the patient (SEX).

Please refer to the MarketScan Database Enrollment Summary and Detail Tables in the Database Dictionary for a full list of population-supported variables.

**Q13. How do I calculate utilization rates and payments by procedure?**

When calculating a utilization rate by procedure, using the count of claims as the number of procedures overstates the number of procedures. This is because a particular procedure on a given service date can generate more than one claim (e.g., a surgeon’s claim, an anesthesiologist’s claim, and a facility claim). A day-episode for the procedure must be constructed to collapse the related services for each of the procedures of interest.

Using the variable PROC1, subset the Inpatient Services Table and/or the Outpatient Services Tables for the procedures of interest.

To eliminate multiple claims, aggregate the data on ENROLID, PROC1, and SVCDATE in order to create one record per patient, per procedure for a single service date. Sum any other variables of interest (e.g., PAY, NETPAY). The number of procedures performed equals the record count in the resulting subset.

Divide the procedure count by the number of covered lives to calculate a utilization rate.

**Q14 Can a diagnosis be linked to drug claims (and vice versa)?**

The Outpatient Pharmaceutical Claims Table does not contain diagnosis variables, because this information is not provided regularly by the physician on a prescription form. Therapeutic Class (e.g., corticosteroids) is provided on the pharmaceutical claims representing the broad classification of the drug. However, in order to impute the diagnosis, one must access the related medical claims for the individual—usually the claims filed within a specific time window around the prescription:

Subset to the National Drug Code (NDC) of interest on the Outpatient Pharmaceutical Claims Table.

Use ENROLID and SVCDATE as the selection criteria to subset all services from the medical tables (I, S, O) that fall within a pre-defined time window around the SVCDATE. The resulting diagnoses on the medical claims may be associated with the pharmaceutical claim.

These steps may be modified to identify the prescriptions associated with a specific diagnosis. First, subset on a diagnosis in the medical claims. Then, select all pharmaceutical claims for each person with the diagnosis (using ENROLID as the linkage variable) within a pre-defined time window around the date of the prescription.

**Q15. How do I count emergency room (ER) visits, which can occur in the inpatient or the outpatient table?**

The SVCSCAT field can be used to identify most types of service. The field is structured so that the first three digits describe the facility type and the last two digits identify service type. To select Emergency Room visits, select from the S or O table any records with a SVCSCAT value that ends in ‘20’.

Because multiple claim records can be generated for a single ER visit, count the number of ER visits by creating day-episode records from the data table produced by aggregating on ENROLID/SVCDATE combinations. Accumulate all analytic variables of interest.

**Q16. The national drug code in the MarketScan database is 11 digits long, but the codes from my**
Food and Drug Administration (FDA) search are only 10 digits long. How can I convert?

The 10-digit codes should be padded with zeros in the appropriate places until the 11-digit, 5-4-2 format is established. The user needs to pad the number with zeros in the appropriate places until it has the 5-4-2 format. See schematic below:

4-4-2 XXXX-XXXX-XX -> 0XXXXX-XXXX-XX
5-3-2 XXXXXX-XXX-XX -> XXXXX-0XXX-XX
XXXXX-XXXX-X -> XXXXX-XXXX-0X

Q17. How are individuals eligible for Medicare determined in the Medicare database?

Primary contract holders are sorted into the MarketScan Medicare Supplemental and COB Database based on employment status and age. The primary contract holder becomes part of the Medicare Supplemental and COB Database if a record for a primary contract holder indicates either: (1) age of 65 or older, or (2) age 18 or older and employment status of Medicare-eligible retiree.

Dependents are sorted into the MarketScan Medicare Supplemental and COB Database based on age. Dependents aged 65 years or older become part of the Medicare Supplemental and COB Database regardless of the contract holder’s status.

Members of an individual family may be split between the Commercial Claims and Encounters Database and the Medicare Supplemental and COB Database. Users conducting family-based analysis or per-employee rates will need to take this into account.

It is also possible for a single individual to appear both in the Commercial Claims and Encounters Database and the Medicare Supplemental and COB Database if they are a primary contract holder experiencing a change in Medicare-eligible retiree status during the year, or for any member regardless of contract holder status who attains the age of 65 during the year.

Q18. What is the relationship between procedures on the Facility Header (F) table and procedures on the corresponding Inpatient Services (S) or Outpatient Services (O) claim lines?

The MarketScan facility-claims data structure is designed to be similar to the UB04 facility-claim data model. The UB04 claim has a header portion (containing fields reported once per claim) and a revenue center or line-item portion (one or more lines per claim). Multiple ICD-9 or ICD-10 procedures are reported at the header level (once per claim). These correspond to PROC1–PROC6 in the Facility Header table. A CPT-4 or Healthcare Common Procedural Coding System (HCPCS) procedure is reported at the line item level (one per line item). This procedure corresponds to PROC1 in the facility records of the O and S tables.

The rules for what procedures must be reported on a facility claim and where they should be reported (ICD-9 or ICD-10 header or CPT/HCPCS line item) vary based on the type of service, geographic areas, and who is paying for the claim. You may see claims where all procedures are reported only at the header level, others where they are reported only at the line-item level, and still others where they may be reported in both places.

Generally, PROC1–PROC6 on the Facility record should be different from PROC1 on any of the corresponding Outpatient or Inpatient records, because the procedures on the Facility record should be ICD-9 or ICD-10 and the procedures on the O/S records should be CPT/HCPCS. This will not always be true in the MarketScan databases, because not all of the data come from actual UB04 type claims. Some data contributors or suppliers may have CPT/HCPCS procedures on their Facility records.

Q19. Why do some payments show more than two decimal places (e.g., 256.9999999877)?

SAS stores numeric variables in floating point format. Not all values can be represented exactly in floating point format; rather, they can only be approximated. The values of the financial variables in the MarketScan SAS datasets are not formatted (i.e., they do not have a permanent SAS format associated with them). When non-formatted values are printed or displayed by SAS, it is SAS that determines how many decimal places will be shown. If a value can only be approximated, SAS may display many decimal places. This can be overcome by applying either temporary or
permanent formats to the financial variables. For example format 12.2, will display the value with 2 digits to the right of the decimal point and up to 9 digits to the left of the decimal point for a total width of 12 characters (including the decimal point). The value is rounded by the format so that a value that may display unformatted as 123456.499999 will display as 123456.50 formatted. Formatting only affects how the variable is displayed by SAS procedures or viewers; it does not change the stored value.

