**CTSA Common Metric (CM) Operational Guideline: Interoperable Clinical Data Availability and Completeness**

<table>
<thead>
<tr>
<th>Template Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Operationalized Metric Title</td>
<td>Interoperable clinical data availability and completeness</td>
</tr>
<tr>
<td>2. Rationale</td>
<td>To accelerate translation, researchers need to be provided with access to a broad range of data (electronic health records, omics, imaging, genetics, behavioral, etc.). These data can come from different sources such as clinical databases, research datasets, sensors, mobile technology, patient generated data, and publicly available data sets. The sharing and pooling of data within and across CTSA Program hubs requires that data be represented in a format that can be queried and adheres to commonly accepted standards. A longer-term goal of the CTSA Program is to harmonize data and data standards so that a query written by any site can be run unaltered against all CTSA Program data repositories. Before we can complete such a data harmonization, we first need an understanding of what types of data are being collected, managed, and stored in each hub’s clinical data repository and how much of this data is in a standard format. This Common Metric will provide a baseline scan of the level of coverage of the types of data that each hub should have in their clinical data repository. The purpose of this common metric is to identify clinical data gaps and opportunities for improvement. This common metric will improve local as well as network capacity to efficiently use data to conduct research. Specifically, the metrics on different data types will aid local hubs in prioritizing additions to their data repositories. Improving the CTSA Program’s clinical research data ecosystem can enhance the effectiveness of collaborative initiatives within and outside of the CTSA Program to provide tools to identify patient cohorts (i2b2/ACT, SHRINE, All of Us, PCORI).</td>
</tr>
<tr>
<td>3. Operational Specification</td>
<td>Level of availability and completeness of the baseline types of data in a standard (CTSA-interoperable) format within a clinical data repository at the CTSA Program primary institution (hub).</td>
</tr>
</tbody>
</table>
The list of data types has been generated by the Informatics Domain Task Force (iDTF); Additional types of data will be considered in the future.

4. Technical Description

**Data Scope:** Data from centralized primary resources available to researchers at the primary hub institution. If patients can be unambiguously counted across multiple data repositories, the aggregate counts may be reported for the hub.

**Key Definitions:**

- **Clinical Research Data Repository:** For the purposes of the CTSA Program and this metric, a clinical research data repository is a standards-based repository of clinical and/or research data (at the individual subject, client, or patient level) available to investigators for research at a CTSA Program hub. This generic term includes clinical data repositories, research data repositories, or enterprise data warehouses.

- **Data Model:** A data model is a descriptive design of data and its relationships. The data model organizes data elements and standardizes how they relate to one another and to properties of the real-world entities. Data models approved for use by the informatics Domain Task Force (iDTF) for this metric include OMOP, PCORnet, i2b2/ACT, or other i2b2 data model. Hubs wishing to use other data models should suggest additions (along with a list of other adopting hubs) to the coordinating center, who will coordinate with the iDTF for consideration.

