The All-Payer Claims Database

Release 2.0
Documentation Guide

Product File

December 2013

Deval L. Patrick, Governor
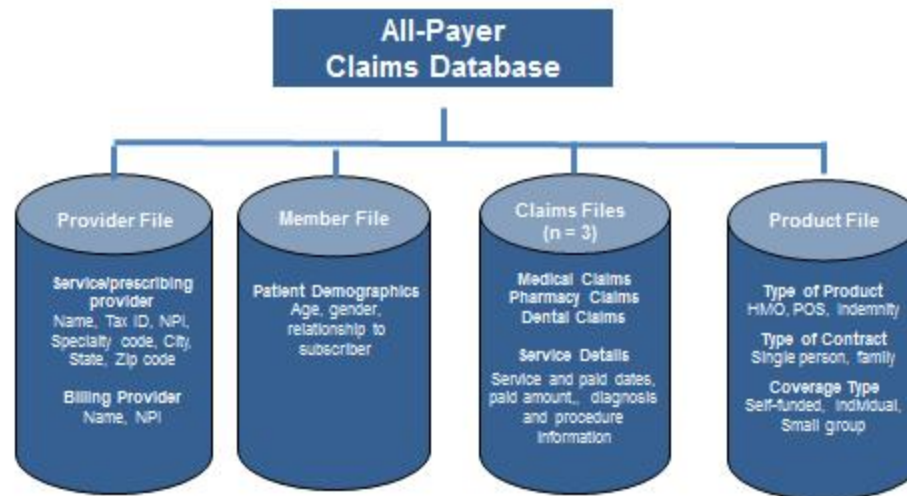**Commonwealth of Massachusetts**

Áron Boros, Executive Director
Center for Health Information and Analysis

Table of Contents

Contents

APCD Files and Selected Data Elements

```
                    All-Payer
                 Claims Database
```

Provider File	Member File	Claims Files (n = 3)	Product File
Service/prescribing provider Name, Tax ID, NPI, Specialty code, City, State, Zip code	Patient Demographics Age, gender, relationship to subscriber	Medical Claims Pharmacy Claims Dental Claims	Type of Product HMO, POS, Indemnity
Billing Provider Name, NPI		Service Details Service and paid dates, paid amount, diagnosis and procedure information	Type of Contract Single person, family
			Coverage Type Self-funded, individual, Small group

For ease of use, the Center for Health Information and Analysis (CHIA) has created separate documents for **each** APCD file type and one for the appendices—for a total of seven separate documents. All are available on the CHIA website.

INTRODUCTION

The Center for Health Information and Analysis (CHIA) was created to be the hub for high quality data and analysis for the systematic improvement of health care access and delivery in Massachusetts. Acting as the repository of health care data in Massachusetts, CHIA works to provide meaningful data and analysis for those seeking to improve health care quality, affordability, access, and outcomes.

To this end, the **All-Payer Claims Database (APCD)** contributes to a deeper understanding of the Massachusetts health care delivery system by providing access to accurate and detailed claims-level data essential to improving quality, reducing costs, and promoting transparency. This document is provided as a manual to accompany the release of data from the APCD.

The **APCD** is comprised of **medical**, **pharmacy**, and **dental claims**, and information from the **member eligibility**, **provider**, and **product** files, that is collected from health insurance payers operating in the Commonwealth of Massachusetts. This information encompasses public and private payers as well as insured and self-insured plans.

APCD data collection and data release are governed by **regulations** which are available on the APCD website (see http://www.mass.gov/chia/gov/laws-regs/chia-regulations.html).

APCD DATA COLLECTION

History

Establishment of the Massachusetts APCD

The first efforts to collect claim-level detail from payers in Massachusetts began in 2006 when the Massachusetts Health Care Quality and Cost Council (HCQCC) was established, pursuant to legislation in 2006, to monitor the Commonwealth's health care system and disseminate cost and quality information to consumers. Initially, data was collected by a third party under contact to the HCQCC. On July 1, 2009, the Division of Health Care Finance and Policy (DHCFP) assumed responsibility for receiving secure file transmissions, creating, maintaining and applying edit criteria, storing the edited data, and creating analytical public use files for the HCQCC. By July 2010, Regulations 114.5 CMR 21.00 and 114.5 CMR 22.00 became effective, establishing the APCD in Massachusetts.