Q20. How does Truven Health determine which claims get sorted into which data year? Why do I see service dates outside of the calendar year of my data?

Data are included in the database for a given year based on:

- Enrollment: date of enrollment (DTSTART)
- Inpatient Admissions and Inpatient Services: admission date (ADMDATE)
- Outpatient and Drug Claims: service date (SVCDATE)

Inpatient admissions may include inpatient service claims from the day before the admission date. These claims may be for emergency room or pre-admission testing services. The earliest service date for inpatient services claims in a database for a given year is 12/31 of the prior year. Admissions that start late in the year or admissions with very long lengths of stay may have discharge dates that are in the next year. The inpatient services claims that correspond to these admissions will have some service dates that occur in the next year. The facility header claims that correspond to the inpatient services facility claims will also have service dates from 12/31 of the prior year through the next year.
APPENDIX A. NEW IN 2016

In our efforts to continuously improve the analytic value and ease of use of the MarketScan databases, we are pleased to announce the following changes effective with the 2016 v1.0 update.

The Diagnosis Related Group (DRG) variable is assigned using grouper version 34. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

New fields POAPDX and POADX1-POADX15 have been added to the I tables, and POADX1-POADX9 to the F tables, to indicate whether the diagnosis codes appearing in the PDX and DX1-DX15 fields were present on admission. These are character fields of length 1. Valid values are:

- Blank: Missing/Unknown
- 1: Unreported/Not Used
- N: No, not present at admission
- U: Unknown
- W: Clinically Undetermined
- Y: Yes, present at admission

The AGE field is being modified to accommodate increased privacy concerns. Beginning in 2016, AGE will be reported as follows:

- Age 0-6 - actual age as of the Date of Service/Enrollment Start Date/Admission Date. This is unchanged from the current MarketScan format.
- Age 7-16 - age as of the 15th of the month of the Date of Service/Enrollment Start Date/Admission Date.
- Age 17+ - age as of the July 1 of the year of the Date of Service/Enrollment Start Date/Admission Date.

The DSTATUS field is also being modified to accommodate increased privacy concerns. DSTATUS values indicating death or transfer to court/law enforcement (DSTATUS=20, 21, 40, 41, 42, 87) will now be coded as Missing.

The following new Therapeutic Class (THERCLS) values have been added:

- 266 - Antidiabetic Ag, Meglitinides
- 267 - Antidiabetic Ag, SGLT Inhibitr
- 268 - Antidiabetic Ag, TZD
- 271 - Kallikrein Inhibitor
- 272 - COMT Inhibitors
- 273 - Per-Act Mu Op Rcp Ant (PAMORA)
- 290 - Antifungal, EENT

The following Lookup values for Therapeutic Group have been edited to include the new THERCLS values cited above:

- 07 - Cardiovascular Agents (Classes 46-56, 245, 250, 271)
- 08 - Central Nervous System (Classes 57-77, 272)
- 16 - Eye, Ear, Nose Throat (Classes 132-146, 240, 290)
- 17 - Gastrointestinal Drugs (Classes 147-162, 273)
- 20 - Hormones & Synthetic Substitutes (Classes 165-180, 246 252-253 256 266-268)

A new value, 02-Telehealth, has been added to Place of Service (STDPLAC).

A new value, 21420 Enid, OK, has been added to Metropolitan Statistical Area (MSA).
APPENDIX B. HISTORICAL DATA RELEASES

Truven Health strives to deliver consistent data variables from year to year. However, periodic revisions are made to the MarketScan databases to improve and enhance the data. The revisions can include renaming variables or aliases, revising variable definitions, creating new variables, and deleting variables. Following is a list of data changes that could produce anomalies when using several years of data.

Changes in 2015

The Diagnosis Related Group (DRG) variable is assigned using grouper version 33. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

The Populations Table (P) has been discontinued. This table no longer provides value in favor of the Annual Enrollment Summary and Enrollment Detail tables.

The length of all diagnosis code variables (PDX, DX1-DX15) is being increased from 5 to 7 characters to accommodate the implementation of ICD-10-CM.

The length of all procedure code variables (PPROC, PROC1-PROC15) is being increased from 5 to 7 characters.

A diagnosis code version indicator variable (DXVER) is being added to the Facility Header (F), Admissions (I), Outpatient (O), and Inpatient Services (S) tables. DXVER is one character in length and has values “9” = ICD-9-CM and “0” = ICD-10-CM. In the Admissions table, DXVER indicates the ICD version of PDX and DX1. In the other tables DXVER indicates the ICD version of all of the diagnosis code variables in the record.

A new value is being added for the procedure code type variable (PROCTYP) in the Outpatient (O) and Inpatient Services (S) tables to identify ICD-10-PCS procedure codes. This new value is “0” = ICD-10-PCS.

The fields PLANKEY and PLNKEY1-12 are no longer being included. These fields linked to the MarketScan Benefit Plan Design Database, which has been restructured to link on ENROLID effective with the 2015 data year.

The field WGTKEY is no longer being included. This field linked to the MarketScan National Weights which were based on the Medical Expenditure Panel Survey (MEPS). The source for the National Weights has been changed effective with the 2015 data year to the American Community Survey (ACS), and a new field MSWGTKEY has been added to contain the key to link to the new ACS-based MarketScan National Weights.

A new field UNITS has been added to the Inpatient Services and Outpatient Claims tables. This field is intended to capture units of service (as opposed to the quantity of services captured in the QTY field). For example, for an injectable drug, QTY would contain the number of injections while UNITS would contain the amount of substance injected. It is only valued for some data contributors.

A new field MSCLMID has been added to the Inpatient Services, Facility Header and Outpatient Claims tables. This field, when used in conjunction with ENROLID and FACPROF (O and S tables), can enable the user to reconstruct which services were submitted as part of the same claim from a claims administration standpoint.

A new field NPI has been added to Facility Header, Inpatient Services and Outpatient Claims tables. This is an encrypted version of the National Provider Identifier. It is only valued for some data contributors.

