



## **Suggested Data Warehouse Fields for Robust Analysis**

**Introduction:**

HCI3 is working with health plans throughout the country on pilot implementations to test the validity of the Evidence-informed Case Rate® (ECR) models, as a means to fairly and equitably price health care services and encourage physicians, hospitals and all other clinicians to collaborate in delivering effective and efficient care. In working with these health plans on creating data extracts and analyzing data, the necessity of a robust and accurate data warehouse is paramount.

Prior to implementing PROMETHEUS, an initial data analysis of a health plan's claims data must be performed as it is essential for determining the areas in which quality of care can be improved and costs can be reduced. If the health plan's data warehouse is inaccurate, incomplete, or lacking member level details/identifiers, the results of this initial data run or any data analysis would be compromised, and depending on the extent of the inaccuracies, considered meaningless.

The suggested data warehouse specifications outlined in this document were derived from the experience HCI3 has gained working with health plans to analyze their data. The specifications below are not designed to be a gold standard, but rather a helpful tool for those who are trying to assemble all payer data sets in an effort to create a community dashboard for costs of care.

**Context:**

This document is intended for use by anyone who wishes to create a data warehouse to be used for robust data analysis. Data should be stored in a way that allows it to be easily extracted and formatted for analysis. A detailed data dictionary with data field specifications and value definitions should be created to assist with data submissions. The data specifications below are considered the minimum that should be captured in order to perform analyses, and the data warehouse administrator should apply their best professional judgment to create additional data fields in their data warehouse where additional data are available.

**Data Field Check List by Field Type:****Member Fields**

- Consistent member ID over time and unique to each plan member (not employee)
- Gender Code (e.g. F=Female, M=Male)
- Year of Birth (e.g. YYYY, e.g. 1956)
- Member Zip Code

**Non-Essential But Important Additional Fields**

- Race
- Ethnicity

**Enrollment Fields**

- Consistent member ID
- Begin Date- member enrollment effective date
- End Date - member enrollment termination date; if member is still enrolled, leave this field blank

If a member has multiple begin and end dates these should be captured in the data. For example, you could create multiple rows for a member with multiple begin and end dates so each is captured in order to understand continuity of eligibility and breaks over time.

**Inpatient Stay Claim Fields**

- Consistent member ID
- Allowed Amount
- From Date - first date of service
- Through Date - last date of service
- Discharge Status code
- Admit Type code (mark admission through emergency room =1, else 0)
- Place of Service Code (mark emergency room service = 23)
- Type of Claim (e.g. mark it = 1 for all inpatient stay claims)
- Bill Type
- Revenue Code (Nx)
- HCPCS Proc Code
- HCPCS Proc Modifier Code
- CPT Code (N1)

- CPT Modifier Code (N1)
- Principal ICD-9 Diag Code
- Secondary ICD-9 Diag Code (N2)
- Principal ICD-9 Proc Code
- Secondary ICD-9 Proc Code (N3)
- TIN
- UPIN
- NPI
- Other Provider ID
- Specialty

Where:

- N1, N2, N3, Nx values could be any number from 1 through 999 and could be different values for CPT codes, revenue codes, ICD-9 procedure codes or ICD-9 diagnosis codes
- N1 represents the number of CPT Codes that will be provided for each claim.
- N2 represents the number of ICD-9 Diagnosis Codes that will be provided for each claim.
- N3 represents the number of ICD-9 Procedure Codes that will be provided for each claim.
- N1, N2, and N3 will be determined during extract based on available data.
- Nx (Revenue codes are mostly important to help identify emergency room visits in stay records: 45x and certain exclusionary criteria such as use of dialysis: 80x, oncology services 28x and labor & delivery: 72x)

It is very important to designate the value of N1, N2, N3, and Nx in the Inpatient Stay Claims file.

**Professional Claim Fields**

- Consistent member ID
- Allowed Amount
- From Date
- Thru Date
- Place of Service Code (mark emergency room service = 23)
- Type of Claim (ex. 2 = Outpatient Facility Claims; 3 = Professional, Ancillary and Other Claims)
- Bill Type
- HCPCS Proc Code
- HCPCS Proc Modifier Code
- CPT Code (N1)
- CPT Modifier Code (N1)
- Principal ICD-9 Diag Code
- Secondary ICD-9 Diag Code (N2)
- Principal ICD-9 Proc Code
- Secondary ICD-9 Proc Code (N3)
- TIN
- UPIN
- NPI
- Other Provider ID
- Specialty

Where:

- N1, N2, N3, Nx values could be any number from 1 through 999 and could be different values for CPT codes, revenue codes, ICD-9 procedure codes or ICD-9 diagnosis codes
- N1 represents the number of CPT Codes that will be provided for each claim.
- N2 represents the number of ICD-9 Diagnosis Codes that will be provided for each claim.
- N3 represents the number of ICD-9 Procedure Codes that will be provided for each claim.
- N1, N2, and N3 will be determined during extract based on available data.