Chapter 224 of the Acts of 2012, "An Act Improving the Quality of Health Care and Reducing Costs Through Increased Transparency, Efficiency and Innovation," created the Center for Health Information and Analysis (CHIA) which assumed many of the functions – including management of the APCD – that were previously performed by the Division of Health Care Finance and Policy (DHCFP).

According to Chapter 224, the purpose of the Massachusetts APCD is **Administrative Simplification:**

> "The center shall collect, store and maintain such data in a payer and provider claims database. The center shall acquire, retain and oversee all information technology, infrastructure, hardware, components, servers and employees necessary to carry out this section. All other agencies, authorities, councils, boards and commissions of the commonwealth seeking health care data that is collected under this section shall, whenever feasible, utilize the data before requesting data directly from health care providers and payers. In order to ensure patient data confidentiality, the center shall not contract or transfer the operation of the database or its functions to a third-party entity, nonprofit organization or governmental entity; provided, however, that the center may enter into interagency services agreements for transfer and use of the data. "

A Preliminary Release of the APCD – covering dates of service CY 2008-2010 and paid through February 28, 2011 – was released in 2012. Release 1.0 covered dates of service CY 2009-2011 and paid through February 2013. Release 2.0 covers dates of service CY2009-2012 and paid through June 2013.

APCD Data Collection Process

The data collected from the payers for the APCD is processed by the **Data Compliance and Support** team. Data Compliance works with the payers to collect the data on a regular, predetermined, basis and ensure that the data is as complete and accurate as possible.   The **Data Quality Assurance** and **Data Standardization and Enhancement** teams work to clean and standardize the data to the fullest extent possible. Data Standardization relies on **external source codes**[1] from outside government agencies, medical and dental associations, and other vendors to ensure that the data collectors properly utilized codes and lookup tables to make data uniform.

Edits

When payers submit their data to CHIA for the APCD, an **Edits process** is run on each file to check that the data complies with requirements for the file and for each data element in the file.

The automated edits perform an important data quality check on incoming submissions from payers. They identify whether or not the information is in the expected format (i.e. alpha vs. numeric), contains invalid characters (i.e. negative values, decimals, future dates) or is missing values (i.e. nulls). If these edits detect any issues with a file, they are identified on a report that is sent to the payer.

Data elements are grouped into four categories (A, B, C, and Z) which indicate their relative analytic value to the Center and APCD users. Refer to the **File Layout** sections of each document to view the Edit Level for each Data Element:

- '**A**' level fields must meet their **APCD threshold percentage** in order for a file to pass. There is an allowance for up to a 2% variance within the error margin percentage (depending on the data element). If any '**A**' level field falls below this percentage it will result in a failed file submission for the payer and a discussion with their liaison regarding corrective action.
- The other categories (**B, C, and Z**) are also **monitored**, but the thresholds are not presently enforced.

Variances

The **Variance process** is a collaborative effort between the payer and CHIA to reach a mutually agreed upon **threshold percentage** for any data element which may not meet the APCD standard. Payers are allowed to request a lower threshold for specific fields, but they must provide a business reason (rationale) and, in some cases, a remediation plan for those elements. CHIA staff carefully reviews each request and follows up with a discussion with the payers about how to improve data quality and possibly suggest alternative threshold rates or possibly "ramping up" overtime to the threshold. CHIA's goal is to work with payers to improve the quality of the APCD overtime.

Once this process is complete, the variance template is loaded into production so that any submissions from the payer are held to the CHIA standard thresholds and any approved variances. The payer receives a report after each submission is processed which compares their data against the required threshold percentages. 'Failed' files are reviewed by the Center liaisons and discussed with the payer for corrective action.[2]

---

[1] For more information on External Source Codes, refer to **Appendix 9** in the Appendices Release Document.
[2] For more information on variance see **Appendix 6**.

Broad Caveats

Researchers using the APCD Release 2.0 data should be aware of the following:

- Release files include data submitted to the Center through June 2013. Data submitted to the Center after June 2013 is **NOT** included in the files.
- Due to the variance process, data quality may vary from one payer to another. Consult Appendix 6 for more information.
- Claim Files submitted **through June 2010** were accepted with **relaxed edits**. (Refer to the edits section of this document.)
    - The release files contain the data submitted to the Center including valid and invalid values.
- Certain data elements were cleaned when necessary. Detail on the cleaning logic applied is described at the end of each file layout.
- Certain data elements were redacted to protect against disclosure of sensitive information.[3]
- Some Release Data was manipulated for compliance with HIPAA:
    - Assignment of linkage IDs to replace reported linkage identifiers (see **Appendix 4**).
    - Member Birth Year is reported as 999 for all records where the member age was reported as older than 89 years on the date of service.
    - Member Birth Year is reported as Null for all records where the member was reported as older than 115 years on the date of service.