Two fields were added to the REDBOOK file to denote the route of administration of a drug: ROACD - Route of Administration Code (2 characters), and ROADS - Route of Administration Description (30 characters). ROADS is the text description for the value appearing in ROACD.

In keeping with changes in CMS coding, the Outpatient Hospital place of service has been split into two parts.
• STDPLAC value 19 has been added as “Outpatient Hospital-Off Campus”.

• The lookup for STDPLAC value 22 has been changed to “Outpatient Hospital-On Campus”.

• A new value for THERCLS has been added. The new value is 259 “Blood Form/Coagul Agents, Misc”.

• The lookup for THERGRP value 06 has been changed to “Blood Form/Coagul Agents (Classes 35-45, 259)”.

**Changes in 2014**

The Diagnosis Related Group (DRG) variable is assigned using grouper version 32. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

Six new values for Therapeutic Class (THERCLS) have been added:

260—Interferons, Antineoplastic
261—Chemotherapy
262—Hormone-Modifying Therapy
263—Molecular Targeted Therapy
264—Radiopharmaceu/Antineoplastic
265—Antineoplastic Agent, Misc.

The mapping of the MSA field has been updated to conform with the US Census Bureau’s 2013 mapping. Some slight changes have occurred regarding the boundaries of individual MSAs as well as to the labels applied to them. For a complete listing of updated MSAs effective with the 2014 data year, see the Data Dictionary.

**Changes in 2013**

The Diagnosis Related Group (DRG) variable is assigned using grouper version 31. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

A new value for STDPLAC (Place of Service) has been added: 18-Place of Employment-Worksite.

New and revised values for DSTATUS (Discharge Status): the descriptions for values 50, 51 have changed, and values 69, 81-95 have been added:

50—Discharged/Transferred to Hospice medical facility
51—Discharged/Transferred to Hospice medical facility
69—Transfer to disaster alternative care site
81—Discharge to home/self care w plan IP readmit
82—Transfer to short-term general hosp w/ plan IP readmit
83—Transfer to SNF w/ plan IP readmit
84—Transfer to custodial/supportive care w/ plan IP readmit
85—Transfer to cancer center/child hosp w/ plan IP readmit
86—Transfer to home health service w/ plan IP readmit
87—Transfer to court/law enforce w/ plan IP readmit
88—Transfer to federal facility HCF w/ plan IP readmit
89—Transfer to Medicare swing bed w/ plan IP readmit
90—Transfer to IRF w/ plan IP readmit
91—Transfer to LTCH w/ plan IP readmit
92—Transfer to Medicaid nursing facility w/ plan IP readmit
93—Transfer to psych unit/hospital w/ plan IP readmit
94—Transfer to CAH w/ plan IP readmit
95—Transfer to other facility NEC w/ plan IP readmit

**Changes in 2012**

The Diagnosis Related Group (DRG) variable is assigned using grouper version 30. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

The YEAR field is being added to the Enrollment Detail (T) table. It currently appears on all other claims and enrollment tables.

**Changes in 2011**

Periodically we subject the MarketScan Databases to review by an external independent consultant to ensure that the databases are indisputably categorized as deidentified data with respect to HIPAA. We have just completed such a review, and as a result we are making the following changes to the Commercial and Medicare Supplemental data structure, effective with the 2011 v1.0 database release in December. These changes will be reflected in all database releases moving forward.
1) The following geographic variables will no longer be included:
   a. EMPCTY (County Employee), HOSPCTY (County Hospital), PHRMCTY (County Pharmacy), PROVCTY (County Provider)
   b. EMPZIP (3-digit Zip Code Employee), HOSZIP (3-digit Zip Code Hospital), PHRMZIP (3-digit Zip Code Pharmacy), PROVZIP (3-digit Zip Code Provider)

All other geographic variables (MSA, state, and region) remain

2) The following clinical/provider variables will no longer be included:
   a. UNIHOSP (Standard Hospital ID)
   b. STDSVC (Service Type)

Provider ID (PROVID) remains in the database. Service subcategory code (SVCSCAT), a more current and detailed variable grouping of services, also remains in the database.

3) The EGEOLOC (Geographic Location Employee) field will no longer report values of 98 (Virgin Islands) and 99 (Other International). Records for these values will be recoded to EGEOLOC=1 (Nation Unknown Region)

4) The STDPLAC (Place of Service) field will no longer report values of 5 (Indian Health Services Free Standing Facility), 6 (Indian Health Services Provider-Based Facility), 7 (Tribal 638 Free Standing Facility), 8 (Tribal 638 Provider-Based Facility) or 9 (Prison-Correctional Facility). Records for these values will be recoded to STDPLAC=99 (Other Unlisted Facility).

5) A Family Identifier field (EFAMID) will be added. This will enable users to study family members enrolled together under a single subscriber policy.

Protecting the health information of the individuals represented in the MarketScan Databases is central to maintaining our ability to offer the data to our customers and is at the forefront of our priorities as we continue to improve and enhance the data.

The diagnosis-related group variable (DRG) is assigned using grouper version 29. This variable appears on the I and S tables. Searches are included in the SAS format file delivered with the databases.

**Changes in 2010**

The Diagnosis Related Group (DRG) variable is assigned using grouper version 28. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

**Changes in 2009**

The Diagnosis Related Group (DRG) variable is assigned using grouper version 27. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

DX3 and DX4 (Diagnosis Code 3 and Diagnosis Code 4) have been added to the S and O tables.

INDSTRY (Industry Code) has three new values. These values are A: Agriculture, Forestry, Fishing; C: Construction; and W: Wholesale.

**Changes in 2008**

The Diagnosis Related Group (DRG) variable is assigned using grouper version 26. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

Therapeutic Class (THERCLS) has two new values. These only appear in RED BOOK and are not yet present in the claims data. The new values are 248-Leukotriene Modifiers, and 249-Uricosuric Agents.

Plan Indicator (PLANTYP) has a new value of 9, representing High Deductible Health Plan (HDHP).

