Improving Healthcare Interoperability - Project Report

Grant support provided by the Pew Charitable Trusts | **08DEC2018**

Improving Healthcare Interoperability - Project Report
Duke Clinical Research Institute
Grant support provided by the Pew Charitable Trusts

EXECUTIVE SUMMARY

The healthcare ecosystem is in a state of flux with respect to the interoperability of clinical data. The call for healthcare data interoperability – and the promise of clinical data informing the learning healthcare system – is neither new nor novel, with work toward accomplishing the goal of data liquidity involving countless stakeholders extending across multiple decades. Despite the general acknowledgement of the need for interoperability, a plethora of initiatives, and federal leadership, the current state is quite distanced from the envisioned goal.

This project, ***Improving Healthcare Data Interoperability*** (the Project), intended to evaluate the state of healthcare data semantic interoperability and suggest recommendations for the roadmap forward, specifically from the perspective of clinical registries. The focus on clinical registries was purposeful – these registries produce actionable data for quality improvement and patient safety and should thus be at the center of healthcare data interoperability. The key technical goals were to 1) determine the scope of the problem, by identifying core common clinical concepts captured across a large sample of registries and determining the degree of concordance of registry database (physical) representations with common data models and national / international data standards, and 2) create an authoritative implementation guidance targeting database developers for the programming of the common core clinical concepts such that interoperability would be intrinsic to those data elements and not require a translation layer to communicate between computer systems. The findings of the technical work were then to inform recommendations for registry owners (and the larger healthcare ecosystem) regarding the roadmap forward.

Singularly, the key finding was that the registry community has not benefitted from, is not aligned with, and does not contribute to interoperability efforts. For example, 11 registries captured information about tobacco use – with each using a different set of questions, and only 1 conforming to the USCDI. Furthermore, the construct of the USCDI for capturing tobacco use was felt by all to be less than ideal from the perspective of clinical utility of the data. The limited efforts by professional societies to contribute to data standards in service of registry needs have not been sustained. As a result, the predominant model for obtaining data for registries is still overwhelmingly forms-based manual chart abstraction and data re-entry – not electronic data capture at the point of care and transmission of that data by direct electronic mechanisms. This model of data capture essentially precludes the sharing of standardized data among registries – there is no cross-registry standardization at the data element level. Furthermore, there have been only limited initiatives to re-engineer the processes of healthcare documentation to capture data (rather than documents) at the point of care. Interpretation of the far-ranging discussions occurring across the project suggest that the failure to achieve healthcare data interoperability stems at least in part from 2 erroneous constructs: 1) that the interchange of clinical documents (e.g., clinical notes, procedure reports) can be equated with and accomplishes clinical data interoperability, and 2) that interoperability can be accomplished by standards that specify how data should be exchanged between entities, rather than at a native (database) level within entities. The current state of (the lack of) healthcare data interoperability seems sufficient evidentiary proof that those constructs are fundamentally flawed.

The following three domains further describe the key findings and recommendations of the Project.

The Implementation Guidance

The primary work product of the Project was the identification of the clinical concepts (and concept classes, specifically medications and laboratory data) commonly shared across the contributing registries, and the specification of 13 key metadata for each of these concepts. As to the form of the implementation guidance, it is purposefully tailored for two sets of stakeholders: clinicians and database developers. For the clinician perspective, the key metadata for each clinical concept include a clinical concept label (e.g., text to print on a case report form or display as a prompt), the clinical definition of the concept (e.g., for a data dictionary), clinical allowed values (i.e., the permissible responses, or "value set" that captures the possible answers to the concept prompt), and definitions of each of the allowed values. In addition to the clinical metadata, database developers (programmers) are anticipated to need an additional 4 metadata to program these concepts as discrete data fields in databases: a database field label (i.e., the database "address" of the data element), data field type, data field business rules (e.g., indexing, cardinality, edit and range checks), and database allowed values (these may be different than the clinical allowed values). The final 5 of 13 metadata (object identifier, or "OID", reference concept binding, reference allowed value bindings, FHIR references, and notes) are included for reference purposes. The latter metadata are typically not needed to program databases but become relevant when validating and verifying the electronic interchange of data between computer systems. A copy of the implementation guidance is available at https://dcri.org/registry-data-standards.

The approach of the Project was to author the implementation guidance to achieve maximal utility with minimal unnecessary content for clinician and database developer stakeholders. We acknowledge that the representation is somewhat unconventional and incomplete, at least from the perspective of formal informatics modeling. Instead, we erred on the side of developing a practical, authoritative reference resource. The exercise itself identified the following. 1) The authoring of the implementation guidance proved much more difficult than anticipated, particularly since these were base concepts common across healthcare. Particularly revealing was that there was no single source or set of sources where all of the metadata were housed – even our 13 limited metadata had to be culled from multiple sources. 2) Many of the core common data elements and corresponding allowed values lacked clear-cut clinical definitions. 3) The "right" answers for some of metadata elements were not readily apparent, particularly for reference bindings. Oftentimes different reference sources had redundant or inconsistent recommendations.

Based on the experience of the Project, we believe the following can be recommended. First, a single, unifying process should be identified and specified for the development of clinical concepts into common data elements, and the instantiation of those data elements as data standards, focusing on the 13 metadata identified in the Project. This process must enable the stewarding of clinical vocabularies by the appropriate stewards (e.g., professional societies). Second, coupled with this unifying process, an open-access repository should be assigned and resourced that houses the data elements and information models thereof. Within the repository, views of the common data elements relevant to clinicians and database developers should be the built-in, default views - rather than the complex and intimidating "more is more" table views we encountered. The 2 default views include the clinically relevant perspective (the first 4 metadata – this should simulate clinical data standards publications from professional societies and other clinical data standards initiatives) and the database developer perspective (particularly the first 8 metadata). Third, specification of the 13 metadata should be advanced as the technical tooling required of professional societies (i.e., registry owners) in their respective registry initiatives. Our work suggest that these 13 metadata may be sufficient for registry purposes and may be all that are needed by larger electronic health record and health information technology communities to support native interoperability. In addition, the 13 metadata provide an excellent starting point for formal informatics modeling of the concepts. Fourth, stewardship of common data elements is necessary – for the core common clinical data elements identified via the Project, this could be a federal agency, standards development organization, or other business entity. In this specific example, with completion of the Project, resourcing for long-term

stewardship of the Project terminology set was also eliminated. And finally, the issues underlying the adage "the great thing about standards is that there are so many to choose from" should be addressed. Specifically, conflicting and overlapping approaches should be deprecated with domains of responsibility better identified, particularly among the standards development organizations. This will obviously require cooperation and coordination across the standards community.

Implications for Registry Owners

From the perspective of clinical registries, semantic interoperability remains a distant goal. The Project identified multiple issues facing registries and their owners. Most critically is the multifaceted enormity of the problem – i.e., the complexity of healthcare and the numerous potential insertion points for registries and registry owners. The lack of a unifying framework for those insertion points (e.g., identification and definition of domain-specific clinical terminologies, an authoritative process for specifying data elements in a manner that can be easily and consistently shared among HIT companies, structured reporting for the capture of data integrated into clinical workflows, non-conflicting and singular terminology bindings, data interchange containers that serve the explicit needs of registries for electronic data submission, etc.) undermines efforts to build solutions that satisfy more than just local registry needs. In addition, participation in frameworks external to the registry requires initiative, leadership, governance, and resources. Then there is the misalignment of benefit versus expenditure – while working within a larger framework has long-term benefits for all, it can be anticipated that those benefits will only be partially recovered and received by the registry owners themselves. Furthermore, there is lack of federal leadership and funding to accomplish and enable participation of registries in an interoperability framework.

The current state – fit to purpose registries built and managed largely independent of an interoperability framework, with data submission dependent on manual chart abstraction – is not financially viable in the long-run. The obvious opportunity presented to professional societies is to shape healthcare delivery such that high quality, clinically relevant data servicing the data needs of both clinical care and of registries are captured at the point of care as a routine process integrated into care delivery. This model could dramatically improve the availability of clinical data for a myriad of other purposes, particularly the capture of real-world data to be used to generate real-world evidence. A recommended set of responsibilities for registry owners includes the following:

- Develop leadership and governance within each registry organization that values and prioritizes interoperability of healthcare data, establishing the goal of having healthcare enterprises manage well-defined, high quality, interoperable clinical data fit that serves registry purposes – from the capture of that data integrated into source documentation all the way to submission to registry databases – embracing the "capture once, use many times" mantra
- Recognize, contribute, and steward those components of the interoperability framework that depend upon clinical subject matter expertise possessed by registry owners, particularly the development of domain-specific clinical concepts as standardized common data elements
- Develop, organize, and manage in-house expertise in informatics, and particularly the informatics of interoperability, and require the use of interoperability standards for the capture of source data in EHR systems for submission of data to registries – inclusive of assignment of resources that specifically participate and contribute to the interoperability framework

Implications for Healthcare

The work of the Project exposed many difficulties in using resources currently available regarding common data elements for the purpose of building database systems. The expectation was that it would be relatively straightforward to author the implementation guidance for the core concepts identified in the Project, given the limited number of common data elements and their longstanding use across healthcare. This assumption proved erroneous – it took a large community of informatics experts to arrive at our (still imperfect) implementation guidance.

