Data Specification for the ITM Leaf Query Tool

The Institute of Translations Medicine (ITM) relies on components of the PCORnet Common Data Model (CDM) to define the database utilized for its Leaf Query Tool. The ITM Leaf dataset is a subset of the PCORnet CDM (http://www.pcornet.org/data/) comprised of select data domains frequently used for patient cohort discovery. The select domains are detailed in this document.

General notes:

- Included data domains: Demographic, Encounter, Diagnosis, Procedures, Death
- This document's domain specification tables were sampled from the official PCORnet documentation.
- Data element shown in tables are color coded as follows, based on availability.
    - Elements available through the Leaf website via query parameters or visualizations are coded Green.
    - Elements included in the underlying Leaf database, but not available through the website are coded Yellow.
    - Elements from the PCORnet model which are not available in the Leaf database are coded Grey.
- Missingness of individual data elements may vary between sites based on availability at their source. Additional Loyola specific information is included in the Definition/Comments section of each data element where applicable. Researchers requesting data should refer to local data curation teams for specifics about each site's missing data.

Data Specification for the ITM Leaf Query Tool

TABLE OF CONTENTS

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Table: DEMOGRAPHIC

Demographics record has direct attributes of individual patients.

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
PATID	.	Arbitrary person-level identifier used to link across tables. PATID is a pseudoidentifier with a consistent crosswalk to the true identifier retained by the source data partner. For analytical data sets requiring patient- level data, only the pseudoidentifier is used to link across all information belonging to a patient. The PATID must be unique within each PCORnet data mart. Creating a unique identifier within a network would be beneficial and acceptable. The PATID is not the basis for linkages across data partners.
BIRTH_DATE	.	Date of birth.
BIRTH_TIME	.	Time of birth. NOTE: values are all 00:00/12:00AM
SEX	A=Ambiguous F=Female M=Male NI=No information UN=Unknown OT=Other	Sex assigned at birth.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
SEXUAL_ORIENTATION	AS=Asexual BI=Bisexual GA=Gay LE=Lesbian QU=Queer QS=Questioning ST=Straight SE=Something else MU=Multiple orientations DC=Decline to answer NI=No information UN=Unknown OT=Other	Sexual orientation.
GENDER_IDENTITY	M=Man F=Woman TM=Transgender male/Trans man/Female-to-male TF=Transgender female/Trans woman/Male-to-female GQ=Genderqueer/Non-binary SE=Something else MU=Multiple categories DC=Decline to answer NI=No information UN=Unknown OT=Other	Current gender identity.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
HISPANIC	Y=Yes N=No R=Refuse to answer NI=No information UN=Unknown OT=Other	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.
RACE	01=American Indian or Alaska Native 02=Asian 03=Black or African American 04=Native Hawaiian or Other Pacific Islander 05=White 06=Multiple race 07=Refuse to Answer NI=No information UN=Unknown OT=Other	Please use only one race value per patient. **Details of categorical definitions:** American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. Black or African American: A person having origins in any of the black racial groups of Africa. Native Hawaiian or Other Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
BIOBANK_FLAG	Y=Yes N=No	Flag to indicate that one or more biobanked specimens are stored and available for research use. Examples of biospecimens could include blood, urine, or tissue (e.g., skin cells, organ tissues). If biospecimens are available, locally maintained "mapping tables" would be necessary to map between the DEMOGRAPHIC record and the originating biobanking system(s). If no known biobanked specimens are available, this field should be marked "No".
PAT_PREF_LANGUAGE_SPOKEN	.	Preferred spoken language of communication as expressed by the patient.
RAW_SEX	.	Field for originating value of field, prior to mapping into the PCORnet CDM value set.
RAW_SEXUAL_ORIENTATION	.	Field for originating value of field, prior to mapping into the PCORnet CDM value set.
RAW_GENDER_IDENTITY	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_HISPANIC	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_RACE	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_PAT_PREF_LANGUAGE_SPOKEN	.	Field for originating value, prior to mapping into the PCORnet CDM value set.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Table: ENCOUNTER

Encounters are interactions between patients and providers within the context of healthcare delivery.