**Changes in 2007**

The Diagnosis Related Group (DRG) variable is assigned using grouper version 25. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

Three new variables have been added. Capitated Service-Claim Indicator (CAP_SVC) is an indication of whether the individual service or claim was paid on a capitated basis. Valid values are ‘Y’ for Yes if the claim was paid on a capitated basis and ‘N’ for No if the claim was not paid on a capitated basis. This field appears on the D, F, O, and S tables.
Network Provider Indicator (NTWKPROV) is an indication of whether the provider of an individual service was a member of the payer’s network. Valid values are ‘Y’ for Yes if the provider was a member of the network and ‘N’ for No if the provider was not a member of the network. This field appears on the D, F, O, and S tables.

Network Paid Indicator (PAIDNTWK) is an indication of whether an individual claim was paid as in-network. Valid values are ‘Y’ for Yes if the claim was paid as in-network and ‘N’ for No if the claim was not paid as in-network. This field appears on the D, F, O, and S tables.

Changes in 2006
The following changes were effective with the 2006 v1.0 update.

The Diagnosis Related Group (DRG) variable is assigned using grouper version 24. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

The Revenue Code (REVCODE) variable has changed from 3-digit codes to 4-digit codes. This variable appears on the O and S tables. Both 3- and 4-digit values are included in the SAS format file delivered with the databases.

The Diagnosis Related Group (DRG) variable is assigned using grouper version 23. This variable appears on the I and S tables. Lookups are included in the SAS format file delivered with the databases.

The RX[year] and PHY[year] variables have been removed from the A and T tables. Instead, the variables RX and PHYFLAG have been added to the A and T tables. The year-specific flags were originally implemented when Enrollment information was delivered in a cumulative, all-time-period table; because the format was changed to one enrollment table per database year, these year-specific variables are no longer necessary.

The Payments Total Case (TOTPAY) variable has been dropped from the S table. It still appears on the I table and is easily associated with the individual services of an inpatient admission using the CASEID variable.

A new variable, Truven Health Service Sub-Category Code (SVCSCAT) has been added to O and S tables. The lookup for this new field appears in the Data Dictionary. SVCSCAT is a highly detailed service category code with over 570 values.

Changes in 2004
Data Expansion: Inclusion of Health Plan Data Contributors.

The 2004 MarketScan files include data obtained from our Health Plan contributors, combined with the data from our Employer customers. Two new variables have also been added to the data files.

Historically, we have delivered data from contributors capturing full carve-out services. In the current deliverable, contributors with potentially incomplete Mental Health and Substance Abuse (MHSA) coverage have been included in the data files. To identify enrollees in plans where MHSA coverage may not be fully captured, we have included a MHSA coverage variable (MHSACOVG). This variable can be used to exclude enrollees from mental health-related analyses or to further investigate the utilization rates of the sub-population.

To easily identify which enrollees come from our new Health Plan data contributors, we have created a health plan indicator variable (HLTHPLAN). This variable allows the user to distinguish between data source types; it is set to 1 for Health Plan lives and 0 for employer lives.
Note: Health Plan data contributors were also added retrospectively to the 2002 and 2003 data years. Missing values of MHSACOVG in these years should be interpreted as “1,” and missing values of HLTHPLAN in these years should be interpreted as “0.”

**New Fields**

Health Plan Indicator (HLTHPLAN): I,S,O,D,P,A,T

Coverage Indicator MHSA (MHSACOVG): I,S,O,D,P,A,T

New SAS formats

The format listing has been updated and new formats have been included for the new categorical fields.

**New DRG Grouper Version**

The 2004 release uses Diagnosis Related Group (DRG) Grouper 22.0. The 2003 MarketScan Databases used Grouper 21.0.

**Changes in 2003**

**New Table: Facility Header (F)**

The records in this table represent facility claim information that occurs at the overall claim level (once per claim). (Facility records in the O and S tables represent facility claim detail lines at the UB04 revenue center or individual service level.) Facility header variables include: 9 ICD-9-CM diagnosis codes (DX1-DX9), 6 ICD-9-CM procedure codes (PROC1-PROC6), Net Payment Amount (NETPAY), Copayment Amount (COPAY), Deductible Amount (DEDUCT), COB Amount (COB), Coinsurance Amount (COINS), UB04 Bill Type (BILLTYP), Facility ID (PROVID and UNIHOSP), Place of Service (STDPLAC) and Provider Type (STDPROV). The facility header financial variables repeat the amounts contained in those variables in the facility detail records in the O and S tables. Facility header records may be linked to their corresponding facility detail records in the O and S tables by the FACHDID variable that appears in the F, I and S tables. (FACHDID is missing in the F table for non-inpatient claims.)

The inclusion of the F table allows users to access up to 9 diagnosis and 6 procedure variables on a facility claim (as opposed to the 5 diagnosis and 1 procedure currently retained in the S and O tables). The inclusion of the F table provides an easier correspondence of complete diagnosis and procedure associated with facility detail records.

The new F table renders the DX3-DX5 fields on the S and O tables superfluous, so these have been removed.

See the MarketScan Data Dictionary for a complete listing of fields included on the F table. All but one (BILLTYP) currently appear on other tables.

**New Table: Annual Enrollment Summary (A)**

A new annual enrollment summary (A) table is included in the CCAE and MDCR Databases. This table is structured as one record per person (ENROLID) enrolled during the year. The annual summary contains monthly arrays of certain variables such as indicators of enrollment (yes/no), days enrolled, data type, and plan type in each month during the year. There are also variables indicating the number of months during the year with enrollment and the total annual enrollment days.

Demographics variables in the new A table fall into two categories.

Monthly arrays—twelve fields give the value of the variable applicable for each month during the calendar year. This treatment is used for the DATATYP, PLANTYP and PLANKEY fields (DATATYP1-DATATYP12, PLANTYP1-PLANTYP12, PLNKEY1-PLNKEY12).

Modal values—the value that appears in the largest number of enrollment months during the year. (This is how the values of these variables are set in the current enrollment summary [E] table.) This treatment is used for fields such as MSA, EECLASS, etc.

The current monthly enrollment detail (T) table for a year of data will continue as currently structured.
The enrollment summary (E) table as it appears in data releases 2002 and prior will no longer be produced.