The gaps (and corresponding recommendations) appear to include at least the following. First, there is a lack of an overall interoperability framework (perhaps better thought of as a framework of frameworks) that identify and delineate how the component pieces (e.g., data elements, database systems, standards, registries, quality analysis and reporting, clinical decision support, governance, etc.) fit together. An interoperability framework – with explicit expectations and rules for the components – could only enhance the probabilities of success, scale, and sustainability of work involving any component of the framework. Second, a single, authoritative repository of common data elements with appropriate views (e.g., clinician view, database programmer view) is needed – instead, we found multiple repositories, each with sets and subsets of the requisite information, with some in conflict with others. Third, several of the clinical concepts in the USCDI are not fit for purpose from the registry (and clinical) perspective. Tobacco use is a prime example that has already been described. A process for revision of these data elements, focusing on their clinical use, is needed, particularly for tobacco abuse and the illicit use of other substances. Fourth, there are substantial overlaps, inconsistencies, and incongruities among the ontologies of standards development organizations. The explicit request made of standards development organizations is for better coordination across their respective domains. Fifth, a consistent finding was the lack of clinically relevant standardized definitions of the clinical concepts – depending on the steward of a clinical concept, a responsibility of the steward should be the development of clinically relevant definitions of terms and their respective allowed values. And sixth, the lack of participation of registries in the interoperability framework must be addressed. Registries (and their respective owners) are in unique position to dramatically and favorably impact the capture of real-world data that can enable clinical evaluation, research and discovery, drug and device safety and surveillance, quality and performance assessment, risk modeling and resource allocation, and clinical decision support – but only if fundamental changes are made to the healthcare ecosystem to enable and resource those efforts.

OVERVIEW

The premise of discrete clinical data informing the practice of medicine is nearly as old as the computer itself. In 1969, Eugene A. Stead, Jr., MD, established the Duke Databank for Cardiovascular Disease to capture cardiovascular procedural and care data and compile the "computerized textbook of medicine". (1) Federal efforts of the past several decades have stimulated the adoption of health information technology to record clinical information across all of health care. In 2004, the position of the National Coordinator for Health Information Technology was created by Executive Order. This led to the convening of the Health Information Technology Standards Panel and the National Health Information Network, two key initiatives intended to drive the exchange of health information via standards. The landmark 2009 Health Information Technology for Economic and Clinical Health (HITECH) legislation authorized a financial impetus for the adoption of certified electronic health record (EHR) systems for the more than 75% of clinical enterprises that at the time were predominantly using paper medical records. HITECH established 2 federal advisory committees, the HIT Policy Committee and the HIT Standards Committee. Critically, the Clinical Operations Workgroup of the HIT Standards Committee recommended that standards be specified to accomplish interoperability *between entities, rather than within an entity*. Explicitly, the Workgroup recommended that standards should *not* apply to internal processes for data capture, storage, and use, but instead should apply only at the level of data exchange. (2)

While the HITECH act can be credited for the dramatic increase in the adoption and use of EHR systems across US healthcare, interoperability of clinical data therein has proven elusive. The 2015 report from the Office of the National Coordinator, *Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap* recognized that despite the investment in EHR infrastructure and the specification of frameworks and standards for the exchange of clinical information, the "interoperability experience remains a work in progress". (3) Just 1 year later, in 2016, the 21st Century Cures Act legislation signaled the intent of Congress to accelerate the interoperability agenda. Specifically, the legislation included a definition of interoperability as health information technology that "(A) enables the secure exchange of electronic health information with, and use of electronic health information from, other health information technology without special effort on the part of the user; (B) allows for complete access, exchange, and use of all electronically accessible health information for authorized use under applicable State or Federal law; and (C) does not constitute information blocking as defined in section 3022(a) of the PHSA as amended." (4) Further signaling the need to accomplish interoperability, the Centers for Medicare and Medicaid Services has now abandoned the term "meaningful use" (with respect to stage 3 of the EHR Incentive Program requirements) in favor of "promoting interoperability." (5)

There are daily reminders that the potential benefits of healthcare data liquidity have not been realized. Consider the more than 1 billion secure and private transfers of confidential banking data that occur daily, or the behavioral targeting based on personal data that "enhances" the consumer experience at e-commerce sites. Within healthcare, however, the movement of clinical information occurs mostly at the document level – not at the data level of granularity that the banking or e-commerce industries utilize. Where retail has long utilized the Universal Product Code (UPC) to uniquely identify merchandise and electronically manage inventory, the medical device Unique Device Identifier (UDI) is only in the early stages of adoption. Within healthcare, the only notable exception in terms of discrete, explicit data is the electronic submission of billing data from healthcare organization to payer. While attractive as a potential source of data, the discordance between claims and clinical data (and the lack of highly granular clinical data) has yet to be overcome. (6)

For several decades, professional societies and other organizations have been developing registries for the evaluation of metrics of quality and process, to enable site-level performance improvement, and to support health outcomes research. A prototypic example is the portfolio of the American College of Cardiology National Cardiovascular Data Registries (NCDR). (7) Since the inception of the NCDR in 1997, the clinical data concepts captured in the NCDR databases have been explicitly defined and

published via (NCDR-specific) data dictionaries. This standardization of the nomenclature facilitates the capture of clinically consistent data regardless of size or composition of the healthcare entity and has been instrumental in the analysis of the data across those entities (e.g., quality and performance benchmarks). The nomenclature, however, has been unique to the NCDR, without formal development via an external, standards development organization (SDO) standardization process or other informatics formalisms. As such, NCDR data elements are not defined at the level of programming into database systems, and there is no standardization across EHR or other electronic health information (EHI) documentation systems as to *how* the individual data elements are stored. Instead, capture of data in an NCDR registry is based on traditional paper-based forms, with human chart abstraction of the information from electronic health information (EHI) system and hand keying of the data for registry submission. With multiple hundreds of registries today, the burden of chart abstraction is now costing healthcare well into the billions of dollars per year. This has raised the question of whether clinical data could be captured at the point of care integrated into clinical workflow, and then transmitted electronically without abstraction to registries – i.e., structured capturing and reporting of semantically interoperable clinical terminology from point of care through submission of data to those registries. (8)

A foundational premise of this project is that *native data interoperability* describes the ideal, desired state of clinical data as stored in healthcare databases – i.e., storage of data using identical physical (programming) representations among databases. With standardization of the programming representation, the need for proprietary processes and unique solutions at each data transfer point (termed Extract, Transform, Load, or ETL) would be reduced or even eliminated. The result would be true semantic interoperability implemented at the native database level – i.e., the ability of two information systems to readily exchange the data without ETL while retaining the semantic meaning of the data. Instead, ETL remains an expensive operational barrier that reduces data liquidity and contributes to the increasing cost of healthcare.

Key to accomplishing true semantic interoperability is standardization of nomenclature, not just at the interfaces between entities, but at both the data source (the healthcare system) and the data recipient (in the case of the Project, registries). Recognition of the value of standardized nomenclature to describe the findings of medicine dates to at least the 1850's work of Florence Nightingale and her systematic description of the causes of battlefield mortality. Despite a generation plus of work by SDOs, the goal of standardized healthcare information – and in particular, interoperable clinical data – remains elusive. While the "Big 5" healthcare terminologies (International Classification of Diseases – Clinical Modification [ICD-10 CM], Current Procedural Therapy [CPT], Systematized Nomenclature of Medicine - Clinical Terms [SNOMED CT], Logical Observation Identifiers, Names and Codes [LOINC], and RxNorm) have been incorporated into EHI systems, these terminologies are primarily positioned as secondary encoding schemas, rather than as data standards used in the primary capture of clinical data, or even as meta-data to enable the semantically interoperable transfer of data *between* EHI systems (including the transmission of clinical data to registries).

The Project hypothesized that there is little compliance at the operational (physical database) level in EHI systems, specifically registry databases, with the terminologies and interchange approaches developed by SDO organizations. We then sought to reduce barriers to compliance by creating a database programmer's implementation guidance for achieving native interoperability. To evaluate this hypothesis, we sought to identify a set of common clinical concepts (e.g. smoking status, medications data, unique device identifiers) shared by registries, and to evaluate compliance with terminology standards in the representations of those concepts in the respective database registries. An environmental scan along with evaluation of national data models was used to identify existing relevant data standards. To resolve the (expected) issue of non-use of data standards, we also formalized our findings by developing a guide for the physical (programming) implementation of data elements for the common concepts identified in the project. Specifically, the project had 3 aims:

Aim 1: To compare the case report forms (CRFs) and other registry database artifacts of at least 20 different disease and device registries of professional societies, to identify the clinical data elements common (>50% prevalence) across those registries.

Aim 2: To evaluate the identified common registry data elements in the context of the "Big 5" healthcare data standards, US Core Data for Interoperability, Fast Healthcare Interoperability Resources (FHIR) draft standards, and national data models (Observational Medical Outcomes Partnership (OMOP) / Observational Health Data Sciences and Informatics (OHDSI), FDA Sentinel, Patient-Centered Outcomes Research Network (PCORnet)), determining concordance of the data elements and suggesting the most appropriate metadata for those data elements per existing data standards.

Aim 3: To describe the current state of the clinical data standards environment, identifying the limitations of that framework, for the purpose of producing a roadmap (implementation guidance for database programmers) that catalyzes the governance, structural, operational, and technical transformations needed to implement a common clinical data element set across EHI systems, registry systems, and national data models.

WHY TARGET REGISTRIES?

