

PCORnet® Data Checks, version 20

Introduction

These checks are included in PCORnet’s foundational data curation package and applied to the core tables in PCORnet® Common Data Model v7.0. All network partners must pass the required data checks. Exceptions to investigative data checks are permitted and explained in an online annotated data dictionary.

Data Model Conformance Checks

Check	Description	Classification	Changes from v19
DC 1.01	Required tables are not present	Required	None
DC 1.02	Required tables are not populated. DEMOGRAPHIC, ENROLLMENT, ENCOUNTER, DIAGNOSIS, PROCEDURES, and HARVEST are required for all network partners; LAB_RESULT_CM, PRESCRIBING, and VITAL are required for network partners with electronic health record data	Required	None
DC 1.03	Required fields are not present	Required	None
DC 1.04	Required fields do not conform to data model specifications for data type, length, or name	Required	None
DC 1.05	Tables have primary key definition errors	Required	None
DC 1.06	Required fields contain values outside of data model specifications	Required	None
DC 1.07	Required fields have non-permissible missing values	Required	None
DC 1.08	Tables contain orphan PATIDs (PATIDs that are not present in the DEMOGRAPHIC table)	Required	None
DC 1.09	Tables contain orphan ENCOUNTERIDs (ENCOUNTERIDs that are not present in the ENCOUNTER table) for more than 5% of records	Required	None
DC 1.10	Replication errors between the ENCOUNTER, PROCEDURES and DIAGNOSIS tables. Replication errors are ENCOUNTERIDs in the DIAGNOSIS or PROCEDURES table where the encounter type or admit date does not match the corresponding value in the ENCOUNTER table	Required	None
DC 1.11	More than 5% of encounters are assigned to more than one patient	Required	None
DC 1.12	Tables contain orphan PROVIDERIDs (PROVIDERIDs that are not present in the PROVIDER table)	Required	None
DC 1.13	More than 5% of CPT®/HCPCS, CVX, ICD, NDC, LOINC, RXNORM or SNOMED codes do not conform to the expected length or content based on terminology-specific heuristics	Required	Added SNOMED
DC 1.14	Patients in the DEMOGRAPHIC table are not in the HASH_TOKEN table	Investigative	None
DC 1.15	Fields with undefined lengths that are present in more than one table (PATID, ENCOUNTERID, PRESCRIBINGID, PROCEDURESID, PROVIDERID, MEDADMIN_PROVIDERID, OBSGEN_PROVIDERID, OBSCLIN_PROVIDERID, RX_PROVIDERID, and VX_PROVIDERID) do not have harmonized field lengths	Required	None
DC 1.16	Laboratory results or clinical observations are recorded in the wrong table based on the LOINC® classtype	Investigative	None
DC 1.17	Zip codes in the ENCOUNTER or LDS_ADDRESS_HISTORY table have non-conforming values (alphabetical characters or the wrong number of digits)	Required	None
DC 1.18	Table refresh dates are not documented in the HARVEST table for populated tables	Required	None

PCORnet® Data Checks, version 20

Data Model Conformance Checks, cont.

Check	Description	Classification	Changes from v19
DC 1.19	More than 10% of the values in any hash token field are assigned to multiple patients	Investigative	None
DC 1.20	More than 5% of LOINC records in the LAB_RESULT_CM, PRO_CM, and OBS_CLIN tables are panel codes based on the LOINC® panel type	Investigative	None
DC 1.21	STATE_FIPS, COUNTY_FIPS and RUCA_ZIP fields have non-conforming values (alphabetical characters or discrepancies between STATE_FIPS and COUNTY_FIPS)	Required	None
DC 1.22	Patients have multiple current physical addresses in the LDS_ADDRESS_HISTORY table	Required	New