**New Fields**

- Facility Bill Type Code (BILLTYP): F
- Date Service Ending (TSVCDAT): Historically included on the S table, it now also appears on the O table.
- Coinsurance (COINS): S,O,F,D
- Date of Discharge (DISDATE): I,S
- Facility Header Record ID (FACHDID): S,O,F
- Facility-Professional Claim Indicator (FACPROF): S,O
- Net Payments Hospital (HOSPNET): I
- Net Payments Physician (PHYSNET): I
- COB and Other Savings Total Case (TOTCOB): I
- Coinsurance Total Case (TOTCOINS): I
- Copayment Total Case (TOTCOPAY): I
- Deductible Total Case (TOTDED): I

**Fields Removed**

- Diagnosis 3 through Diagnosis 5 (DX3–DX5) removed from S and O tables only
- Days from Prior Discharge (LASTADM)
- Days to Next Admission (NEXTADM)
- Payment Indicator (PAYIND)
- Primary Care Physician ID Number (PCPID)
- Primary Care Physician Specialty (PCPSPEC)
- Physician Classification (PHYCLAS)
- Primary Medical Group ID (PMGID)
- Record Flag (RECFLAG)
- Referral Indicator (REFIND)
- Referral Type (REFTYP)
- Treatment Group (TG)
- Trim Flag Length of Stay (TRIMLOS)
- Trim Flag Per Diem (TRIMPDM)

**New SAS formats**

The format listing has been updated and new formats have been included for the new categorical fields. Formats for fields no longer delivered have been removed. There are also some new values for STDPLAC and THERCLS.

**New DRG Grouper Version**

The 2003 release uses Diagnostic Related Group (DRG) Grouper 21.0. The 2002 MarketScan Databases used Grouper 19.0.

**Changes in 2002**

The 2002 Commercial Claims and Encounters and Medicare Supplemental and COB databases are larger this year because of our ability to add several new data contributors. The datasets now represent 25% to 50% more covered lives than in 2001.

We have implemented an audit of the Length of Stay (DAYS) field on the Inpatient Services (S) Table. Previously, there was a possibility of discrepancy between the sum of service-level DAYS for an inpatient admission and the length of stay listed on the corresponding admission record in the Inpatient Admissions table. We have been able to correct the discrepancy so that approximately 90% of admissions will have no discrepancy between length of stay on the admission- and service-level records and an additional 5% will be within 1 or 2 days. The remaining 5% are not correctable and we recommend using the admission-level length of stay in those instances.

**Changes in 2001**

**DRG Grouper Update**

The 2001 release uses Diagnostic Related Group (DRG) Grouper 19.0. The 2000 MarketScan Databases used Grouper 17.0.

**Encryption of Provider Fields**

The provider identifying fields in the MarketScan Databases are now encrypted to better protect the confidentiality of the data contributors. The fields affected are Provider ID, Pharmacy ID, Uniform Hospital ID TRUVENT HEALTH, Physician ID,
Primary Care Physician ID, Primary Medical Group ID.

Rx Mail Order-Retail Indicator Field

A new field has been added to the Outpatient Pharmaceutical Claims file, Rx Mail Order-Retail Indicator (RXMR), to denote whether the prescription was filled by a retail pharmacy or through a mail-order program.

National Weights

MarketScan person-level national weights are constructed using the Household Component of the Medical Expenditure Panel Survey (MEPS). MEPS provides estimates of the number of people with employer-sponsored private health insurance. The estimates are used to weight individuals in MarketScan to reflect the national ESI distribution, as captured by the most relevant year of MEPS data.

To construct the weights, MEPS respondents are stratified using combinations of demographic variables that account for substantial differences in utilization and expenditures. They include:

Region (Northeast, North Central, South, West)
Age (three groups: 0–17, 18–44, 45–64)
Sex (male, female)
MSA classification (MSA, non-MSA)
Insurance policy holder status (policy holder, spouse/dependent).

Not all combinations of these demographic categories are used. We collapse the policyholder/non-policyholder status for non-MSA strata in the Northeast and West regions because of small cell sizes for these areas. We do not distinguish between policyholder and non-policyholder for the 0–17 age group. In all, 72 strata are used to construct the weights.

The person-level weight is the ratio of MEPS-based estimates in the different age/sex/region categories and the MarketScan number in the same category.

Note: Person-level weights are assigned to the MarketScan data on all tables by means of a numeric key pointer (WGTKEY) to a stand-alone table of weights values. The weights table itself is not delivered as part of the standard MarketScan databases. Interested parties should contact Truven Health regarding purchase of the weights table.

Change in Medicare-Eligible Classification Methodology

Primary contract holders are sorted into the MarketScan Medicare Supplemental and COB Database based on Employment Status. If a record for a primary contract holder indicates Medicare Eligible Retiree, the primary contract holder becomes part of the MarketScan Medicare Supplemental and COB Database.

Dependents are sorted into the MarketScan Medicare Supplemental and COB Database based on age. Dependents aged 65 years or older become part of the Medicare Supplemental and COB Database regardless of the contract holder’s status.

Members of an individual family, therefore, may be split between the MarketScan Commercial Claims and Encounters Database and the Medicare Supplemental and COB Database. Users conducting family-based analysis or per-employee rates will need to take this into account.

Previously, the data were divided based on the age and employment status of the primary contract holder; thus, non-Medicare eligible dependents of Medicare-eligible contract holders formerly appeared in the Medicare Supplemental and COB Database, and Medicare-eligible dependents of non-Medicare eligible contract holders formerly appeared in the Commercial Claims and Encounters Database.

Enrollment File Structure Change

Beginning with the 2001 data release, the Enrollment Detail Table changed in structure. A single record will represent one month of enrollment for an individual. Persons enrolled continuously for the entire calendar year 2001 will have 12 records in the 2001 Enrollment Detail Table. Databases will be delivered with monthly enrollment records that are applicable to that particular database; time periods of enrollment prior to the time period of the medical claims data will no longer be included.

The structure of the Enrollment Summary Table has not changed, but the file will now only contain enrollment records for calendar year 2001 with one
record per period of continuous enrollment per enrollee and the prevailing demographics. Continuously enrolled individuals will have one record in the Enrollment Summary Table; however, enrollees may still have multiple records per year in the summary file if they have discontinuous enrollment.