APCD Release 2.0 Overview

The **APCD** is comprised of data elements collected from **all Private and Public Payers**[4] of eligible **Health Care Claims** for Massachusetts Residents.[5] Data is collected in six file types: **Product (PR)**, **Member Eligibility (ME)**, **Medical Claims (MC)**, **Dental Claims (DC)**, **Pharmacy Claims (PC)**, and **Provider (PV)**. Each is described separately in this user manual.

Highlights of the release include:

- Data is available for dates of service from January 1, 2009 to December 31, 2012 as paid through June 2013.
- Release 2.0 contains more comprehensive and recently updated data, including resubmissions from several large carriers.
- Data elements are classified as either Level 2 or Level 3 data elements. Level 2 include data elements that pose a risk of re-identification of an individual patient. Level 3 data elements are generally either Direct Personal information, such as name, social security number, and date of birth, that uniquely identifies an individual or are among the 18 identifiers specified by HIPAA. Refer to the **File Layout** sections for listings of Level 2 and Level 3 data elements for each file.[6]
- Public Use Files (PUFs), which are de-identified extracts of the Medical Claims (MC) and Pharmacy Claims (PC) files, will be release separately. The PUFs incorporate certain levels of aggregation and a much more limited list of elements to help ensure data privacy protection.
- Certain identifying or sensitive data elements are **Masked** in the release in order to protect personally identifiable information and allow for the linkage of data elements within the same file.
- Some data elements have been derived by CHIA from submission data elements or have been added to the database to aid in versioning and identifying claims (e.g. Unique Record IDs and status flags). Refer to the **File Layout** sections for detail.

---

[3] Detail on the redaction process is available in **Appendix 3.**

[4] Medicare data is only available to state agencies. Medicaid data requires separate approval from the Massachusetts Executive Office of Health and Human Services.

[5] In certain instances out of state residents are included. Most notably enrollees in the State's Group Insurance Commission medical programs and enrollees in plans subject to the Massachusetts risk adjustment program for the Affordable Care Act.

[6] Note that Level 1 (de-identified) extracts of the Medical Claims (MC) and Pharmacy Claims (PC) APCD files will be released by CHIA in the coming months.

PRODUCT FILE

As part of the APCD, payers are required to submit a **Product** file. Unlike the Preliminary Release which had one Product File per year of submission, Release 1.0 only has one Product File that consists of aggregated and unduplicated records across multiple years.

Below are details on business rules, data definitions, and the potential uses of this data. **For a full list of elements refer to the File Layout section**.

Types of Data Collected in the Product File:

Product Identifiers

The Center has made a conscious decision to collect elementary identifiers that may be associated with a Product. The data in fields PR002 through PR008 can be used when analyzing Product data across payers. The identifiers will be used to help link Product data to the Member's Eligibility File.

Deductibles

The Center collects deductible information using bands. Additional data elements such as Coinsurances and Co-pays are reported in other file types.

Dates

The Center collects two date fields for each Product record.

The Begin and End Dates for each Product describes the dates the Product was active with the payer and usable by eligible members. For Products that were still active at the end of 2011, the End Date should be Null. For Products that were not active, but may still have claims being adjudicated against them, the End Date should be the End Date reported to the Division of Insurance OR the date the license was terminated.

Product Release File Structure:

File Characteristic	Details
Frequency of Submission by Payers to CHIA	Quarterly
Release File Format	Release files will be in an **asterisk delimited text file** in the same order as found in the File Layout sections. Empty or null data elements will have no spaces or characters between the asterisks. Each user will only receive the data elements requested and approved.
Rows	Each row is supposed to represent a unique instance of a Product. However, some payers have reported products on separate rows that differ only in aspects that are not specified in the Product file layout. Therefore, for some payers there may be appear to be duplicate rows, when in fact they are district product.
Product Definition	A **Product** starts as a base offering, often described by a business model that it conforms to, e.g.: HMO, PPO, Indemnity, etc. Product Line of Business Model (PR004) is collected by the APCD to define the type of business model. The data must be submitted using a CHIA-provided lookup table, which is located in the Product File Lookup Table section below.