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
ENCOUNTERID	.	Arbitrary encounter-level identifier. Used to link across tables, including the ENCOUNTER, DIAGNOSIS, and PROCEDURES tables.
PATID	.	Arbitrary person-level identifier used to link across tables.
ADMIT_DATE	.	Encounter or admission date.
ADMIT_TIME	.	Encounter or admission time. NOTE: values are all 00:00/12:00AM
DISCHARGE_DATE	.	Discharge date.
DISCHARGE_TIME	.	Discharge time. NOTE: values are all 23:59/11:59PM
PROVIDERID	.	Code for the provider who is most responsible for this encounter. As with the PATID, the provider code is a pseudoidentifier with a consistent crosswalk to the real identifier.
FACILITY_LOCATION	.	Geographic location (5-digit zip code).

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
ENC_TYPE	AV=Ambulatory Visit ED=Emergency Department EI=Emergency Department Admit to Inpatient Hospital Stay (permissible substitution) IP=Inpatient Hospital Stay IS=Non-Acute Institutional Stay OS=Observation Stay IC=Institutional Professional Consult (permissible substitution) TH=Telehealth OA=Other Ambulatory Visit NI=No information UN=Unknown OT=Other	Encounter type. Details of categorical definitions: Ambulatory Visit: Includes visits at outpatient clinics, physician offices, same day/ambulatory surgery centers, urgent care facilities, and other same-day ambulatory hospital encounters, but excludes emergency department encounters. Emergency Department (ED): Includes ED encounters that become inpatient stays (in which case inpatient stays would be a separate encounter). Excludes urgent care facility visits. ED claims should be pulled before hospitalization claims to ensure that ED with subsequent admission won't be rolled up in the hospital event. Does not include observation stays, where known. Emergency Department Admit to Inpatient Hospital Stay: Permissible substitution for preferred state of separate ED and IP records. Only for use with data sources where the individual records for ED and IP cannot be distinguished. Inpatient Hospital Stay: Includes all inpatient stays, including: same-day hospital discharges, hospital transfers, and acute hospital care where the discharge is after the admission date. Does not include observation stays, where known. Observation Stay: " Hospital outpatient services given to help the doctor decide if the patient needs to be admitted as an inpatient or can be discharged. Observations services may be given in the emergency department or another area of the hospital." Definition from Medicare, CMS Product No. 11435, https://www.medicare.gov/Pubs/pdf/11435.pdf. Institutional Professional Consult: Permissible substitution when services provided by a medical professional cannot be combined with the given encounter record, such as a specialist consult in an inpatient setting; this situation can be common with claims data sources. This includes physician consults for patients during inpatient encounters that are not directly related to the cause of the admission (e.g. a ophthalmologist consult for a patient with diabetic ketoacidosis) (guidance updated in v4.0). Non-Acute Institutional Stay: Includes hospice, skilled nursing facility (SNF), rehab center, nursing home, residential, overnight non-hospital dialysis, and other non-hospital stays. Telehealth: Includes telemedicine or virtual visits, which can be conducted via video, phone or other means. Other Ambulatory Visit: Includes other non-overnight AV encounters such as hospice visits, home health visits, skilled nursing visits, other non-hospital visits, as well as ~~telemedicine,~~ telephone and email consultations. May also include "lab only" visits (when a lab is ordered outside of a patient visit), "pharmacy only" (e.g., when a patient has a refill ordered without a face-to- face visit), "imaging only", etc.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
FACILITYID	.	Arbitrary local facility code that identifies the hospital or clinic. Used for chart abstraction and validation. FACILITYID can be a true identifier, or a pseudoidentifier with a consistent crosswalk to the true identifier retained by the source data partner.
DISCHARGE_DISPOSITION	A=Discharged alive E=Expired NI=No information UN=Unknown OT=Other	Vital status at discharge.
DISCHARGE_STATUS	AF=Adult Foster Home AL=Assisted Living Facility AM=Against Medical Advice AW=Absent without leave EX=Expired HH=Home Health HO=Home / Self Care HS=Hospice IP=Other Acute Inpatient Hospital NH=Nursing Home (Includes ICF) RH=Rehabilitation Facility RS=Residential Facility SH=Still In Hospital SN=Skilled Nursing Facility NI=No information UN=Unknown OT=Other	Discharge status.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
DRG	.	3-digit Diagnosis Related Group (DRG). The DRG is used for reimbursement for inpatient encounters. It is a Medicare requirement that combines diagnoses into clinical concepts for billing. Frequently used in observational data analyses.
DRG_TYPE	01=CMS-DRG (old system) 02=MS-DRG (current system) NI=No information UN=Unknown OT=Other	DRG code version.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
ADMITTING_SOURCE	AF=Adult Foster Home AL=Assisted Living Facility AV=Ambulatory Visit ED=Emergency Department ES=Emergency Medical Service HH=Home Health HO=Home / Self Care HS=Hospice IP=Other Acute Inpatient Hospital NH=Nursing Home (Includes ICF) RH=Rehabilitation Facility RS=Residential Facility SN=Skilled Nursing Facility IH=Intra-hospital NI=No information UN=Unknown OT=Other	Admitting source.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
PAYER_TYPE_ PRIMARY	.	Categorization of payer type for primary payer associated with the encounter
PAYER_TYPE_ SECONDARY	.	Categorization of payer type for secondary payer associated with the encounter
FACILITY_TYPE	.	Description of the facility where the encounter occurred.
RAW_SITEID	.	Field for locally-defined identifier intended for local use; for example, where a network may have multiple sites contributing to a central data repository. This attribute may be sensitive in certain contexts; the intent is for internal network use only, and not to enable site quality comparisons.
RAW_ENC_TYPE	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_DISCHARGE_DISPOSITION	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_DISCHARGE_STATUS	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_DRG_TYPE	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_ADMITTING_SOURCE	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_FACILITY_TYPE	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_PAYER_TYPE_PRIMARY	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_PAYER_NAME_PRIMARY	.	Primary payer name as denoted in the source system. Used to derive PAYER_TYPE_PRIMARY if validated process does not exist.
RAW_PAYER_ID_PRIMARY	.	NOTE: values are all null/missing
RAW_PAYER_TYPE_SECONDARY	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_PAYER_NAME_SECONDARY	.	Secondary payer name as denoted in the source system. Used to derive PAYER_TYPE_SECONDARY if validated process does not exist.
RAW_PAYER_ID_SECONDARY	.	NOTE: values are all null/missing