Addition of Age and Age Group to Enrollment Tables

Age and Age Group of each enrollee now appears on the Enrollment Detail and Summary Tables. This represents age as of the start of the enrollment period indicated on the record.

Addition of MSA to the Populations Table

The Metropolitan Statistical Area field (MSA) is now valued wherever possible on the Populations Table.

Deleted Identifier Fields

The family identifiers and member identifiers of both the Enrollee and Patient identification systems (EFAMID, EMEMID, FAMID, MEMID) have been removed. This is to conform with the requirements of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and to reduce the risk of implicit patient identification through other demographic fields.

Changes in 2000

New Variables

Dx3, Dx4, and Dx5 (S, O)

Procedure code modifiers and revenue codes are now available for a subset of MarketScan data contributors:

- Procedure Code Modifier (S, O). A procedure code look-up file (including CPT and modifier codes) is available upon execution of the American Medical Association (AMA) CPT license agreement.
- Revenue Code (S, O). A revenue code look-up file is included on the documentation CD.

Variable Changes

Standard Place of Service (STDPLAC) and Standard Provider Type (STDPROV) have new values that are now consistent with new Truven Health company-wide standards. Place of Service values now correspond to CMS standard values. Provider Type values were expanded in order to represent the breadth of provider types now covered by medical benefit programs. We have provided a map of old values to new values for your convenience.

Facility, professional, and other providers are now identified according to the following values:

001–099 Facilities
100–799 Physicians
100–199 Non-admitting Physicians
200–799 Admitting Physicians
500–599 Surgeons
800–899 Professionals (Non-Physician)
900–999 Agencies

Financial variables: Effective with the 2000 data year, Truven Health has a new standard format for financial data. Inpatient, outpatient, and prescription drug financial variables are now represented in dollars and cents with an explicit decimal point. Some customer databases continue to reflect financial data for inpatient and outpatient claims in whole dollars. The percentage of these claims will diminish over time. Databases delivered in SAS format will contain the explicit decimal point. There will be no change in field format for databases delivered in DataProbe®.

PPROC Assignment: The new standard is to assign the principal procedure (PPROC) only when the procedure is part of the DRG/MDC assignment. PPROC will have missing values when the DRG/MDC is for a non-surgical admission.

Quarterly Updates Released

In an effort to release to our customers the most current data available while still maintaining the highest level of data quality, MarketScan data releases follow a quarterly schedule. Only data contributors with at least 3e months of paid runoff (the lag time between a service being incurred and a claim being paid) are included with each interim quarterly release. Each December we will continue to release a complete version of the prior year’s
data, with at least 6 months paid runoff (considered to be analytically complete).

Quarterly updates are released in March, June, September, and December. These databases include all tables that normally are found in a yearly database: Inpatient Admissions, Inpatient Services, Outpatient Services, Outpatient Pharmaceutical Claims, and Enrollment. The Benefit Plan Design Database is released annually.

Included with each quarterly update is a Quarterly Comparison Report, which shows changes in overall covered lives, continuous covered lives, claim volume by quarter, and claim payments by quarter. The volume of each quarterly data release depends upon the update cycle of the individual data contributors and the level of completeness of the data.

Enrollee Identifier Transition

Historical MarketScan databases contain two sets of person identifiers. Enrollee identifiers (ENROLID, EMEMID, EFAMID) were derived solely from eligibility data prior to 1999. These identifiers then were assigned to corresponding claims using the eligibility data as the source. Patient identifiers (PATID, MEMID, FAMID), which identify unique claimants, are based on information available on the claim without reference to an eligibility record. The use of these identifiers has not been straightforward and we have taken steps to simplify their use.

With the 1999 MarketScan data release, we began a new system of person identification that, over time, will eliminate the need to maintain two types of identifiers. MarketScan data now contain an enrollee identifier that is assigned to all patients regardless of whether enrollment data are present. This “universal” identifier provides continuous person identification for data contributors with prior years of enrollment data in the MarketScan databases and is more reliable than the historical patient identifier (PATID) assignment method. For data contributors without enrollment data (about 9% of covered lives in 2000), an enrollee identifier is derived. A person identifier flag variable (EIDFLAG) describes the source and quality of the enrollee identification derivation and assignment. The method for deriving the enrollee identifier differs depending upon whether enrollment or claims data are used and whether the data contributor reports patient date of birth on the claim.

The current patient identifier variables (PATID, MEMID, FAMID) are being maintained for an indefinite period for compatibility with prior year deliverables and analyses. We plan to replace these variables entirely with the universal enrollee identifier variables when practical for our database users.

Please refer to Section 4: Person-Level Identifiers for further information on the development of the enrollee identifier variables.

Changes in 1999

Adjustment Records

Adjustment records result from corrections made to a paid claim. These records may contain negative amounts in financial or other variables (e.g., QTY). Historically, the MarketScan databases have applied an adjustment algorithm to claims on the Outpatient Services Table in an effort to resolve records with negative financial amounts. This algorithm combines financials on the original record with financials on the adjustment record. The financial variables used are PAY, DEDUCT, COPAY, COB, NETPAY.

This year, the adjustment algorithm was reviewed and applied to the Outpatient Pharmaceutical Claims Table. Some negative records remain. These records represent voided claims where the original claim is missing. Users should use discretion in deleting these “orphan” voids, because these were intended to cancel other positive values where we could not link the void and original.

DRG Grouper 17.0

DRG values are now assigned using HCFA Grouper 17.0 values. Sixteen new values have been added.

New Table

An extensive list of RED BOOK variables is now available on the MarketScan databases. These variables have been included in a separate table (RED BOOK Supplement) to enhance prescription drug analyses. Licensed users of MarketScan Research Databases may use these variables to
develop internal reports. The RED BOOK variables are linked to the Outpatient Pharmaceutical Table by the National Drug Code (NDC). Many RED BOOK variables have text lookup values in corresponding “description” variables, allowing text searches. We have removed the NDCNUM1 and NDCNUM2 variables from the Outpatient Pharmaceutical Claims Table because manufacturer, product name, and package size information can now be linked from the RED BOOK Table.