- **Registries Have a Central Role in Healthcare**

A clinical registry has been defined as "an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes." (9) For the past several decades clinical registries have played a critical role in understanding how care is provided across clinician and organizational boundaries and over time. However, submission of data to clinical registries has typically been labor intensive, costly, and focused on retrospective analysis. Despite these challenges, professional society registry stewards and other organizations have continued to invest in registry infrastructure because of the value of consistently coded national-level data for quality measurement, improvement, research and other uses.

There are several published frameworks which identify ranges of functionality, data sources, and business objectives of registries used for health care and research. (10, 11) Registries often are designed for or support a single business purpose (12), with those supporting patient-centered outcomes research usually also support operational needs (13). Central to the success and value of registry data is the uniformity of the clinical and administrative data definitions adhered to by the registry as defined by a group of clinical subject matter experts. A deep understanding of the body of literature in the clinical domain, the workflow in which those data are generated during patient care activities, and how the data might be used by providers and other team members are critical elements. Registries have traditionally been highly effective at creating these definitions within their own specialty and scope of practice.

Data submitted to registries has typically not been encoded in EHR systems and captured as structured data integrated into care delivery workflow. Instead, the typical model is manual chart abstraction and data re-entry into electronic systems. A 2017 survey on the status of US registries notes that while the predominant purposes of registries are for quality improvement and clinical research informing value-based payment models, 88% percent utilized manual data entry to accomplish same. (12) Lack of data interoperability was cited as the top barrier to registry development and improvement.

Not all registries make their data dictionaries available to the public, a factor that contributes to the proliferation of similar but non-identical clinical content concepts in registry data models and further degrades data liquidity. Without transparency and harmonization of related but non-aligned content, the path to EHR-enabled data extraction across registries remains difficult. Thus, despite the promise of

health IT-enabled data reuse to inform a learning health system, a large gap remains between technical implementation and the desired state. As currently constructed, data standards simply do not enable data interoperability at a native, data element by data element level. A significant number of registries and clinician-focused organizations have recognized this as an opportunity to begin efforts to clinically define common clinical data elements. (14-16) While these efforts aim to enable the seamless, universal, minimal-resource extraction and exchange of data between EHRs and registries, even these initiatives still typically lack sufficient technical specification to accomplish native data interoperability. Indeed, the PCPI National Quality Registry Network and the Council of Medical Specialty Societies have identified this paradigm – native data interoperability – as a keystone of their respective efforts to encourage interoperability between source EHI systems and registries.

- **Registries Are Positioned to Facilitate Semantic Interoperability**

Efforts by the United States government to facilitate interoperability (e.g., the Medicare EHR Incentive Program, commonly known as Meaningful Use) have markedly increased the amount of information captured and stored electronically. However much of this information is in the form of documents, with the result being that EHR systems are essentially sophisticated electronic filing systems designed to optimize remuneration and compliance. Only a very limited amount of discrete clinical information is captured and stored as structured data. Even there, linkage of data still often requires manual intervention to be useful at the point of care. There have been numerous efforts across both the public and private sectors to define common healthcare terms that achieve semantic interoperability, but implementation remains disjointed and poorly coordinated. While the goal is to retain the semantic meaning across systems, essentially none of these efforts have achieved this goal at scale. The registry use case explicitly requires data from end to end – the capture of high-quality, well-defined, clinically relevant discrete data on the front end of clinical care through the analyses conducted by registries in support of performance measurement and quality assessment. The terminologies thereof are founded in the clinical realm – consistent with clinical perspectives to be used for clinical purposes. With registries at the forefront, the impetus is already present for end to end semantic interoperability.

- **Registries Facilitate Development of Real-World Evidence**

A key differentiator of the registry paradigm compared with clinical trials is the opportunity to capture real-world data across the entirety of an eligible population. To actualize the goal of a learning health system, highly accurate, on-demand data are needed to answer questions. Envisioned are not just registries, but registry networks that share interoperable data electronically. (17) Where prospective, empirical evidence is lacking, registry data could, if built upon interoperable data sourced from core common data elements, create a suitable, curated, and validated data substrate to support missing findings via observational research or pragmatic trials. Registries are increasingly providing the data to perform both observational and prospective research studies in ways that substantively reduce the time from hypothesis generation to findings and publication. (18, 19) Higher quality, more comprehensive data across the clinical enterprise can further accelerate opportunities for understanding disease processes and management while reducing data acquisition costs.

- **Registries Enable Healthcare Surveillance**

Another potential enabled by data standards derived from the registry domain is the aggregation of real-world data for purposes of surveillance. A limited number of registries already capture longitudinal data useful for drug or device surveillance. Registries could be used alone or combined with other data sources to examine the relationship between adverse events and medical device usage or performance, individual clinician or provider organization performance, medication use or social determinants of health. These analyses could be automated to create active, prospective real-time signals for key stakeholders (e.g., public health department, drug or device manufacturer, regulatory bodies). (20) Current approaches to this kind of tracking are rarely standards-based but are instead retrospective and require manual processes.

Significant time and cost savings could be realized using an automated surveillance methodology. In one evaluation of the return on investment of a registry-based approach to transcatheter aortic valve replacement device surveillance, the registry approach was estimated to save device manufacturers $75 million compared with a formal post-approval study approach, a four-fold reduction in surveillance cost (21). This savings was achieved even with the use of manual chart abstraction to capture and submit registry data. By reducing manual chart abstraction, the potential cost savings of interoperable data flowing from point of care directly to registries across the entirety of the US registries portfolio could easily be in the billions of dollars.

- **Registries Are a Key to Performance Measurement and Population Health Management**

For medical and health care professional specialties seeking to evaluate detailed measures of quality and performance relevant to member clinicians and provider organizations, registries have long provided one of the only practical methods for acquiring the needed data at a national level. Entities including the Centers for Medicare and Medicaid Services (CMS), the Joint Commission, and the National Committee for Quality Assurance are committed to moving their quality reporting programs to electronic capture and reporting. By establishing and stewarding the respective terminology standards, the registry community can catalyze a "collect once, use many times" approach to standardized clinical data using a common data dictionary as their foundation. CMS acknowledges this potential via several programs; for example, the CMS Quality Payment Program provides a mechanism for registries to offer quality measurement and reporting as a service for their clinician and facility participants. (22) Attempts by registries and specialty societies to create electronic clinical quality measures have been limited by the same issues that registries are challenged by in general – a lack of consistent process, structure and standardization of registry data models and workflows resulting in data quality issues. (23) A solution addressing these larger data standardization and quality issues is expected to ease burdens associated with data capture for meaningful quality measurement and improvement.

Registries also have an important role in population health management as they facilitate high-risk patient cohort identification and risk stratification. Accountable care organizations (ACOs), created to distribute financial risk across multi-disciplinary care teams, utilize registries for population and care management, patient monitoring and quality metrics, controlling costs and improving efficiency. ACOs often focus on a population that receives care from a specific set of providers, an important point of contrast for registries with additional or other primary business models. This registry use benefits strongly from the use of established quality and performance standards definitions, leading again to the previously described challenges faced by organizations seeking to measure and aggregate quality data across specialty and organization boundaries.

- **Registries Facilitate Clinical Process Improvement**

Many do not understand why the efficiencies IT brings to other industries have not yet been realized in healthcare. One reason is the lack of consistency in workflows across clinical care. Other industries that have successfully improved quality, safety and efficiency have embraced standardization of work processes to drive the tooling thereof. In health care, workflow has often been designed to conform to tooling, rather than the opposite approach. (24) The construction of use cases, workflows, and dataflows around the registry-enabled care processes has the potential to streamline care activities and improve billing and reimbursement processes. Improvements in reimbursement are frequently noted with the implementation of standardized workflows for the electronic coding of implantable medical devices. (24) In another example, data entered at the point of care by the appropriate clinicians, if shared, can eliminate hours of time spent by downstream clinicians and staff requesting records, reading charts and asking patients for information that should already be available.

METHODS

The Project had 3 specific Aims as listed above and outlined below. Those Aims were designed to accomplish 2 key goals:

- Determine the scope of the problem: identify clinical concepts commonly captured across a selection of registries and determine the degree of concordance of their database representations across the registries and with national data standards
- Create a solution: author a guide for the programming of the common core clinical concepts in databases such that native data interoperability could be inherently built into databases

To accomplish the Aims of the project, a team of informatics experts at the DCRI (Anne Heath, Mary Williams, Davera Gabriel, Asba Tasneem, Rebecca Wilgus, James Tcheng) executed the following work.

Aim 1: To compare the case report forms (CRFs) and other registry database artifacts of at least 20 different disease and device registries of professional societies, to identify the clinical data elements common (>50% prevalence) across those registries.

To accomplish Aim 1, we solicited the contribution of CRFs, data capture forms, data dictionaries, and electronic database representations of registries that were members of the two national organizations of registry owners: the PCPI National Quality Registry Network (NQRN) and the Council of Medical Specialty Societies (CMSS). Recruitment was achieved through a combination of face to face meetings, webinars, direct recruitment, personal communications, and word of mouth. A total of 38 registry owners contributed content (Table 1).

All registry materials were kept confidential at the DCRI and not further distributed. Since several registry owners managed multiple registries, a limit of 4 registries from any one given registry owner was implemented. The non-domain specific data elements captured by each registry were abstracted and anonymized. For efficiency, only clinical data elements that could be reasonably expected to appear across multiple registries, such as such as patient identifiers, comorbidities, common physical examination findings, procedures, medications, laboratory results, and patient outcomes were tabulated.