- **Standard Value:** The rules for values of data elements in a database; will ensure consistency.
  - LOINC ID: Logical Observation Identifiers Names and Codes: [https://loinc.org/](https://loinc.org/)
  - SNOMED: A clinical terminology created by a range of healthcare specialists to support clinical decision-making and analytics in software programs:
<table>
<thead>
<tr>
<th><strong>Informatics Common Metric (ICM) OG</strong>&lt;br&gt;<strong>Updated October 14, 2019</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>○ Notes/Narrative: Free text documentation entered during a clinical encounter that refers to a meeting between a person and a health care provider whose services are provided.</td>
</tr>
<tr>
<td>○ Observations: A measurement of a single variable or a single value derived logically and/or algebraically from other measured or derived values. A test result, a diastolic blood pressure, and a single chest x-ray impression are examples of observations. In certain circumstances, tracings and images may be treated (by Health Level 7 (HL7), which is an application protocol for electronic data exchange in health care environments) as individual observations and sent as a single observation message. Observations are intended to cover all types of patient specific observation reports except pharmacy.</td>
</tr>
<tr>
<td>● <strong>OMOP Data Model:</strong></td>
</tr>
<tr>
<td>○ Documentation for the OMOP CDM is found on Github (<a href="https://github.com/OHDSI/CommonDataModel/wiki/Background">link</a>) and the OMOP CDM Wiki (<a href="https://www.ohdsi.org/">link</a>). The OMOP Common Data Model is an open-source, community standard for observational healthcare data that is managed by Observational Health Data Sciences and Informatics (OHDSI). OHDSI has been established as a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics. The OHDSI collaborative includes all of the original OMOP research investigators, and will develop its tools using the OMOP Common Data Model. Learn more at <a href="https://www.ohdsi.org/">https://www.ohdsi.org/</a>. OMOP CDM contains person, condition, drug, procedure and visit information, and provider and cost information. This will support health economics use cases and medical treatment outcome studies, including medical device safety, comparative effectiveness and healthcare quality. For more information see <a href="https://github.com/OHDSI/CommonDataModel/wiki/Background">https://github.com/OHDSI/CommonDataModel/wiki/Background</a>.</td>
</tr>
<tr>
<td>Use of Data Models with TriNetX:</td>
</tr>
<tr>
<td>---------------------------------</td>
</tr>
<tr>
<td>- <strong>PCORnet Common Data Model:</strong></td>
</tr>
<tr>
<td>○ The PCORnet CDM (see <a href="http://www.pcornet.org/pcornet-common-data-model/">http://www.pcornet.org/pcornet-common-data-model/</a>) is based on the FDA Sentinel Initiative Common Data Model and has been informed by other distributed initiatives such as the Health Care Systems Research Network, the Vaccine Safety Datalink, various AHRQ Distributed Research Network projects, and the ONC Standards &amp; Interoperability Framework Query Health Initiative. The PCORnet CDM leverages standard terminologies and coding systems for healthcare (including ICD, SNOMED, CPT, HCPSC, and LOINC) to enable interoperability with and responsiveness to evolving data standards. See: <a href="https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model/sentinel-common-data-model">https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model/sentinel-common-data-model</a></td>
</tr>
<tr>
<td>- <strong>i2b2/ACT Data Model:</strong></td>
</tr>
<tr>
<td>○ Informatics for Integrating Biology and the Bedside (i2b2) was an NIH-funded National Center for Biomedical Computing (NCBC) based at Partners HealthCare System. The i2b2 NCBC developed a scalable informatics framework that is designed to bridge clinical research data and the vast data banks arising from basic science research in order to better understand the genetic bases of complex diseases. I2b2 was designed for cohort identification. The i2b2 framework employs a simple, yet powerful data model. It consists of facts and dimensions. A fact is the piece of information being queried, and the dimensions are groups of hierarchies and descriptors that describe the facts. The i2b2 database utilizes a star schema that consists of one fact table surrounded by numerous dimension tables. Facts in i2b2 are observations about a patient, including things like diagnoses, demographics, laboratory results, etc. See: <a href="https://www.i2b2.org/">https://www.i2b2.org/</a>.</td>
</tr>
<tr>
<td>○ Accrual to Clinical Trials (ACT) is a network of 21 CTSA Program hubs that use a common informatics platform for cohort discovery. See: <a href="https://www.act-network.org/node/29">https://www.act-network.org/node/29</a></td>
</tr>
</tbody>
</table>
TriNetX is a global health research network connecting healthcare organizations (including 25 CTSA Program hubs), biopharma and contract research organizations. The TriNetX platform enables cohort identification and hypothesis generation based on clinical data that can currently be sourced from a common data model (i2b2, OMOP, NAACCR, etc.), flat files, or via NLP of narrative documents. See: https://www.trinetx.com/

<table>
<thead>
<tr>
<th>Data Domain</th>
<th>Standard Value</th>
<th>Numerator</th>
<th>Denominator</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td>N/A</td>
<td>count of unique patients with an age/DOB value</td>
<td>Count of all patients in the Data Repository</td>
<td>% patients with age/DOB value</td>
</tr>
<tr>
<td>Patient</td>
<td>Administrative Sex</td>
<td>count of unique patients with administrative sex value</td>
<td>Count of all patients in the Data Repository</td>
<td>% patients with administrative sex value</td>
</tr>
<tr>
<td>Labs</td>
<td>LOINC ID</td>
<td>count of unique patients with a LOINC ID value</td>
<td>Count of all patients in the Data Repository</td>
<td>% of patients with LOINC ID value</td>
</tr>
<tr>
<td>Medications / Drugs</td>
<td>RxNorm ID</td>
<td>count of unique patients with a RxNorm ID value</td>
<td>Count of all patients in the Data Repository</td>
<td>% of patients with RxNorm ID value</td>
</tr>
<tr>
<td>Conditions / Diagnosis</td>
<td>ICD 9/10 or SNOMED</td>
<td>count of unique patients with an ICD 9/10 or SNOMED value</td>
<td>Count of all patients in the Data Repository</td>
<td>% of patients with ICD 9/10 or SNOMED value</td>
</tr>
<tr>
<td>Procedures</td>
<td>ICD 9/10 CPT</td>
<td>count of unique patients with an ICD 9/10 or CPT Procedure Code</td>
<td>Count of all patients in the Data Repository</td>
<td>% of patients with ICD 9/10 or CPT Procedure value</td>
</tr>
<tr>
<td>Notes / Narrative</td>
<td>N/A</td>
<td>count of unique patients with free text data</td>
<td>Count of all patients in the Data Repository</td>
<td>% of patients with free text data</td>
</tr>
<tr>
<td>Observations</td>
<td>N/A</td>
<td>Presence of Observations or Absence of Observations</td>
<td>N/A</td>
<td>Presence of Observations or Absence of Observations</td>
</tr>
<tr>
<td>--------------</td>
<td>-----</td>
<td>-----------------------------------------------------</td>
<td>-----</td>
<td>-----------------------------------------------------</td>
</tr>
</tbody>
</table>

5. Metric Type, Score(s), Numerator and Denominator Statements or Continuous Variable Statement, Inclusion/Exclusion Criteria

In addition to the metrics, hubs will be asked to report:
- Data model they used to report this metric: [OMOP, PCORnet, i2b2/ACT, or other i2b2 data model]
- If they worked with TriNetX to report this metric.
- How many unique patients are represented in the data model.