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Table: DIGANOSIS

Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery.

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
DIAGNOSISID	.	Arbitrary identifier for each unique record.
PATID	.	Arbitrary person-level identifier. Used to link across tables.
ENCOUNTERID	.	Arbitrary encounter-level identifier. Used to link across tables.
ENC_TYPE	AV=Ambulatory Visit ED=Emergency Department EI=Emergency Department Admit to Inpatient Hospital Stay (permissible substitution) IP=Inpatient Hospital Stay IS=Non-Acute Institutional Stay OS=Observation Stay IC=Institutional Professional Consult (permissible substitution) TH=Telehealth OA=Other Ambulatory Visit NI=No information UN=Unknown OT=Other	Please note: This is a field replicated from the ENCOUNTER table. See the ENCOUNTER table for definitions.
ADMIT_DATE	.	Please note: This is a field replicated from the ENCOUNTER table. See the ENCOUNTER table for definitions.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
PROVIDERID	.	Identifier associated with the provider most responsible for the diagnosis.
DX	.	Diagnosis code. Some codes will contain leading zeroes, and different levels of decimal precision may also be present. This field is a character field, not numeric, to accommodate these coding conventions.
DX_TYPE	09=ICD-9-CM 10=ICD-10-CM 11=ICD-11-CM SM=SNOMED CT NI=No information UN=Unknown OT=Other	Diagnosis code type. We provide values for ICD and SNOMED code types. Other code types will be added as new terminologies are more widely used. Please note: The "Other" category is meant to identify internal use ontologies and codes.
DX_DATE	.	Date diagnosis was recorded, if known. NOTE: values are all null/missing
DX_SOURCE	AD=Admitting DI=Discharge FI=Final IN=Interim NI=No information UN=Unknown OT=Other	Classification of diagnosis source. We include these categories to allow some flexibility in implementation. The context is to capture available diagnoses recorded during a specific encounter.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
DX_ORIGIN	OD=Order/EHR BI=Billing CL=Claim DR=Derived NI=No information UN=Unknown OT=Other	Source of the diagnosis information. Billing pertains to internal healthcare processes and data sources. Claim pertains to data from the bill fulfillment, generally data sources held by insurers and other health plans.
PDX	P=Principal S=Secondary NI=No information UN=Unknown OT=Other	Principal discharge diagnosis flag.
DX_POA	Y = Diagnosis present N = Diagnosis not present U = Insufficient documentation W = Clinically undetermined 1 = Unreported / not used NI=No information UN=Unknown OT=Other	Flag to denote whether diagnosis was present on inpatient admission. Y = Diagnosis present at time of inpatient admission N = Diagnosis not present at time of inpatient admission U = Documentation insufficient to determine if the condition was present at the time of inpatient admission W = Clinically undetermined. Provider unable to clinically determine whether the condition was present at the time of inpatient admission 1 = Unreported / not used. Exempt from present-on-admission reporting.
RAW_DX	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_DX_TYPE	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_DX_SOURCE	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_PDX	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_DX_POA	.	Field for originating value, prior to mapping into the PCORnet CDM value set.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Table: PROCEDURES