We intended to evaluate at least 20 common clinical data elements in this context across the registries. Specifically excluded from the tabulation were administrative data (e.g., social security number, address, insurance information) as well as disease, device, and / or procedure-specific concepts that would be nominally expected to be the unique focus of a specific registry. The resulting data elements were analyzed first by grouping them into logical domains. Within a domain, each element was assigned an informal match status (identical, nearly identical, similar, no match) with data elements found across the other registries.

Details about the process of abstracting data elements from registry artifacts are as follows. We parsed applicable content of each registry resource by logical domain and then recorded in a spreadsheet the clinical concept (e.g., Date of Birth, Race, Ethnicity), data element name (i.e., question text on the CRF), data type, and permissible values as represented in the registry artifact. Each registry resource was abstracted several times across each logical domain and discrete data element within domains to assure completeness and eliminate abstraction errors. The columns of the spreadsheet included Source ID (blinded random number), Domain, Clinical Concept Label, CRF Label, Clinical Concept Definition, Data Type / Format, Allowed Values, and Allowed Value Definitions. In the abstraction, we did not encounter any resources that referenced OIDs or SDO reference bindings. Domain and Clinical Concept Label were later merged into a single concept field called Clinical Concept and CRF Label was later renamed Data Element Name. While Clinical Concept Definition, Data Type / Format, and Allowed Value Definition are all critical metadata, these components were not used in the assessment of concordance.

Our original intent was to identify a set of core data elements defined as being present in more than 50% of the registry artifacts, standardizing the representation of the core data elements through the implementation guide. However, across the 38 registries, there were no clinical concepts represented in 50% or more of the provided registry artifacts. Furthermore, an exceedingly few were collected data via precisely the same technical representation across registry sources. With these initial findings, we shifted our efforts to quantifying the degree of discordance. The approach we took was to evaluate in detail a pre-determined set of commonly collected clinical concepts. With the input of clinicians and ontologists, we decided to concentrate our analysis on the following concepts: Patient Name, Date of Birth, Sex, Ethnicity, Smoking Status, Body Height, Body Weight, Heart Rate, Blood Pressure, Procedures, UDI, Death, Substance Use, Alcohol Use, Care Team Members, and Labs.

This process identified clinical concepts shared across the registries in 2 classes: first, data where a single prompt : value pair captures the information (e.g., date of birth, sex, vital signs, procedures, UDI); and second, data where a set of prompt : value pairs is required to capture the information (specifically, medication data and laboratory data). Note that metadata to describe context of any given data element (e.g., past medical history, current encounter, procedural, post-procedure) is typically defined by the structure of the data collection instrument itself, and so a separate metadata concept to capture context is NOT included in this framework.

Focusing our attention on these clinical concepts allowed us to assess concordance at a more granular level. We defined concordance as an exact match of data element name and permissible values, these being the essential components needed for interoperability. Any disagreement on either of those planes was counted as discordance. We did not consider implementation decisions in determining concordance.

Concordance was documented in tables generated by filtering the spreadsheet of raw abstraction data. We filtered the spreadsheet on the Clinical Concept column for each concept under focused review as described earlier. For each filtered concept, we placed those results in a separate spreadsheet and manually collapsed and counted the data elements that were concordant on data element name and allowed values as was earlier described in the definition of concordance. Those were then formatted and placed in a Word document and titled Registry Source Concordance Tables.

Table 1. Organizations Contributing Registry Content

- American Academy of Ophthalmology
- American College of Cardiology NCDR
- American College of Gastroenterology
- American College of Obstetricians and Gynecologists
- American College of Radiology
- American Optometric Association
- American Orthopedic Association
- American Physical Therapy Association
- American Podiatric Medical Association
- American Society for Clinical Pathology
- American Society for Radiation Oncology
- American Society of Anesthesiologists
- American Society of Echocardiography
- American Society of Nuclear Cardiology
- American Urogynecologic Society
- American Urological Association
- Americas Hernia Society
- Arthritis Research Center Foundation
- Creaky Joints Patient Powered Research Network
- Cystic Fibrosis Foundation
- Michigan Surgical Quality Collaborative
- National Osteoporosis Foundation
- Neuropoint
- North American Association of Central Cancer Registries
- Outpatient Endovascular and Interventional Society
- Plastic Surgery Registries Network (GRAFTS, TOPS)
- Renal Physicians Association
- Society for Vascular Surgery
- Society of Interventional Radiology
- Society of Thoracic Surgeons
- University of Massachusetts Function & Outcomes Research for Comparative Effectiveness in Total Joint Replacement
- United Network for Organ Sharing
- Venous Access: National Guideline and Registry Development (VANGUARD)
- Vermont Oxford Network
- Women's Health Initiative

Aim 2: To evaluate the identified core clinical concept data elements in the context of the "Big 5" healthcare data standards, the USCDI, and national data models (OMOP/OHDSI, Sentinel, PCORnet), determining the matches and suggesting the most appropriate match between registry data elements and existing data standards.

The goal of Aim 2 was to evaluate the degree of utilization of data standards by registries for the core clinical concepts identified via the Aim 1 processes. The premise was that while there might be modest compliance with data standards, there would be insufficient rigor to accomplish native data interoperability at the physical database layer per se. The task was to compare the representations of data

elements in national data models and national data standards to the registry core clinical concepts. Specifically, for data elements that had at least a "similar" concordance status, a master concept would be identified drawn from reference data standards, and then the match status of data elements to master concept was assigned. The intent was to produce a crosswalk where the master concept was identified, including a limited amount of meta-data about the concept (e.g., concept name, concept definition, allowed values, and allowed value definitions), and the SDO representations of the concept. Descriptive annotations of the master concepts were culled from the "Big 5" SDO initiatives and the USCDI. The registry concepts of interest were compared to each of these models and terminologies to determine the degree of concordance with data standards.

Unexpectedly, in the work of Aim 2 we determined that national data models also did not completely conform to established data standards. We therefore pivoted the project to evaluate the ways those national data models represented the core concepts identified via the registry survey process. Each representation of a core concept was documented by source (national data standard or data model) in a Microsoft Excel (2016) spreadsheet and a concordance table was created. Concordance was defined as the degree of agreement between sources on data element name and each item in the permissible value set; 100% concordance was defined as having complete agreement between sources in terms of data element name and each item in the permissible value set. If there was a difference in data element name or permissible value set, then the two sources were deemed not concordant.

Aim 3. To describe the current state of the clinical data standards environment, identifying the limitations of that framework, for the purpose of producing a roadmap that catalyzes the governance, structural, operational, and technical transformations needed to implement a common clinical data element set across EHI and registry systems, and national data models.

A key premise of the Project is that database developers (e.g., of registry databases) are foundational to accomplishing interoperability, and that increasing data liquidity requires reducing barriers to data transfer at the level of the data itself. To this end, the database developer community has been hampered by the lack of authoritative source documentation for the building of data fields in databases. While knowledge of standards work cannot be ignored, standards themselves are somewhat peripheral to the construction of the data fields per se. The work of Aims 1 and 2 identified deficiencies and limitations of the current state of interoperability with respect to common core clinical concepts and their representation in registry databases systems. The intent of Aim 3 was to memorialize the work of Aims 1 and 2 in the form of the key deliverable of this project, namely this report inclusive of a programmer's implementation guidance to building the common core clinical concepts as database elements in registry databases.

Single Prompt:Value Concepts

The creation of the programmer's implementation guide explicitly evaluated representation of the core clinical concepts of interest across the participating registries and national common data models, resulting in a harmonized set of parameters for each of the concepts. A Microsoft Excel (2016) spreadsheet containing the recommendations regarding the concepts was generated. These concepts were described using the following 13 key metadata elements:

- Clinical concept label (human prompt – for CRF, data entry screen)
- Clinical definition
- Clinical allowed values (human prompt – for CRF, data entry screen)
- Clinical allowed values definitions
- Database field label (as listed in db)
- Database field data type / format (e.g., char, date, float, integer, list)

- Database field business rules (e.g., cardinality, edit checks, range checks)
- Database allowed values (as stored in db)
- Reference ontology clinical concept binding
- Reference ontology allowed values bindings
- OID
- Fast Healthcare Interoperability Resources (FHIR) profile and / or resource
- Sources / references / notes

In terms of selection of recommended metadata, priority was given to predicate work, particularly the ONC USCDI, the NIH National Library of Medicine (NLM) Value Set Authority Center (VSAC), and HL7 FHIR profiles and resources (especially content listed in the FHIR Detailed Description tabs and FHIR Implementation Guides). Of note, the metadata elements that were least consistent were the clinical definitions and clinical allowed values definitions. Where there was no single, unifying definition, a definition compilation was created focusing on clarity and ease of understandability. For other metadata, where no reference representation was identified, the recommendation was made based on the Aim 2 concordance tables where a majority of the national common data models agreed on data element representation; in these cases, metadata fields were populated using information from the national common data models. Reference sources are noted in the FHIR Profile and Resource and Sources/References/Notes metadata elements of the programmer's implementation guidance document.