Truven Health licenses the following variables from RED BOOK:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEACLAS</td>
<td>DEA Class Code</td>
</tr>
<tr>
<td>DEACLDS</td>
<td>DEA Class Description</td>
</tr>
<tr>
<td>DESIDRG</td>
<td>DESI Drug Indicator</td>
</tr>
<tr>
<td>EXCDGDS</td>
<td>Exceptional Drug Description</td>
</tr>
<tr>
<td>EXCLUDRG</td>
<td>Exceptional Drug Indicator</td>
</tr>
<tr>
<td>GENERID</td>
<td>Generic Product ID</td>
</tr>
<tr>
<td>GENIND</td>
<td>Generic Indicator</td>
</tr>
<tr>
<td>GENNME</td>
<td>Generic Drug Name</td>
</tr>
<tr>
<td>GNINDDS</td>
<td>Generic Indicator Description</td>
</tr>
<tr>
<td>MAINTDS</td>
<td>Maintenance Indicator Description</td>
</tr>
<tr>
<td>MAINTIN</td>
<td>Maintenance Indicator</td>
</tr>
<tr>
<td>MANFNME</td>
<td>Manufacturer Name</td>
</tr>
<tr>
<td>MASTFRM</td>
<td>Master Form Code</td>
</tr>
<tr>
<td>METSIZE</td>
<td>Metric Size</td>
</tr>
<tr>
<td>MSTFMDS</td>
<td>Master Form Description</td>
</tr>
<tr>
<td>NDCNUM</td>
<td>National Drug Code</td>
</tr>
<tr>
<td>ORGBKCD</td>
<td>Orange Book Code</td>
</tr>
<tr>
<td>ORGBKDS</td>
<td>Orange Book Code Description</td>
</tr>
<tr>
<td>ORGBKFG</td>
<td>Orange Book Standard Flag</td>
</tr>
<tr>
<td>PKQTYCD</td>
<td>Package Quantity Code</td>
</tr>
<tr>
<td>PKSIZE</td>
<td>Package Size</td>
</tr>
<tr>
<td>PRDCTDS</td>
<td>Product Category Description</td>
</tr>
<tr>
<td>PRODCAT</td>
<td>Product Category Code</td>
</tr>
<tr>
<td>PRODNME</td>
<td>Product Name</td>
</tr>
<tr>
<td>SIGLSRC</td>
<td>Single Source Indicator</td>
</tr>
<tr>
<td>STRNGTH</td>
<td>Strength</td>
</tr>
<tr>
<td>THERCLS</td>
<td>Therapeutic Class</td>
</tr>
<tr>
<td>THERDTL</td>
<td>Therapeutic Detail Code</td>
</tr>
<tr>
<td>THERGRP</td>
<td>Therapeutic Group</td>
</tr>
<tr>
<td>THRCLDS</td>
<td>Therapeutic Class Description</td>
</tr>
<tr>
<td>THRDTDS</td>
<td>Therapeutic Detail Code</td>
</tr>
<tr>
<td>THRGRDS</td>
<td>Therapeutic Group Description</td>
</tr>
</tbody>
</table>

Database Renaming

1998
The databases formerly known as “Private Pay Fee-for-Service and Encounter” have been combined and renamed to the MarketScan Commercial Claims and Encounters Database. The “Medicare Database” has been renamed the “MarketScan Medicare Supplemental and COB Database.”

Introduction of New Variables

1999
A five digit State-county variable is now available describing the county of the employee, hospital, provider, and pharmacy. These variables are based on FIPS State code and county name, where State code is two digits and FIPS county code is three digits. For example, 06013 where 06 = California and 013 = Contra Costa county.

Other new variables are:

- County Employee (EMPCTY)
- County Hospital (HOSPCTY)
- County Provider (PROVCTY)
- County Pharmacy (PHRMCTY)
- Enrollee ID Derivation Flag (EIDFLAG) describes the source of data used to assign ENROLID, EFAMID, and EMEMID as well as the quality of that assignment.
- Date Claim Paid (PDDATE) is assigned to the Inpatient Services, Outpatient Services, and Outpatient Pharmaceutical Claims Tables.
- Diagnosis15 (DX15) replaces DX_N. DX_A through DX_N have been renamed DX2 through DX15. DX1 is now the PDX.
- Procedure15 (PROC15) replaces PROC_N. PROC_A through PROC_N have been renamed PROC2 through PROC15. PROC1 is now the PPROC.
- REGION has been added to the Enrollment Tables.
- Cohort Drug Indicator (RX) has been added to the Populations Table, replacing the three RX(CCYY) variables.
Data Type (DATATYP): Encounter and fee-for-service data now reside in the same database. A data type variable has been created to allow users to easily identify and deal with these data in analyses. DATATYP=1 or 2 identifies fee-for-service and encounter records, respectively, in the Commercial Claims and Encounters Database. DATATYP=3 or 4 identifies fee-for-service and encounter records, respectively, in the Medicare Supplemental and COB Database.

Payment Indicator (PAYIND): Payments in-network and out-of-network for individuals enrolled in managed care plans with network incentives can now be examined. Payment In/Out of Plan values are as follows:

1=Paid in plan; in-plan provider
2=Paid in plan; out of area
3=Paid in plan; referred out
4=Paid in plan; other
5=Paid out-of-plan (opt-out)

Bundled Deliveries Flag (BUNDELV): A flag indicating that some data contributors may bundle infants’ claims with their mothers for normal deliveries; hence, there may be no separate record for the newborn in the Inpatient Admissions or Services Tables (appears only in 1997 data).

Enrollment_Flag (ENRFLAG): May be used to subset data to only those patients and persons eligible for coverage from data contributors for whom we have enrollment information. This flag is available on the Inpatient Admissions, Inpatient Services, Outpatient Services, Outpatient Pharmaceutical Claims, and Populations Tables.

Physician Specialty Coding Flag (PHYFLAG): May be used to subset to data with highly differentiated physician specialty coding (>70%) on claims. This flag is available on the Inpatient Admissions, Inpatient Services, Outpatient Services, Outpatient Pharmaceutical Claims, Populations, and Enrollment Tables.
Therapeutic Group (THERGRP)

1998


1997

Therapeutic Class (THERCLS): See Data Dictionary, Attachment K.
Therapeutic Group (THERGRP): See Data Dictionary, Attachment L.