PCORnet® Data Checks, version 20

Data Plausibility Checks

Check	Description	Classification	Changes from v19
DC 2.01	More than 5% of records have future dates. Future dates are defined as those with dates occurring after the maximum refresh date in the HARVEST table	Investigative	None
DC 2.02	More than 10% of records fall into the lowest or highest categories of age (<0 or >89), height (<21 inches or >76 inches), weight (<0 lbs or >350 lbs), diastolic blood pressure (<40 mmHg or >120 mmHg), systolic blood pressure (<40 mmHg or >210 mmHg), or dispensed days supply (<1 day or >90 days)	Investigative	None
DC 2.03	More than 5% of patients have illogical date relationships. Illogical date relationships are defined as dates of service which occur before a patient's birth date or after a patient's death date; procedure dates occurring more than 5 days before the admit date or 5 days after the discharge date for the same encounter; or stop dates before start dates in the EXTERNAL_MEDS, MED_ADMIN, OBS_CLIN, OBS_GEN, or PAT_RELATIONSHIP tables	Investigative	None
DC 2.04	The average number of encounters per visit is > 2.0 for inpatient (IP), emergency department (ED), or ED to inpatient (EI) encounters. A visit is defined as a unique combination of PATID, ENC_TYPE, ADMIT_DATE, and PROVIDERID in the ENCOUNTER table	Investigative	None
DC 2.05	More than 5% of results for selected laboratory tests do not have the appropriate specimen source	Deprecated	None
DC 2.06	Median lab result values for selected laboratory tests are statistical or clinical outliers	Deprecated	None
DC 2.07	The average number of principal diagnoses per known DX_ORIGIN per encounter is above threshold [2.0 for inpatient (IP) and ED to inpatient (EI)]	Deprecated	Deprecated
DC 2.08	The monthly volume of encounter, diagnosis, procedure, vital, prescribing, medication administration, external medications, laboratory records, or patient-reported outcomes is an outlier. Outliers are defined as months with 0 records or a significant decrease compared to the average volume in the previous 12 months. Encounters, diagnoses, and procedures are limited to ambulatory (AV), telehealth (TH), emergency department (ED), ED to inpatient (EI), or inpatient (IP) settings.	Investigative	None
DC 2.09	Less than 80% of patients with a face-to-face encounter during the past 5 years have at least 1 face-to-face diagnosis and 1 vital measurement. Face-to-face is defined as an encounter type of ambulatory visit (AV), emergency department (ED), emergency department admit to inpatient hospital stay (EI), inpatient hospital (IP), or observation stay (OS).	Investigative	None

PCORnet® Data Checks, version 20

Data Completeness Checks

Check	Description	Classification	Changes from v19
DC 3.01	The average number of diagnoses records with known diagnosis types per encounter is below threshold [1.0 for ambulatory (AV), inpatient (IP), emergency department (ED), ED to inpatient (EI), or telehealth (TH) encounters].	Investigative	None
DC 3.02	The average number of procedure records with known procedure types per encounter is below threshold [0.75 for ambulatory (AV) encounters, 0.75 for emergency department (ED) encounters, 1.00 for ED to inpatient (EI) encounters, and 1.00 for inpatient (IP) encounters].	Investigative	None
DC 3.03	More than 10% of records have missing or unknown values for the following fields: DISCHARGE_DISPOSITION (IP/EI encounters only), DX_SOURCE, SEX, code fields [MEDADMIN_CODE, PRO_CODE], date fields [BIRTH_DATE, DISCHARGE_DATE (IP/EI encounters only), RX_ORDER_DATE, PX_DATE], ENCOUNTERID fields in selected tables [DIAGNOSIS, LAB_RESULT_CM, PRESCRIBING, MED_ADMIN, PROCEDURES, and VITAL], and provenance fields [DEATH_SOURCE, DX_ORIGIN, MEDADMIN_SOURCE, LAB_RESULT_SOURCE, PRO_SOURCE, PX_SOURCE, RX_SOURCE, VITAL_SOURCE)	Investigative	Added PRO_CODE and PRO_SOURCE. Removed 20 fields (all code type fields and all fields in CONDITION, DEATH_CAUSE, DISPENSING, EXTERNAL_MEDS, IMMUNIZATION, OBS_CLIN, and OBS_GEN)
DC 3.04	Less than 50% of patients with encounters have DIAGNOSIS records	Required	None
DC 3.05	Less than 50% of patients with encounters have PROCEDURES records	Required	None
DC 3.06	More than 10% of IP (inpatient) or ED to inpatient (EI) encounters with any diagnosis from a known DX_ORIGIN don't have a principal diagnosis from that source	Deprecated	Deprecated
DC 3.07	Encounters, diagnoses, or procedures in an ambulatory (AV), telehealth (TH), emergency department (ED), ED to inpatient (EI), or inpatient (IP) setting are less than 75% complete two months prior to the current month. Data completeness is calculated by comparing actual volume to the average volume during the previous year	Investigative	None
DC 3.08	Less than 80% of prescribing orders are mapped to a RXCUI which fully specifies the ingredient, strength and dose form (i.e. RXCUI codes that have a Term Type of SCD, SBD,BPCK, or GPCK)	Investigative	None
DC 3.09	Less than 80% of laboratory results are mapped to LAB_LOINC and have either a quantitative result (RESULT_NUM is not null and RESULT_MODIFIER is not NI, UN, OT, or null) or a qualitative result (RESULT_QUAL is not NI, UN, OT, or null)	Investigative	None
DC 3.10	Less than 80% of quantitative results for tests mapped to LAB_LOINC fully specify the normal range in the range and modifier fields	Deprecated	Deprecated