It is very important to designate the value of N1, N2, N3, and Nx in the Professional Claims file.

**Pharmacy Claim Fields**

- Consistent member ID
- Allowed Amount
- NDC code
- Supply Days Num
- Prescription filled Date

**Data Warehouse Field Specifications Checklist:**

Below are some data checks to help ensure the integrity of the data.

- Are NPI numbers 10 digits long?
- Are CPT codes 5 digits long?
- Do date fields contain all four digits for the year?
- Do NDC codes contain leading and trailing zeros?
- Is the member ID unique and accurate at the member level?
- Does the medical claim data eligibility period match the pharmacy eligibility period?
- Are claim header and claim costs **NOT** rolled up? (Rolling these up will double costs)
- How much detail is available at the claim line level?

**EXAMPLE OF DATA WAREHOUSE FIELDS & FORMATS**

Below is an example of a layout for a data warehouse that would enable reliable analyses at the provider level.

<b>Field Mapping and Format for <u>Member File</u></b>			
<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Description</b>
CONSISTENT_MEMBER_ID	Char	(varies by data)	Unique member ID.
SEX	Char	1	Patient gender F = Female M = Male
YOB	Num	8	Member's year of birth. Coded as YYYY (e.g., 1965).
Member Zip Code ( <i>optional</i> )	Char	5	Member's zip code (not needed for running Prometheus standard analysis. Will be needed if user wishes to perform additional regional variation analysis).

<b>Field Mapping and Format for <u>Member Enrollment File</u></b>			
<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Description</b>
CONSISTENT_MEMBER_ID	Char	(varies by data)	Unique member ID.
BEGIN_DATE_s	Num	8	Member enrollment effective date (SAS date).
END_DATE_s	Num	8	Member enrollment termination date (SAS date). If a member is still enrolled, leave this field blank.

<b>Field Mapping and Format for <u>Inpatient Stay Claims File</u></b>			
<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Label and/or Description</b>
CONSISTENT_MEMBER_ID	Char	(varies by data)	Unique member ID.
PROVIDER ID	Char	(varies by data)	The ID that identifies the serving hospital, such as the claim organization NPI number.
ALLOWED_AMT	Num	8	Total Reimbursed Allowed Amount. Please do not include dollar sign. See Section 2.3 for the calculation of Total Reimbursed Allowed Amount.
FROM_DATE_s	Num	8	First Date of Service (SAS date).
THRU_DATE_s	Num	8	Last Date of Service (SAS date).

Field Mapping and Format for <b>Inpatient Stay Claims File</b>			
Variable	Type	Length	Label and/or Description
DISCHARGE_STATUS	Char	2	Code indicating status at discharge from the hospital, as defined in the Medicare Carrier manual for the claim. This field is used to identify left against medical advice (LAMA) or death. Codes that will be used by the Prometheus program are listed below. '07' = Left against medical advice or discontinued care '20' = Expired (or did not recover - Christian Science patient). '40' = Expired at home (hospice claims only). '41' = Expired in a medical facility (e.g. hospital, SNF, ICF or free standing hospice). (Hospice claims only) '42' = Expired - place unknown (hospice claims only). Other discharge status is allowed in this field, as long as above status is coded as indicated.
ADMIT_TYPE_CODE	Num	1	Type of Admission/Source of Admission. This field is used to identify admission through emergency room. 1 = Emergency room service 0 = Not emergency room service  If no such a field is directly available in the original claims data, the user should use other source of information to identify ER services and populate this field as indicated.
PLACE_OF_SVC_CODE	Char	2	Place of service. This field is used to identify emergency room service. '23' = Emergency Room Other place of service codes are allowed in this field, as long as emergency room is coded as indicated.
TYPE_OF_CLAIM	Char	1	Type of claim: this field should be generated by the user. '1' = Inpatient Facility Claims '2' = Outpatient Facility Claims '3' = Professional, Ancillary and Other Claims The inpatient stay claims file should only contain claims that carry a '1' for this field.
REV_CPTX	Char	5	Revenue Code <b>X (Nx)</b>
HCPCS_PROC_CODE	Char	5	HCPCS CPT4 Procedure Code.
HCPCS_CPT_MOD	Char	2	Initial (functional) CPT modifier for HCPCS_PROC_CODE

<b>Field Mapping and Format for Inpatient Stay Claims File</b>			
<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Label and/or Description</b>
CPTCODEX	Char	5	Additional HCPCS CPT4 Procedure Code <i>X (N1)</i>
CPTCODE_MODX	Char	2	Initial (functional) CPT modifier for CPTCODEX <i>(N1)</i>
PRINCIPAL_DIAG_CODE	Char	6	ICD-9 Principal Diagnosis code
SECONDARY_DIAGX_CODE	Char	6	ICD-9 Secondary Diagnosis code <i>X (N2)</i>
PRINCIPAL_PROC_CODE	Char	6	ICD-9 Principal Procedure code
PROC_ICD9_2NDX	Char	6	ICD-9 Secondary Procedure code <i>X (N3)</i>