Maintenance Indicator (MAINTIN):

New Values (1997 forward):
1: Used primarily for long-term treatment of chronic conditions
2: Used for both chronic and acute conditions
3: Used primarily for short-term treatment of acute conditions
4: Other/unavailable

Old Values (prior to 1997):
1: Maintenance drug

Pharmacy Class (PHCLASS):

New Values (1997 forward):
0: Other
1: Independent
2: Chain
3: Hospital
4: Clinic
5: Franchise

Old Values (prior to 1997):
1: Community Pharmacy
2: Chain Pharmacy (4+ stores)
3: Hospital Pharmacy
4: Clinic Pharmacy
5: Nursing home/Ext Care Pharmacy
6: Department Store Pharmacy
7: Grocery Store Pharmacy
8: Other

Generic Indicator (GENIND)

New Values (1997 forward):
1: Single source brand
2: Not used
3: Brand name, generic available
4: Multisource generic
5: Single source generic
6: Over the counter
7: Other/unavailable

Old Values (prior to 1997):
1: Brand—Single Source
2: Brand—Multi Source
3: Original Product—Generic Available
4: Generic Product

1996 and subsequent years

The missing value for ENROLID is actually ‘missing’ for individuals in data contributors/plans without enrollment information. Prior to 1996, all individuals not receiving an enrollee ID were assigned an ENROLID of all zeroes (i.e., 00000000000).

1995 and subsequent years

Diagnosis_A through Diagnosis_N and Procedure_A through Procedure_N are true secondary codes in the 1995 data and subsequent years. Previously, these variables could contain the primary diagnosis or procedure code as well as secondary codes.

On the Outpatient Pharmaceutical Claims Table, the financial variables contain amounts accurate to the penny. The enhancement was made to achieve greater accuracy when handling small charge or payment amounts. On prior years of data, the financial variables on the Outpatient Pharmaceutical Claims Table could contain whole dollar amounts.

1994 and subsequent years

The number of valid definitions for Plan Indicator (PLANTYP) increased from four to seven for 1994 forward. (Refer to the Data Dictionary for the valid values.)
Variable Renames

1999

DX_A through DX_N is now DX2 through DX15 where DX1 is the PDX.

PROC_A through PROC_N is now PROC2 through PROC15 where PROC1 is the PPROC.

On prescription drug variables, the P suffix has been removed from financial variables to simplify variable naming:

- AWPP is now AWP.
- COBP is now COB.
- COPAYP is now COPAY.
- DEDUCTP is now DEDUCT.
- DISPFEP is now DISPFEE.
- INGCSTP is now INGCOST.
- NETPAYP is now NETPAY.
- PAYP is now PAY.
- SALETXP is now SALETAX.

CASEINP/INP

There is now one variable to identify a hospital admission and its related services:

- CASEINP has been renamed CASEID
- INP has been renamed CASEID

1996

<table>
<thead>
<tr>
<th>New Variable Name</th>
<th>Old Variable Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDCCYY*</td>
<td>MEDYY</td>
</tr>
<tr>
<td>RXCCYY</td>
<td>RXYY</td>
</tr>
</tbody>
</table>

CCYY represents the century and year (e.g., 1997).

Variable Renames in DataProbe

In DataProbe, the following variable aliases were renamed with the 1995 database and subsequent years. The variable definitions have not changed.

<table>
<thead>
<tr>
<th>New Variable Name</th>
<th>Old Variable Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX</td>
<td>SEX</td>
</tr>
<tr>
<td>PLANTYP</td>
<td>TYPE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>New Variable Name</th>
<th>Old Variable Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>PATID</td>
<td>PATNT</td>
</tr>
<tr>
<td>MEDyy*</td>
<td>CMEDyy*</td>
</tr>
<tr>
<td>Rxyy*</td>
<td>CDRUGyy*</td>
</tr>
</tbody>
</table>

* yy represents specific year of data

Deletion of Variables

1999

NDCNUM1 and NDCNUM2 have been removed from the Outpatient Pharmaceutical Claims Table and are now available in the Truven Health RED BOOK Table.

MED(CCYY) variables have been eliminated from the database to simplify use. Information on whether the data contributor had medical in a specific year can be derived from the Enrollment Tables.

RX (CCYY) variables have been removed from the Populations Table. In their place, the Cohort Drug Indicator (RX) describes plans with available drug data in 1999. RXCCYY variables (RX1993 to RX1999) are present in the Enrollment Tables for CCAE, and RX1998 and RX1999 are present for Medicare Supplemental and COB data. These variables allow users to subset on enrollees with prescription drug claims for those years.

1998

The Bundled Deliveries Flag (BUNDELV) variable has been eliminated. This variable was delivered for the first time in the 1997 research databases and was intended to indicate claims with a bundled charge for the baby and the mother during normal deliveries. After careful review, we have concluded that the data needed to accurately develop this variable are not available for all data contributors.
1996

The State_Employee (EMPSTAT) variable has been deleted. Please refer to other employee-specific geographic variables. These are Employee Geographic Location (EGEOLOC) and Employee ZIP Code (EMPZIP).

The following variables were deleted in the 1995 database and subsequent years:

- The Disease Staging variables (EXPMORT, LOSCALE, LOSERR, PDXCAT, STAGE, RDSCALE, RDERR, and TRIMRD) are no longer delivered as standard variables unless the Disease Staging application has been licensed.
- AHAID was deleted from the Inpatient Admissions Table and the Inpatient Services Table.
- QCC, QDEATHS, and QTRACER were deleted from the Inpatient Admissions Table.
- STDPLAC was deleted from the Outpatient Pharmaceutical Claims Table because the Place was always set to “outpatient.”
1999

The COHORT Selection Table is one of three methods for selecting data contributors/plans with prescription drug claims. This file was developed because not all data contributors contribute prescription drug information to the MarketScan databases. The table does not ensure that a family opted for that coverage or had claims in a given year.

To simplify the use of the database and reduce the number of redundant variables, we have eliminated this table. Users may continue to use the Cohort Drug Indicator (RX)—now available on the claims and Populations Tables—or the RXCCYY variable on the Enrollment Tables to identify enrollees with drug coverage in a given data year.