File Characteristic	Details
Products or lines of business not included in the lookup table for PR004	For other lines of business the Payers will report the following:

Element	Element Name	Submission Guideline
PR004	Product Line of Business	ZZ
PR007	Other Product Benefit Description	Payers may enter the name of the business model here.

By reporting the Model Code of ZZ (mutually defined by CHIA and the Payers) the Payer will be able to report the name of the business model in PR007. The Center realizes that payers store their Product data in a variety of formats and data structures. The Center feels this methodology will provide the most flexibility to analyze Product data.

Product File Layout

The following sections list **all APCD data elements** for the Product File and grouped by file type and then by level of availability for release.

Restricted Release Elements:

- Each **row** in the release file contains one record of the indicated file type. There is an **asterisk-delimited field** in each row for every data element listed in the Restricted Release sections for each file type.

- Data Elements will be delimited in the order displayed in the File Layout sections of this document.

- **Empty** or **null** data elements will have no spaces or characters between the asterisks.

Lookup Tables:

- **Element-specific** Lookup Tables are included in this document after each File Type Layout section.

- A **Carrier-Specific Master Lookup** table is included with each data extract. Refer to the **Carrier-Specific Reference** and **Linking** sections in this document for more information.

- **External Code Sources** are listed in Appendix 9.

File Layout Section Columns

- **Element**: The code name of the element, with reference to the Regulation and the Submission files received by the Center from Payers. The first two digits refer to the File Type and the following numbers to the ordering in the Submission Files.
- **Data Element Name**: Name of the element.
- **Max Length:** Maximum Length of the data column in the APCD's SQL Server database at the Center.
- **Data Type Guide:** Data Type of the column in the APCD's SQL Server database at the Center. When the APCD Release text file is imported to a database or other file type by the final user of the data, these data types provide a guide to setting up the columns in the receiving file.
- **Description:** Description of the element.
- **Release Notes:** Additional information about the element in the release.
- **Edit Level:** Level of enforcement of the data element's requirements by the Center on Payer Submissions. Refer to the **Edits** section of this document.
- **APCD Threshold:** The expected percentage of validity for instances of the element in each submission file by the Payer.

Release Text File Column Titles

- **Appendix 10: Release File Column Names** included in this document lists the column name for each data element in the Level 2 and Level 3 release files. The text files exported from the APCD SQL Database include these SQL column names in the first row.

The APCD Product File

						Product File – Level 2 Data Elements		

Element	Data Element Name	Max Length	Data Type Guide	Description	Release Notes	Edit Level[7]	APCD Threshold[8]
HD002	Payer	6	varchar	Carrier/Submitter Org ID (from the file Header)	Unique identifier assigned by CHIA to each Carrier/Submitter		
Derived - PR1	Release ID	NULL	int	Unique record ID derived specifically for this release file type	With each release file type table this number is reset to 1 and sequentially incremented by one for every record released	N/A	N/A
PR001	Linking Product ID Number		int	Product Identification Number	Carrier-assigned identifier that uniquely defines this Product. This identifier is used in tandem with Payer Org ID to align Products to Eligibility and the corresponding claims when applicable.		
PR003	Carrier License Type	10	varchar	Carrier License Type (Lookup Table – Redacted due to identifying data)	A code that defines the license type associated with the Product filing with the Massachusetts Division of Insurance.	A0	100
PR004	Product Line of Business Model	2	varchar	The Line of Business / Insurance Model the Product relates to. (Lookup Table)	A code that defines a product's business model as defined by the carrier or its designee. Value of ZZ (Other) should correspond to non-insurance vendors; Claim Re-processors or Re-pricers, Computer Leasing, etc.	A0	100
PR005	Insurance Plan Market	10	varchar	Insurance Plan Market Code (Lookup Table)	A code that defines a product's business model as defined by the carrier or its designee. Value of ZZ (Other) should correspond to non-insurance vendors; Claim Re-processors or Re-pricers; Computer Leasing; etc.	A0	100
PR006	Product Benefit Type	1	varchar	Indicates combinations of offerings (Lookup Table)	Numeric indicator that reports a benefit selection or a product-range offering as defined by the carrier or its designee.	A0	100

---

[7] See pg. 5 for a discussion on Edit Levels.
[8] See pg. 5 for a discussion on APCD Thresholds.