PCORnet® Data Checks, version 20

Data Completeness Checks, continued

Check	Description	Classification	Changes from v19
DC 3.11	Vital, prescribing, or laboratory records are less than 75% complete three months prior to the current month. Data completeness is calculated by comparing actual volume to the average volume during the previous year	Investigative	None
DC 3.12	Less than 80% of quantitative results for tests mapped to LAB_LOINC fully specify the RESULT_UNIT (i.e. RESULT_NUM is not null and RESULT_UNIT is not NI, UN, OT, or null)	Investigative	None
DC 3.13	The percentage of patients with selected lab tests is below threshold	Deprecated	None
DC 3.14	Medication administration, dispensing, or clinical observation records are less than 75% complete three months prior to the current month. Data completeness is calculated by comparing actual volume to the average volume during the previous year	Deprecated	Deprecated
DC 3.15	Less than 80% of medication administrations mapped to RXNORM are mapped to a RXCUI that fully specifies the ingredient, strength and dose form (i.e. RXCUI codes that have a Term Type of SCD, SBD,BPCK, or GPCK)	Investigative	None
DC 3.16	Less than 80% of clinical observations are mapped to an OBSCLIN_CODE and have a quantitative result (OBSCLIN_RESULT_NUM is not null and OBSCLIN_RESULT_MODIFIER is not NI,UN,OT or null), qualitative result (OBSCLIN_RESULT_QUAL is not NI,UN,OT or null) or narrative result (OBSCLIN_RESULT_TEXT is not null)	Investigative	None
DC 3.17	Less than 80% of quantitative results for tests mapped to OBSCLIN_CODE fully specify the RESULT_UNIT (i.e. OBSCLIN_RESULT_NUM is not null and OBSCLIN_RESULT_UNIT is not NI, UN, OT, or null)	Investigative	None
DC 3.18	Less than 80% of patients with a face to face visit in the past 5 years have ADDRESS_ZIP5 and COUNTY_FIPS address data for their current physical addresses. Face-to-face is defined as an encounter type of ambulatory visit (AV), emergency department (ED), emergency department admit to inpatient hospital stay (EI), inpatient hospital (IP), or observation stay (OS).	Investigative	New

Data Persistence Checks

Check	Description	Classification	Changes from v19
DC 4.01	More than a 5% decrease in the number of patients or records in a CDM table between the previous and current DataMart refresh	Investigative	None
DC 4.02	More than a 5% decrease in the number of patients or records for diagnosis, procedures, labs or prescriptions during an ambulatory (AV), telehealth (TH), other ambulatory (OA), emergency department (ED), or inpatient (IP) encounter between the previous and current DataMart refresh	Investigative	None
DC 4.03	More than a 5% decrease in the number of records or distinct codes for CPT/HCPCS, CVX, ICD10, LOINC, NDC, or RXNORM codes between the previous and current DataMart refresh	Investigative	None