This metric results in eight scores of data domain-based values (see Table above):

Metric Type: Percent

Metric: Percent of unique patients with the standard value: [%] (see Table above):

To calculate this metric for each data type domain use the following:
- Count of unique patients with the standard value (numerator)
- Count of unique patients within the clinical data repository (denominator)
- % of unique patients with the standard value (% = [numerator]/[denominator])

Inclusion/Exclusion Criteria:
- Scripts do not contain a timeframe and thus this first year collects the metric for the entire clinical research data repository. In subsequent years the scripts will be modified to contain a shorter time frames (possibly 5 years, to be decided by the iDTF) to highlight improvements over time.

6. Data Sources & Methods of Data Collection

Data Sources: CTSA Program hub clinical research data repository.

- Aggregate counts may be reported for the hub and clinical affiliates, if individuals in the repository can be unambiguously identified and counted across individual data repositories.
Note that the “clinical research data repository” may be a clinical repository, research data repository or dual use repository.

- Each reporting period hubs should choose a single data model that would best represent their research data warehouse.
- Hubs may choose to change which data model they use for reporting in subsequent years if beneficial to the strategic management of this metric at their hub.

Note about data completeness: if a data domain is not available hubs should state “data not available”. Data may not be available for a number of reasons (data model does not capture this data, the institution does not allow this data to be captured due to HIPAA, etc).

- Note about percent targets for data domains:
  - It is not expected that every data domain will achieve 100% as some data domains may not record a value. For example, 100% target assumes that every patient will be on a medication or drug.

**Method of Data Collection:**

- Queries/scripts will be provided to hubs for the OMOP, PCORnet and i2b2/ACT data models. These queries/scripts will enable standardized automated query against the data repository (or repositories) at each of the CTSA hubs. The scripts will be developed, tested, and approved collaboratively by the Informatics Common Metric Development Team with the iDTF. Data model scripts: https://github.com/ncats/CTSA-Metrics

- For hubs that choose to use the approved data models that are also utilized by TriNetX, TriNetX will provide the metric data to your hub directly for incorporation into your Scorecard. [Note: TriNetX supports the CTSA Program Common Metric for Informatics Solutions. As end users do not have direct access to the in-memory databases on their local or hosted appliances, TriNetX has created a report that can be requested by the individual CTSA Program hubs. A designated user from the site should send an email to CTSA@trinetx.com requesting the CTSA Program Common Metric report. TriNetX will then create the report and send it to the requesting site admin.]

- For hubs that choose to use i2b2 with their own hub-specific data model, that hub will be responsible for generating the script for their own purposes of reporting the required metric.
These hubs are encouraged to share the script on the GitHub site, however, there is no assurance that other hubs would be able to validate the script.

- Data models approved for use by the informatics Domain Task Force (iDTF) for this metric include OMOP, PCORnet, i2b2/ACT, or other i2b2 data model. Hubs wishing to use other data models should suggest additions (along with a list of other adopting hubs) to the coordinating center, who will coordinate with the iDTF for consideration (see [here](#)).

<table>
<thead>
<tr>
<th>7. Frequency of Data Collection and Reporting</th>
<th>Data collected 1x a calendar year, by the end of the first quarter. Reported 1x a year – by August 31st.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Unit of Analysis</td>
<td>Data will be collected within each hub from the proposed data type (domain) level and reported at the hub level.</td>
</tr>
<tr>
<td>9. Notes/Comments</td>
<td><strong>Note:</strong> Although 8 metrics will be reported, hubs should develop only 1 turn-the-curve plan.</td>
</tr>
<tr>
<td></td>
<td><strong>How can we use this to strategically manage the hub?</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Strategic Management:</strong> This metric demonstrates the extent to which data are complete and use standard approaches. The metric will allow CTSA Program hubs to take the first step necessary to enable a hub to view the breadth and depth of data and through local strategic management activities and turn the curve exercises, hubs will be able to identify factors behind clinical data gaps (e.g., technological, financial, unavailable?) and opportunities for strategic investment of time, funds, and/or focus.</td>
</tr>
<tr>
<td></td>
<td><strong>Aspirational Considerations:</strong> This metric will evolve over time to enhance the completeness of the data repositories across the CTSA Program consortium and incorporate additional types of data within repositories as the CTSA Program, in collaboration with the iDTF, finds useful and appropriate. Hubs are encouraged to test the collection of additional data domains to inform iDTF discussions about the evolution of this metric. If hubs have collected a “stretch metric” to inform the consortium</td>
</tr>
<tr>
<td>about the future of this metric please inform CLIC so they can convey to the iDTF: <a href="https://clic-ctsa.org/common_metrics/help-desk">https://clic-ctsa.org/common_metrics/help-desk</a></td>
<td></td>
</tr>
</tbody>
</table>