

Derivation of the Diagnoses Public Use File

Version 1.0 Produced March 21, 2016

The data included in the Diagnosis-based Public Use File (PUF) are derived from detailed health insurance claims submissions by payers, Third Party Administrators (TPAs) and Prescription Benefit Management (PBM) firms to the Minnesota All Payer Claims Database (MN APCD).

To prepare final data extracts for the MN APCD and develop PUFs, MDH and its data aggregator perform a series of steps processes designed to ensure the robustness of the data. These steps involve four primary processes: data consolidation, deduplication, aggregation and redaction, each of which are described below.

Consolidation

Consolidation is the process of combining multiple versions of a claim that may be submitted by a single insurance company or other payers. When a company responsible for paying a claim reviews it, it determines whether the claim can be paid and if so, the amount owed by the plan and the patient's share. This claim record is then submitted to the MN APCD. In some instances, this process may be repeated and a new or revised claim could be submitted, resulting in multiple versions of the claim. MDH's data aggregation manager uses sophisticated, extensive algorithms to consolidate these multiple transactions into a single, consolidated claim record.

Deduplication

In addition, multiple insurance companies and other payers sometimes share responsibility for paying a single claim. Once all transactions are consolidated into a single claim record for each company involved in paying the claim, additional processing is undertaken to resolve all duplicate records into a single, de-duplicated claim record that provides the most complete and accurate description of the service event possible. For the MN APCD, this takes the form of a "flag" to give analysts the most control over which costs to count. For summary data, aggregation effectively rolls up duplicate claims.

Aggregation

Aggregation is the process of grouping and summarizing multiple de-duplicated claim records into the form that is reported in the PUF, where each line of data represent all records (volume of claims and associated costs) for the combination of data elements in the PUF like patient age, patient ZIP code and diagnostic information.

Redaction

Redaction is the process used to ensure that each PUF on its own and in combination with each other or third-party data stays within legislatively established guard rails concerning protecting provider, health insurance carrier and individual patient identities. Redaction criteria were selected with the objective of retaining as much data by as possible while providing necessary protections.

Redaction takes the following iterative steps:

1. For cells with identification risk, geographic detail is redacted first. Health plan members and paid amounts associated with one or more redacted geographic cells for a given diagnosis are aggregated in a redacted cell, i.e., information is not “lost”.
2. When recalculated cells continue to pose potential identification risk, member and paid amounts for one or multiple age groups are redacted next. Again, information is appropriately aggregated in cells combining redacted geography and redacted age for a given diagnosis.
3. In some instances, the remaining diagnosis information still poses potential identification risk and is therefore redacted as well. At this points, PUFs do not retain information for cells where geography, age, and the diagnosis code is redacted.

The following table summarizes the data at each step of the processing described above.

Total medical claims transactions submitted with service date 2013	211,795,459	
Total consolidated medical claims records with service date 2013	184,200,895	
Total Deduplicated Claims	176,681,673	
Claim records not included in raw PUF due to disqualifying factors (e.g., non-Minnesota residents, missing diagnostic code, inaccurate diagnostic information, negative paid amounts)	1,467,659	
	Number of Claim Records	Total Dollars
In Pre-redaction PUF	175,214,014	\$ 25,861,096,170
Completely removed from Raw PUF to meet patient cell size limits	31,950	\$ 12,072,591
Completely removed from Raw PUF to meet provider and health plan cell size limits	422,357	\$ 111,565,615
Remaining in Published PUF	174,759,707	\$25,737,457,964
Percent remaining in Published PUF	99.7%	99.5%

Users should be aware that unique health plan member information is just unique to each individual cell in the PUF. Because health plan members over a given year of health care use can have multiple primary diagnoses, they could appear at minimum in multiple diagnoses cells for the same age-by-geography combination. Thus, unique member counts are not additive.