Multiple Prompt:Value Concept Sets (Medications and Laboratory Data)

The identification, description, and characterization of data elements to capture medication data and laboratory data requires the capture of multiple data elements to characterize each observation. For medications, 78% of registries captured data about medications. Across the registries, the prompts were found to fall into four general contexts (complete list of medications, use of a domain-specific medication, list of medications by class, and medications administered in the context of a procedure). Typical prompts were as follows:

1. Complete list of medications (n=12) – e.g., "List all medications patient is currently taking"
2. Domain-specific medications (n=12) – e.g., "Does the patient have a TOBI Podhaler?"
3. Medications by class (n=18) – e.g., "Is the patient receiving an anticoagulant?"
4. Medications administered related to a procedure (n=8) – e.g., "Indicate if patient was prescribed Ciprofloxacin (Cipro) antibiotic after biopsy"

Once these general categories were discerned, we compared the specific prompts from the medication representations used by each registry with the AHRQ Health Information Technology Standards Panel (HITSP) medication data model. Medication prompts that could not be categorized per the HITSP model were noted and contributed to the medication model of the project. A survey of existing standards noted multiple potential standards contributors to medication modeling, including HL7, ICD, NCI, LOINC, SNOMED CT, RxNorm (including RxClass), NDF-RT, NDC, and UNII. To arrive at the final medication model of this procedure, we solicited comments from across the informatics experts and registry stakeholders contributing to the project. The final model was designed to capture the data sufficient to respond to the 4 medication contexts listed above.

Similar to medications, registries frequently captured laboratory data. Across our survey, 100% of registries prompted for laboratory data. Unlike medications, the laboratory data concept is managed in databases as a series of individual observations – with the need for multiple prompts related to timing of measurement (i.e., context) and units of measure. As with the other data elements, the evaluation of laboratory data included an assessment of the conformance of laboratory data elements with existing data standards. A survey of existing standards noted multiple potential standards contributors to laboratory data modeling, specifically LOINC, SNOMED-CT, and UCUM. We leveraged the proceedings of two

multi-agency sponsored public workshops (25, 26) and a multi-agency/ stakeholder public-private collaborative (Systemic Harmonization and Interoperability Enhancement for Laboratory Data (SHIELD), of government, industry, EHR vendors, laboratories, the College of American Pathologists, standards developers, professional organizations, MDIC/NEST and academia) (27) to arrive at the laboratory data model.

Proceedings and Validation

To provide external validation and drive the Aims of the project, a series of face to face meetings and webinars with stakeholders, along with frequent team meetings were held. Key stakeholder engagements included the following:

- March 18 & 19, 2018: PCPI/NQRN Registries on FHIR meeting in Washington DC. Initial introduction of the project to NQRN members and issuing of an open call to registry owners to participate by submitting registry case report forms, data dictionaries, and other registry artifacts.
- May 10, 2018: CMSS Registry Summit in Chicago, IL. Introduction of the project to CMSS members and issuing of an open call to CMSS members to participate by submitting registry case report forms, data dictionaries, and other registry artifacts.
- May 11, 2018: Posting of the project summary on the Interoperability Proving Ground.
- June 27, 2018: Go-live date of website describing the project (https://dcri.org/registry-data-standards/)
- July 12, 2018: Project webinar with stakeholders to further describe the project, provide an update on progress, and solicit feedback. Presentation of preliminary findings regarding (lack of) concordance across registries with clinical core concepts of interest, and (lack of) utilization of data standards in the capture and representation of those concepts.
- July 29 – August 1, 2018: Health Services Platform Consortium-Clinical Information Interoperability Council (HSPC-CIIC) meeting at NIH Lister Hill, Bethesda, MD. Presentation of the project to the annual meeting of the HSPC-CIIC, including planning for the follow-on phase of formal terminology modeling of the project data element work product.
- August 21, 2018: In-person stakeholder meeting held in Washington DC at the Pew Charitable Trusts, with presentation of additional findings regarding (lack of) concordance across registries with clinical concepts of interest, finalization of the list of the concepts of interest, and engagement of discussion groups focusing on planning work to incorporate the programmer's implementation guidance into registries and other EHI systems.
- September 24, 2018: 2nd project webinar with stakeholders to review final clinical core concepts of interest, along with associated metadata, via a walk-through of each of the data elements, followed by request for feedback (via Qualtrics survey tool).
- September 27, 2018: Spreadsheet of registry concepts of interest complete with metadata distributed to 306 stakeholders with survey tool.
- October 15, 2018: Stakeholder comment period closed.

Feedback was provided by external stakeholders through 3 primary mechanisms: direct comment (e.g., email correspondence, phone discussions, face to face discussions), in group breakout sessions during the August 21, 2018 face to face meeting held at the Pew Charitable Trusts facility, and via a Qualtrics (Qualtrics, Inc.) survey issued September 27, 2018 with the first production draft of the implementation

guidance (the "spreadsheet). Where appropriate, direct comments have been incorporated into the work product. Proceedings of the group breakout sessions were processed using mixed-methods research software (NVivo, QSR International) and summarized below and incorporated into the Project where possible; recordings and transcripts of the group breakout sessions are posted to the Project website (https://dcri.org/registry-data-standards/). Responses to the Qualtrics survey were individually evaluated for inclusion where appropriate into the Project work project; the survey results and Project team responses are also posted to the Project website (https://dcri.org/registry-data-standards/).

FINDINGS

A. Environmental Scan

An extensive environmental scan was performed of the predicate work in the health IT standards and interoperability space felt to be of relevance to the registry community. Key findings and interpretations are as follows.

- **Health IT Standards for Data Interchange**

As alluded to earlier, many of the components needed for semantic interoperability exist; what was clearly identified was the lack of single, explicit and communal operational approaches to the "how", especially from the registry database programming perspective. This is needed to allow quality data to be not only be clinically captured and shared, but also queried, analyzed, and aggregated across computer systems. While the EHR Incentive Program ("Meaningful Use") required certain interchange standards to be implemented in certified EHR systems, those standards did not focus on the physical representation of granular clinical data per se. The Meaningful Use framework was based on the premise that the key contribution of interoperability was the interchange of largely narrative documents for team-based care. The specifics for coding discrete clinical data within documents were left largely to the implementer, with no guidance as to the storage of granular clinical data at the database level. As a result, the use of proprietary terminologies, disparate form and format requirements, optional metadata, and free-form concept coding has flourished. Within current document-based standards, there is nearly unlimited freedom to code the granular clinical content per the discretion of the programmer / implementer. For example, in the Health Level Seven International (HL7) Consolidated Clinical Document Architecture (C-CDA) standard, it is specified that lab tests should be coded using LOINC. However, this instruction statement alone does not specify which LOINC codes to use, what units of measure to apply, whether and how reference ranges are to be included, and when to allow the use of free text. As a result, while C-CDA documents can contain discrete, coded elements, systems receiving those C-CDA documents frequently cannot consume the coded elements without additional translation. In addition, narrative text occurring in the midst of coded data is often simply discarded. As a result, lab test data are not easily shared, with most EHR systems essentially ignoring much of the content of a C-CDA while still qualifying for certification per the EHR Incentive Program for having the capability to send and receive C-CDA documents.

Even with an interoperability construct focused on document exchange, ONC has recognized the need for greater specificity of clinical data. In the 2014 and 2015 editions of the EHR Incentive Program, the Common Clinical Data Set (CCDS) specified certification standards for nearly 2 dozen clinical concepts including race, sex, medications, laboratory data, and vital signs. (28) The 21st Century Cures Act set a goal of specifying a common set of data classes that are required for interoperable data exchange. (4) In furtherance of this goal, the ONC created the USCDI, an expansion of the CCDS that delineated processes for adding to the core set of standardized data elements. (29) However, the USCDI specifies clinical content at the data class level, not at the level of fully-specified data elements. Given this, it is unclear whether content conforming with the USCDI will reach the needed level of specificity required to accomplish native data interoperability. The work of the Project intentionally targets the specification of metadata to the level of granularity necessary to incorporate (some of) the data elements in the USCDI in health IT systems.

HL7 FHIR is an interoperability approach that structures the interchange of healthcare data. The FHIR standard includes specifications of data structures for common concepts within the FHIR resources as part of the standard itself. This base set of healthcare concepts is configured for interoperability via FHIR profiles, which are groups of data structure definitions (e.g., constraints, extensions) and corresponding

value sets defined for a common purpose. FHIR profiles extend and constrain applicable resources to make them suitable for specific use cases and workflows. (30, 31) These FHIR profiles model the information, with the specifications published in implementation guides for software developers who wish to implement FHIR-based information interchange in their products. FHIR profiles covering some of the content of the USCDI have been published, including the US Core Implementation Guide. (32) However, the details for the specific types for capture and storage of the data of observations and procedures again are not defined in the base standard, nor does the US Core Implementation Guide delineate all of the clinical content to the same level of specificity.

In summary, while current health IT standards provide a level of format and structure for the interchange of data between systems, standardization of the content itself that preserves semantic meaning and is natively computable is still lacking. Specific to registries, attempts to extract information from documentation systems via FHIR and other translation-based approaches have simply not succeeded to date.