Where *X* could be any of the values below:

- N1 represents a fixed number of multiple CPT Codes on claims.
- N2 represents a fixed number of multiple ICD-9 Diagnosis Codes on claims.
- N3 represents a fixed number of multiple ICD-9 Procedure Codes on claims.
- N1, N2, and N3 should be determined by the user during extract based on available data.
- Nx (Revenue codes are mostly important to help identify emergency room visits in stay records: 45x and certain exclusionary criteria such as use of dialysis: 80x, oncology services 28x and labor & delivery: 72x)

<b>Field Mapping and Format for Professional, Outpatient Facility, Ancillary and Other Claims file</b>			
<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Label and/or Description</b>
CONSISTENT_MEMBER_ID	Char	(varies by data)	Unique member ID.
ALLOWED_AMT	Num	8	Total Reimbursed Allowed Amount. Do not include dollar sign. See Section 2.3 for the calculation method.
PROVIDER ID	Char	(varies by data)	This ID identifies the performing physician, such as the claim performing physician NPI number on a professional claim and the claim organization ID number on an outpatient facility claim.

<b>Field Mapping and Format for Professional, Outpatient Facility, Ancillary and Other Claims file</b>			
<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Label and/or Description</b>
SPECIALTY	Char	2	CMS specialty code used for pricing the line item service on the noninstitutional claim. Specialty Codes that will be used by the Prometheus program are listed below. 01 = General practice 08 = Family practice 11 = Internal medicine 16 = Obstetrics/gynecology 37 = Pediatric medicine 38 = Geriatric medicine 84 = Preventive medicine 06 = Cardiology 46 = Endocrinology 29 = Pulmonology 10 = Gastroenterology 13 = Neurology
FROM_DATE_s	Num	8	First Date of Service (SAS date).
THRU_DATE_s	Num	8	Last Date of Service (SAS date) – usually FROM_DATE and THRU_DATE is the same in PFO claims.
PLACE_OF_SVC_CODE	Char	2	Place of service. This field is used to identify emergency room service. ‘23’ = Emergency Room Other place of service codes are allowed in this field, as long as emergency room is coded as indicated.
TYPE_OF_CLAIM	Char	1	Type of claim: this field should be generated by the user. ‘1’= Inpatient Facility Claims ‘2’= Outpatient Facility Claims ‘3’= Professional, Ancillary and Other Claims The professional, outpatient facility, ancillary and other claims file should only contain claims that have a ‘2’ and ‘3’ for this field.
HCPCS_PROC_CODE	Char	5	HCPCS CPT4 Procedure Code.
HCPCS_CPT_MOD	Char	2	Initial (functional) CPT modifier for HCPCS_PROC_CODE
CPTCODEX	Char	5	Additional HCPCS CPT4 Procedure Code <b>X (N1)</b>
CPTCODE_MODX	Char	2	Initial (functional) CPT modifier for CPTCODEX <b>(N1)</b>
PRINCIPAL_DIAG_CODE	Char	6	ICD-9 Principal Diagnosis code
SECONDARY_DIAGX_CODE	Char	6	ICD-9 Secondary Diagnosis code <b>X (N2)</b>
PRINCIPAL_PROC_CODE	Char	6	ICD-9 Principal Procedure code

<b>Field Mapping and Format for Professional, Outpatient Facility, Ancillary and Other Claims file</b>			
<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Label and/or Description</b>
PROC_ICD9_2NDX	Char	6	ICD-9 Secondary Procedure code <i>X (N3)</i>

Where *X* could be any of the values below:

- N1 represents a fixed number of multiple CPT Codes on claims.
- N2 represents a fixed number of multiple ICD-9 Diagnosis Codes on claims.
- N3 represents a fixed number of multiple ICD-9 Procedure Codes on claims.
- N1, N2, and N3 should be determined by the user during extract based on available data.

<b>Field Mapping and Format for Pharmacy file</b>			
<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Label and/or Description</b>
CONSISTENT_MEMBER_ID	Char	(varies by data)	Unique member ID.
ALLOWED_AMT	Num	8	Total Reimbursed Allowed Amount. Please do not include dollar sign. See Section 2.3 for the calculation of Total Reimbursed Allowed Amount.
NDC_CODE	Char	11	NDC Drug Code. NDC codes should be 11 characters. Hyphens should be removed. Leading and trailing zeros should be included. For example, 00000-1435-00 should be formatted as 00000143500.
SUPPLY_DAYS_NUM	Num	8	Number of days of prescription supply
PRESCRIPTION_FILLED_DATE_S	Num	8	Date of prescription filled (SAS date).