Procedure codes indicate the discrete medical interventions and diagnostic testing, such as surgical procedures and lab orders, delivered within a healthcare context.

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
PROCEDURESID	.	Arbitrary identifier for each unique record.
PATID	.	Arbitrary person-level identifier. Used to link across tables.
ENCOUNTERID	.	Arbitrary encounter-level identifier. Used to link across tables.
ENC_TYPE	AV=Ambulatory Visit ED=Emergency Department EI=Emergency Department Admit to Inpatient Hospital Stay IP=Inpatient Hospital Stay IS=Non-Acute Institutional Stay OS=Observation Stay IC=Institutional Professional Consult (permissible substitution) TH=Telehealth OA=Other Ambulatory Visit NI=No information UN=Unknown OT=Other	Please note: This is a field replicated from the ENCOUNTER table. See ENCOUNTER table for definitions.
ADMIT_DATE	.	Please note: This is a field replicated from the ENCOUNTER table. See ENCOUNTER table for definitions.
PROVIDERID	.	Identifier of the PROVIDER most associated with the procedure order.
PX_DATE	.	Date the procedure was performed.
PX	.	Procedure code.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
PX_TYPE	09=ICD-9-CM 10=ICD-10-PCS 11=ICD-11-PCS CH = CPT or HCPCS LC=LOINC ND=NDC RE=Revenue NI=No information UN=Unknown OT=Other	Procedure code type. We include a number of code types for flexibility, but the basic requirement that the code refer to a medical procedure remains. Revenue codes are a standard concept in Medicare billing and can be useful for defining care settings. If those codes are available, they can be included. Medications administered by clinicians can be captured in billing data and Electronic Health Records (EHRs) as HCPCS procedure codes. Administration (infusion) of chemotherapy is an example. We are now seeing NDCs captured as part of procedures because payers are demanding it for payment authorization. Inclusion of this code type enables those data partners that capture the NDC along with the procedure to include the data. Please note: The “Other” category is meant to identify internal use ontologies and codes.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
PX_SOURCE	OD=Order/EHR BI=Billing CL=Claim DR=Derived NI=No information UN=Unknown OT=Other	Source of the procedure information. Order and billing pertain to internal healthcare processes and data sources. Claim pertains to data from the bill fulfillment, generally data sources held by insurers and other health plans.
PPX	P=Principal S=Secondary NI=No information UN=Unknown OT=Other	Principal procedure flag. NOTE: values are all null/missing
RAW_PX	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_PX_TYPE	.	Field for originating value, prior to mapping into the PCORnet CDM value set.
RAW_PPX	.	Field for originating value, prior to mapping into the PCORnet CDM value set.

Data Specification for the ITM Leaf Query Tool

Data Availability:
Leaf Website
Database Only
Not Available

Table: DEATH

Reported mortality information for patients.

Field Name	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*
PATID	.	Arbitrary person-level identifier used to link across tables.
DEATH_DATE		Date of death.
DEATH_DATE_IMPUTE	B=Both month and day imputed D=Day imputed M=Month imputed N=Not imputed NI=No information UN=Unknown OT=Other	When date of death is imputed, this field indicates which parts of the date were imputed.
DEATH_SOURCE	L=Other, locally defined N=National Death Index D=Social Security S=State Death files T=Tumor data DR=Derived NI=No information UN=Unknown OT=Other	
DEATH_MATCH_CONFIDENCE	E=Excellent F=Fair P=Poor NI=No information UN=Unknown OT=Other	For situations where a probabilistic patient matching strategy is used, this field indicates the confidence that the patient drawn from external source data represents the actual patient.