- **Common Clinical Data Element Libraries**

A feature of most registries is that the data captured across the varied and differing healthcare organizations (and the clinicians thereof) focuses on clinical care, healthcare processes, and / or patient outcomes. For this to be effective, a key prerequisite is a common, shared data dictionary that defines the concepts being captured. This gives registries a unique perspective on the somewhat conflicting priorities of data specificity versus parsimony. Unfortunately, most of the common data element and clinical information modeling efforts to date have not been primarily driven by the clinical care use case but instead by research, public health or surveillance requirements. In addition, registry owners have largely avoided informatics formalisms required to develop registry data dictionary content into common clinical data elements. This has hampered implementation in EHR systems and reduced use of well-defined clinical concepts in the clinical care setting. The problem also extends to federal agencies in which agencies or programs often develop their own definitions of common data elements. The overall lack of transparency and governance in the content use space has led to a proliferation of repositories and data sets that are strongly overlapping in content without being interoperable. In some cases, duplicate content is common even within a single repository or dataset. A review of electronic databases for adverse event reporting systems found 1,782 items that corresponded to 33 different data element groupings, the vast majority of which were unique and dedicated to the specific reporting system and inherently not interoperable. (33)

Without leadership and collaboration across the clinical community, government, regulatory bodies, and health IT stakeholders, neither a standard for common clinical data elements nor a process for governing same has emerged. In the remainder of this section we review some of the existing resources and content repositories that could be aligned across the landscape of clinical data elements. However, much of this content has never have been used in or mapped to or from operational clinical systems. Table 2 shows a list of non-registry-based repositories and resources where CDEs, models and terminology content can be found. The work of Ghitza and colleagues to utilize CDEs created for clinical research in a point of care setting is illustrative; significant modifications to the content was required before the CDEs could be implemented. (34) Harmonization of content is clearly needed and has been suggested many times; however, many of the owners of these sets of data elements were developed for a single, specific use case, and incentives or other motivations to invest in CDE modeling and harmonization work have proven insufficient or nonexistent. While many of these repositories contain useful content or functionality, the lack of cohesion in the standards used, the content covered, the use cases served, and the varying levels of content curation means that there is no true gold standard or single location to find content that meets clinical point of care or downstream clinical analytics use cases.

Table 2. CDE, model and terminology repositories for candidate clinical data element content.

Repository	Description and Use Case	Content and Standards
Draft US Core Dataset for Interoperability (USCDI)/ EHR Incentive Program (Meaningful Use) Common Clinical Dataset (CCDS) www.healthit.gov/sites/default/files/draft-uscdi.pdf	Draft regulatory set published by ONC. Goal is to define or constrain content that would facilitate semantic interoperability across EHRs. Use of CDS was required for ONC certification of EHR systems. It appears unlikely that the same regulatory approach will be pursued with the USCDI.	Although a limited number of discrete data elements are included, USCDI does not use a defined standard for common data elements per se. Focuses on "data classes" across clinical care and administrative data. Uses EHR Incentive Program endorsed code systems. The USCDI currently published in pdf form, CCDS is available in VSAC and in FHIR US Core for 2015 Edition.
NLM Common Data Element Repository https://cde.nlm.nih.gov/	Library of CDEs made available through the NLM site. Initial goal is to collect and provide a single point of availability of federal CDEs for federal programs. Could in the future provide tooling, authoring, endorsement or support for USCDI but this is currently aspirational.	Standards depend on the form, manner and terminology associated with each dataset. UMLS login allows full viewing of licensed content. API interface in development.
CIMI Model Repository http://cimi.hl7.org/	The Clinical Information Modeling Initiative (CIMI) repository allows clinical and technical users to create highly specified models that describe clinical concepts. These models should have the level of specificity to prevent interoperability failures related to alternative interpretation of the content.	CIMI provides a logical model framework for clinical content. This model can be used as a physical representation for content and generally would provide a stable representation allowing transformation to another exchange standard.
PCORNet https://pcornet.org/pcornet-common-data-model/	Specific to the patient-centered outcomes research use case, contains most typical classes of clinical and administrative data. It	Based on PCOR-Net Common Data Model which is specific to the patient-centered outcomes research use case. Recodes concepts

	aligns to but requires mapping to other federal and non-federal data models and data representations.	that already exist in terminologies in many cases. Available on the PCOR site as a download.
FHIR Core and US Core http://www.hl7.org/fhir/us/core/	FHIR is not a content standard– thus, there is no "FHIR CDE Repository". Most FHIR resources are described at the class level. There is requisite support of a limited number of resources focused on patient administrative data.	FHIR provides a physical representation for content exchange – it does not have an underlying logical model. FHIR Core includes standardized, required supported content although substantial optionality exists in the inclusion of metadata. FHIR US Core aligns to the MU CCDS/USCDI (primarily at the class level). Content available through FHIR API.
CMS Data Element Library https://del.cms.gov/DELWeb/pubHome	The CMS Data Element Library supports federal long-term and post-acute care (LTPAC) programs that monitor quality of care. Data elements are specific and limited to that use case.	Uses standard terminologies (SNOMED and LOINC) using a standard format for questions and answers from standardized assessments. Includes MDS 3.0, RDF-PAI, OASIS, LCDS, and HIS.
CMS Data Element Catalog https://vsac.nlm.nih.gov/download/ecqm	Housed within the NLM Value Set Authority Center, this repository is updated at least annually and contains all the data elements for the CMS eCQM programs. Certification is offered by ONC for use of the content to report eCQMs.	Uses the Quality Data Model (QDM) to describe datatypes and attributes connected with value sets and codes from standard terminologies as described by HHS in the CMS Blueprint for use in quality measures. Available as a download or through NLMs SVS or FHIR APIs.
AHRQ US Health Information Knowledgebase (USHIK) https://ushik.ahrq.gov/mdr/portals	The United States Health Information Knowledgebase (USHIK) is an on-line, publicly accessible registry and repository of healthcare-related metadata, specifications, and standards. It encompasses	Provides support for NCPDP, X12, and HITSP standards as well as the content of AHRQ Common Formats, Children's EHR Format, CMS eCQMs including draft content and Claims databases for some states. Content standards are determined by the different

	many use cases in a single location.	formats for the content available on the site. Downloads available in multiple formats including an API.
Value Set Authority Center https://vsac.nlm.nih.gov/welcome	Hosted by NLM, this site allows users to author, view, download, and collaborate on value sets using terminologies in the UMLS. Includes content for CMS eCQMs, HL7 C-CDA, and CMS hybrid measures.	Terminologies include those supported by NLM and those supported by other terminology owners. A UMLS license is needed to view and author licensed content. Content is limited to value sets unless codes are identified within a specific use case supported by VSAC. Available through SVS or FHIR APIs.
National Cancer Institute Cancer Data Standards Registry and Repository CDE Browser cdebrowser.nci.nih.gov/	Hosted by NCI, this site collects NIH Institute data element sets for clinical research efforts. Includes content across all the Institute clinical domains.	The form and manner is based on ISO 11179-3; however, the conformance of the individual CDEs isn't tested. There is no governance process so there are many duplicate and conflicting CDEs across the repository.

Largely missing from the matrix in Table 2 is the registry community. Given the emphasis of registries on the capture of granular clinical data for assessment of clinical quality, healthcare processes and performance, and patient outcomes, the registry community (and the professional society registry owners thereof) should not just be represented but arguably should govern and steward the common data element ecosystem. In addition, responsibility for harmonization arguably falls squarely on the registry community itself as subject matter owners. Too few registries have communicated, collaborated or shared their content and attempted to align with nationally-accepted standards (and vice versa). Despite this, the prospect of coordinated registry networks (35) is creating an appetite for cross-registry CDEs derived from registry data dictionaries. Harmonization is an extremely difficult undertaking without a shared understanding of the importance of multiple use cases, but registries, having a unique understanding of data capture, extraction, and reuse in the clinical domain, are best positioned to accomplish just that. Even with shared governance, coordination, and stewardship, the final missing link is technical infrastructure – tooling that minimizes the need for technical expertise and provides content transparency, an inclusive governance and stewardship structure for content development and maintenance, a representative stakeholder community, and resources to accomplish the above. That community must agree on the need for shared definitions, have a willingness to collaborate to harmonize terminologies, and then offer to participate in the effort to create a meta-dictionary of CDEs. Registries are well equipped to provide that leadership along with the expertise to bring their respective data dictionaries into a single repository. This goal – a single, shared interoperability framework – could revolutionize both the sharing of data and catalyze its seamless reuse for safety, quality, research and beyond.

- **Implementation Maturity**

Data standards and the database systems that use them need ongoing stewardship and maintenance. Developing a data standards management strategy that includes a process to maintain semantic alignment over the lifespan of a registry is key in assuring that the registry can support the depth and breadth of potential uses. (36) The most efficient approach is to utilize a human-understandable, detailed logical model to represent the clinical content which can then be translated into a physical representation. This would enable, for example, the concept of peripheral arterial disease and its component metadata to remain stable as machine languages and syntaxes mature and change over time. Aligned common data models provide the basis for a common representation that crosses versions, format and structure. These highly curated and stable data element representations can then be extended to support multiple clinical, research, quality and performance measurement objectives.

A major challenge is that most EHRs are unable to capture registry content as native data, thus requiring abstraction and ETL for transmission to registries. A second challenge is the lack of consistency across registries for loading data electronically into the respective registry databases. There is a clear need to mature the common data element ecosystem to include mechanisms and processes to review, promote, and steward content created by registries to assure at least a basic level of data quality, to help deploy and mature the corresponding data concepts, and to scale across health IT systems to capture data at the point of care integrated into the processes of care. The creation and shared use of a common registry set could create the critical mass needed to allow the registry and clinical community to advocate for alignment of federal and other data element sets, and for EHR and other health IT vendors to support this content.