				Product File – Level 2 Data Elements			
Element	Data Element Name	Max Length	Data Type Guide	Description	Release Notes	Edit Level[7]	APCD Threshold[8]
PR008	Risk Type	1	varchar	Indicates if the product was an at-risk product or self-insured. (Lookup Table)	Numeric indicator that reports the product development attribute that defines a risk assignment.	A2	100
PR009	Product Start Date	8	datetime	Product Start Date	First date that a product is eligible for Member enrollment. (YYYY-MM-DD 00:00:00.000)	A0	100
PR010	Product End Date	8	datetime	Last date on which members could be enrolled in this product	Last date that product is active for Member enrollment. (YYYY-MM-DD 00:00:00.000)	B	100
PR011	Product Active Flag	1	varchar	Indicator to further refine activity status (Lookup Table)	Numeric indicator that reports active vs. inactive products for the date span indicated in Product Start and End Dates.	C	100
PR012	Annual Per Person Deductible Code	3	varchar	Per Person Deductible bandwidth reporting (Lookup Table)	Value that represents the Total Per Person Deductible for all benefits under this product for the date span indicated in Product Start and End Dates.	B	100
PR013	Annual Per Family Deductible Code	3	varchar	Per Family Deductible bandwidth reporting (Lookup Table)	Value that represents the Total Per Family Deductible for all benefits under this product for the date span indicated in Product Start and End Dates	B	100
PR014	Coordinated Care model	1	varchar	Indicates if a patient's care is clinically coordinated or managed. (Lookup Table)	Numeric indicator that reports if Patient care is clinically coordinated or managed, as an attribute of this product, by the carrier or its designee	C	100

				Product File – Level 3 Data Elements			
Element	Data Element Name	Max Length	Data Type Guide	Description	Release Notes	Edit Level[9]	APCD Threshold[10]
PR002	Product Name	70	varchar	Carrier defined Product Name	Unique name assigned to Product by Carrier/Submitter	C	100

---

[9] See pg. 5 for a discussion on Edit Levels.
[10] See pg. 5 for a discussion on APCD Thresholds.

Product File – Level 3 Data Elements							
Element	Data Element Name	Max Length	Data Type Guide	Description	Release Notes	Edit Level[9]	APCD Threshold[10]
PR007	Other Product Benefit Description	80	varchar	Benefit Description	Refining description applied by carrier or its designee when PR006 was reported as 0 (Other).	B	100

APCD Product File: External Code Sources

Refer to Appendix 9: External Code Sources

Product File Cleaning, Standardization, and Redaction

APCD Product File Cleaning Logic, by Element

Element	Data Element Name	Format/Length	Description	Cleaning Logic
PR005	Insurance Plan Market	varchar[10]	Insurance Plan Market Code	Change: 'GMMMK' TO 'GMMK'.
PR007	Other Product Benefit Description	varchar[80]	Benefit Description	Nullify the value of Benefit Description (PR007) when Product Benefit Type (PR006) is not equal to Other ('0').
PR011	Product Active Flag	int[1]	Indicator - Active Product	Change: 'A' to '1', 'Y' to '1'.
PR012	Annual Per Person Deductible Code	char[3]	Per Person Deductible bandwidth reporting	Change: '0' to '000', '1' to '001', '2' to '002', '3' to '003', '4' to '004'.
PR013	Annual Per Family Deductible Code	char[3]	Per Family Deductible bandwidth reporting	Change: '0' to '000', '1' to '001', '2' to '002', '3' to '003', '4' to '004' '5' to '005', '6' to '006', '7' to '007'.

APCD Product File SSN Redaction, by Element[11]

Element	Data Element Name	Format/Length	Description
PR002	Product Name	varchar[70]	Carrier defined Product Name
PR007	Other Product Benefit Description	varchar[80]	Benefit Description

APCD Product File Reidentification, by Element

Element	Data Element Name	Format/Length	Description

---

[11] Please refer to **Appendix 3** for details on the redaction process and **Appendix 4** the reidentification process.

| PR001 | Product ID Number | varchar[30] | Product Identification |