- **Data Heterogeneity and Data Quality**

Data heterogeneity across registries is pervasive and contributes to inconsistent data quality, chart abstraction requirements, and IT technical work to accomplish data translation, mapping, and transmission to registries. Variation across and within systems affects the overall quality of data and consistency of the data over time. Lacking a common core set of data elements, users of registry data are often limited in the analyses they can perform due to the compounding issues of data heterogeneity. (37) If data heterogeneity were reduced, opportunities would increase to combine and aggregate registry data for research and surveillance purposes and support the coordinated registry network approach. A nuance to recognize is that while most concepts can be captured as explicitly structured data (and could thus be interoperable at the native data structure level), a limited number require synthesis of information to derive.

Data quality is a consummate problem in health IT, in large part due to the lack of consistent human-understandable definitions, machine-interpretable representations, immaturity of and rapid changes in standards and terminologies, widespread reliance upon translation and mapping, and failure to include adequate metadata to establish data provenance. Critically, clinical registries inherently enjoy a much higher level of data quality than that available to the average data owner or aggregator; however, this is at the cost of employing cadres of human chart abstractors, implementing numerous quality assurance checks, and the manual entry of missing data. If registries are to maintain their current level of data quality while remaining sustainable, the paradigm of structured reporting - data capture integrated into workflow based on standardized data dictionaries – is paramount. (8) These strategies also alleviate clinician and staff burden without compromising data quality. More, not less information should be coded with elements including data about their use. For example, should an exclusion criterion apply to a group of patients for reporting an otherwise required data element, registries should include in the data definition that criterion.

Data quality assurance is rapidly becoming a requirement of participation in registry programs and organizations. One of the PCOR standards for registries is “to take appropriate steps to ensure data quality” including “range and consistency checks”. (9) Registries also “need to be able to demonstrate the high quality of their data through transparency of measures and quality control techniques”. (38) As

patient-generated data are incorporated into registries and their related clinical use cases, uniform data collection and quality checks and validation measures will be paramount to draw accurate conclusions. (38) Developing data quality plans early in the registry and study planning process, including detailed data auditing and verification plans, are crucial to ensure a robust start to such activities; however, these activities must also be maintained longitudinally. (13)

- **Data Linkage, Consent, and Patient Matching**

Aggregating registry and EHR data in analysis datasets can reduce data collection costs and provide much more detailed information about patients than is available in a single registry. When registries and sites utilize multiple sources of data for reporting, linking patient records appropriately often remains a roadblock despite years of research and policy proposals on the subject. (11) The PCOR standard recommends developing data linkage plans early in the study process inclusive of data descriptions, linking processes, and transformations necessary, and matching these factors to the appropriateness and limitations of the linked data for use with a plan to maintain and test the processes over time. (11) The use of common core data elements for patient matching should support data linkage directly, whereas data elements based on common definitions support analyses which can more easily address data missingness, linkage, reproducibility and quality assertions that could impact research quality and interpretation. Many useful studies have been undertaken to suggest methodologies for patient matching across data sources; however, most fail to reach the level where CDE content is explicitly declared. A shared set of patient matching data elements across registries could provide the first steps towards harmonizing content across registry definitions and also making registry data more easily extractable and aggregated for shared use.

Data sharing also introduces concerns about privacy and security, patient consent, and business use agreements. Local, national and international regulations must also be considered and evaluated when linking data from secondary sources or repurposing data for a study, including applicable legal and privacy conditions in both its original and repurposed data use. It is hoped that the participation in and development of the Trusted Exchange Framework and Common Agreement (TEFCA) will further address privacy and related issues while reducing barriers to sharing. (39) Registry owners should maintain awareness of and participation in the TEFCA as it is developed to help advance TEFCA as a tool that reduces administrative and legal hurdles to data exchange.

- **Missing Data**

Given that healthcare documentation is typically incomplete, it can be anticipated that registries will have an element of data missingness. Because the process of delivering care is uniquely tuned to the individual patient, it can be anticipated that missingness issues may actually worsen as registries move towards point of care capture at scale. Regardless, many missing data methods used in research are unsustainable in the context of the larger healthcare enterprise.

Understanding the types of missing data assists in developing appropriate analytical strategies when using registry data. Fundamental to this approach is understanding the needs of the data entry users at the point of care – only users who prioritize high quality data are likely to capture same. A critical eye must be used when determining whether data are considered required vs. optional. (40) Research on the topic of optionality within a model or data element can be useful to optimizing data utility while maximizing data entry efficiency and should be part of a data review strategy as well as incorporated into training for users. Fortunately, the concept of data persistence is likely to be incorporated into the EHR of the future to reduce unnecessary repeat documentation.

B. Core Clinical Concepts

The work in support of Aim 1 of the Project was to identify shared, core clinical concepts and compare their representations per the supplied registry artifacts. We anticipated that we would identify 25-50 clinical concepts shared across registries; the final tally was actually only 18 unique concepts, 2 of which (medications and laboratory tests) are better represented as models than singular data elements. (We counted dependent concepts such as "Height" and "Height Unit of Measure" together as one concept.) Thus 16 concepts could be considered single (or collections of single) prompt:value concepts. Next, we evaluated for concordance of the corresponding data elements (defined as having exactly the same data element name, allowed values, and data type), anticipating that we would find concordance across at least 50% of registries in which they appeared. Our analyses demonstrated that none of the data elements were concordant across 50% or more of registries. In fact, very few of the data elements were concordant across more than 2 sources. Thus, abstraction and analysis of registry content revealed considerable variation (i.e., little concordance) among the concepts of interest for this project.

Highlights of the analyses follow; for the complete list of tabulations, please refer to Appendix 1.

- **Single Prompt:Value Concepts**

While we had originally intended to identify shared, core common clinical concepts as appearing in 50% or more of registries, in the end this proved too restrictive. Instead, we reset the criterion at 30% to capture additional concepts for analysis and development. The final list of single prompt:value concepts delineated in the Project are:

Given Name	Height
Family Name	Height Unit of Measure
Suffix	Body Length
Sex	Body Length Unit of Measure
Race	Weight
Ethnicity	Weight Unit of Measure
Date of Birth	Heart Rate
Tobacco Product Use	Systolic Blood Pressure
Tobacco Type	Diastolic Blood Pressure
Cigarette Consumption	Date of Death
Cigarette Start Age	Proceduralist National Provider Identifier (NPI)
Cigarette Quit Date	Clinician National Provider Identifier (NPI)
Tobacco Use Duration	Procedure Code
Illicit Drug Use	Procedure Date
Alcohol Use	UDI Device Identifier

Among the most frequently collected data elements were demographics, height and weight. Height and weight are described with other vital signs in the paragraph below. Demographics include patient name (n = 22), sex (n = 21), date of birth (n = 15), race (n =14), and ethnicity (n = 14). Given name, family name, and sex (M/F) had the highest concordance with 5 registries using the same data element name, data type, and for sex, the same value set. If the definition of concordance is expanded to allow for flavors of null and variability in data element labels, "concordance" increased substantially for these concepts. While race and ethnicity were collected in many registries, every CRF data collection instruments had some unique variation in the data element or allowed value list. Most registries used some combination of the values recommended by the Office of Management and Budget, supplemented with additional values. A few had more specificity and/or some flavor of null; another variant (which the Project

endorses) was to allow for the capture of multiple races. One registry combined race and ethnicity into a single data element.

Among the most concordant data elements were vital signs (heart rate, blood pressure, height & weight). Interestingly, heart rate, systolic blood pressure, and diastolic blood pressure were collected by only 4 registries. While these registries did not specify units of measure for heart rate/pulse or blood pressure in the materials provided for this review, we assumed units to be beats per minute (for heart rate or pulse) and millimeters of mercury (for blood pressure measurements), given the universal convention of using same. Registries consistently collected systolic and diastolic blood pressures using individual discrete data elements for each, rather than a text field to capture the expression in one field. Height and weight were collected by approximately half of registries (18/38 and 17/38 respectively). Some specified units of measure (kilograms, pounds, inches, centimeters) in the data element name while others did not specify units in the materials provided for this review. Interestingly, none of the registries captured (in the materials submitted for review) the method by which height and weight were obtained (measured vs reported) and none defined body position (standing/lying down (for infants)) during the measurement; however, the need to specify body position and method of measurement in the stakeholder feedback opportunities.

Similar to data elements related to the other concepts of interest, there was considerable variation in how registries collected data about tobacco use. Nine registries collected a data element that captured currency of smoking status (some flavor of never, current, former). One registry asked specifically about types of 'nicotine use' (smoking, chewing, e-cigarette, patch, gum) vs tobacco type (cigarettes, cigars, pipe or smokeless). Three asked specifically about use of smoked tobacco within a one-year timeframe. Interestingly, one registry combined concepts of currency (former/current), frequency (every day, some days, & never) and level of exposure (light / heavy smoker) into a single data element. This registry was among the few to include a concept ID to link allowed values to a standard terminology.

Alcohol and substance use, particularly narcotics abuse, have become topics of high public health interest. While these concepts were captured in only a small number of registries, we elected to include recommendations for representations of these concepts to accelerate development of these concepts. Of note, there was considerable lack of standardization at even a foundational level (e.g., how to quantify the amount of alcohol ingestion, what constitutes an illicit substance).

Date of death was collected on only four registries without concordance in the approach. 'Death' was included in all four data element labels (date of death x 2, year of death, patient date of death). Date format was used twice; year was collected once and the fourth data type was unspecified.

Care team members were collected by many registries (n= 18) although the data element label (e.g., advanced practice provider, provider, NPI number) and data type (numeric, string, alphanumeric) were inconsistent. The data type for several registries was 'unspecified' (n = 7). National provider index (NPI) number was the allowed value 8 times. Allowed value lists for other data elements were comprised of specific provider types (DO, MD, CNM). Specific roles of interest were also identified in several data element labels (ordering physician, discharge provider) while in others labels were used to specify rank (attending physician, fellow, resident).

As adoption of Unique Device Identifiers for implantable medical devices is in its infancy, only one of the registries collected the Unique Device Identifier (UDI). However, the importance of including UDI in registry and other health information systems to support effective surveillance and evaluation of device-related clinical outcomes cannot be understated. As such, UDI is included and strongly encouraged in these recommendations.

While data about diagnostic and interventional medical procedures was collected by many registries, these data were collected in great detail very specific to the patient population served by the registry. As such, procedure data elements (beyond the generic concept of the billing concept of a procedure) were

considered 'domain specific' and were not abstracted or included in this section of the report. Finally, medication and laboratory data elements are developed separately as these are better considered composite models (see below).

In retrospect, the lack of concordance should not have been surprising. Possible explanations include that the each of the registries was developed for a different use case – often for a highly domain-specific context (e.g., coronary angioplasty procedure, hip or knee replacement). As such, there were specific underlying goals and assumptions to which we were not privy. Of note, about half of the registry artifacts provided us were actual data models – not data collection forms – and those data models stored only the clinical and domain-specific data elements. We could not discern whether demographic or general medical history data might have been captured in portions of the registry not provided us.

Medications

HITSP identified thirteen general categories of medication information along with corresponding source standards in the HITSP medications data model. Several of the categories were not used or only infrequently used in registries, while we identified three additional elements commonly captured in registries not included in the HITSP medication model (Table 4). Derivation of concept definitions and corresponding source standards were identified from ongoing work in a project led by Principal Investigator John R. Windle MD, *Optimizing the Electronic Health Record for Cardiac Care* (AHRQ R01 34-5224-2005-001). Table 4 lists the number of registries that ask questions in the specific fields of the HITSP framework (count column) as well as the number of registries that utilize interoperability standards relevant to the specific medication concept (interoperable count column).

Table 4. Medication Model Concepts

Medication Concept	Definition	Applicable Standard	Count	Interoperable Count
Fill Status	This identifies whether the medication has been fulfilled, such as completed and aborted.	HL7 v3	2	1
Indication	The medical condition or problem intended to be addressed by the ordered medication.	ICD-9; ICD-10; or SNOMED CT	3	2
Product Form	The physical form of the product as presented to the individual.	NCI	2	1
Dose	The dose of the product prescribed by the individual	NCI; LOINC; or HL7 v3	12	2
Route	The method for the medication received by the individual	NCI; or HL7 v3	6	3
Type	This is a classification based on how the medication is marketed	SNOMED CT	0	0
Site	This is the anatomic site where the medication is administered.	(Body Site)	0	0
Brand Name	The product brand name of drugs	RxNorm	0	0
Clinical Drug Name	The product clinical name of drugs	RxNorm; or LOINC	21	9
Drug Class	The pharmacological drug class	RxClass; or NDF-RT	18	2
Packaged Product	The labeler, product, and trade package size	NDC	1	1
Ingredient Name	Drug ingredients	UNII; or RxNorm	0	0

Vehicle	Non-active ingredient(s), or substances not of therapeutic interest, in which the active ingredients are dispersed.	SNOMED CT	0	0
Prescriber*	Clinician that prescribed or administered a drug.	NPI	1	1
Effective Date*	Date of the earliest fill date.	HL7 v3	10	10
Duration*	Amount of time patient take prescription	HL7 v3	5	1

The analysis of registry content determined that only a limited number of registries utilized interoperability standards (primarily for Effective Date, Indication, and Packaged Product fields), while most medications data concepts did not reference standards. Notably, out of the 18 registries that requested information by medication class, only 2 documented the use of defined interoperable standards in their data dictionary. Overall, of the potential 85 fields related to medications across all registries, only 33 (37.5%) adhered to or referenced interoperability standards.

Laboratory Data

We evaluated the use of in-vitro diagnostic (IVD) data interoperability standards in the registry artifacts available to the project and observed no use of those standards. In part, we suspect that the added layer of structure required to directly transmit IVD data from the data source (e.g., analysis equipment, laboratory information system (LIS)) is not designed to include clinical context (e.g., baseline, before a procedure, following a procedure) as captured in registry database systems. Furthermore, while laboratory data is almost completely digital, the inconsistent use of terminology standards to identify IVD tests and the results thereof further hampers efforts to manage these data electronically. For example, a search on the term "serum glucose" returns over 600 LOINC terminology bindings. The absence of context-specific semantic interoperability is recognized as a substantial impediment to use of electronic IVD data by patients, providers, payers, industry and government agencies alike. (41) For diagnostic test information to seamlessly populate electronic databases, at least the construct of sample time will need to be included. Ideally, data would be collected directly from the source (e.g., from the LIS), as opposed to a downstream HIT system (e.g., EHR).

Similar to medications, a set of concepts is needed to capture IVD test results. The concepts comprising the model are listed in Table 5.

Table 5. Laboratory Data Model Concepts

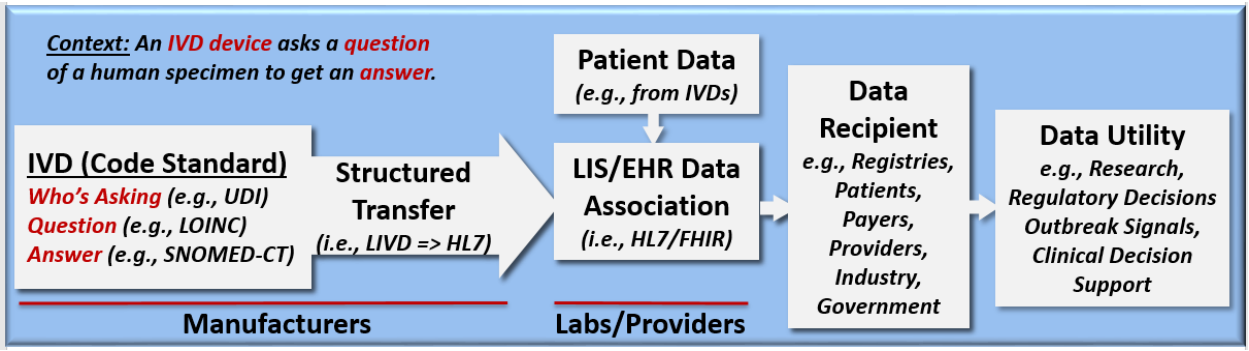
Laboratory Data Concept	**Definition**	**Applicable Standard**
IVD test	Common name of a laboratory test	LOINC
IVD sample date / time	Date / time that the specimen being tested is obtained	
IVD result (quantitative)	Laboratory test result	
IVD result units	Laboratory test result units of measure	UCUM
IVD result (qualitative)	Laboratory test result (non-numeric)	SNOMED-CT

IVD = in-vitro diagnostic

Particularly relevant to this project have been a series of discussions with FDA about the SHIELD framework. First, FDA has recognized LOINC as the reference code set that identifies an IVD test (42) and SNOMED-CT as the reference for codes that describe qualitative test allowed values. (43) Use of these bindings permits the transmission of IVD data in a structured transmission format – LOINC to Vendor IVD Test Results (LIVD). (44) For quantitative tests, Unified Codes for Units of Measure (UCUM) is recognized as the authority for units of measure. Working with HL7 to aid LIVD implementation, SHIELD stakeholders have developed an implementation guide on the specifics of how to convey existing and future LOINC to IVD Test Code mappings using FHIR resources (45) and will not be replicated in the Project report. This will enable laboratories to more efficiently and effectively map lab test/analyte codes between the two code systems. To provide additional regulatory clarity in support of the efforts of the IVD industry, the FDA has published a guidance, *Logical Observation Identifiers Names and Codes for In Vitro Diagnostic Tests*. (46) The programmer is referred to these documents for specific recommendations regarding accomplishing native interoperability for laboratory data.

As illustrated in Figure 1 below, the SHIELD initiative is shepherding plans for IVD test manufacturers to provide harmonized codes to downstream systems. In other words, the same laboratory test concept (regardless of instrumentation or other variance) is to be identified using the same LOINC code across manufacturers. The Regenstrief Institute has been identified as the entity to provide authoritative oversight. Data will be transmitted in LIVD format to LIS systems. For laboratories and other consumers who have implemented FHIR, the data transfer process will be even further simplified (e.g., to EHR systems) given the FHIR-embedded specifications for the data itself.

Figure 1. SHIELD Framework for Laboratory Data Interoperability



In this framework, when laboratory systems pass data to downstream systems (e.g., registries), consumers should expect that semantic harmonization will be achieved via the bindings to LOINC, SNOMED-CT, or UCUM (depending on data type). If these requirements are built into registry databases by the developers, this will improve the ability to electronically receive and utilize laboratory data, driving successful interoperability across data partners.

C. Concordance of Common Data Models, FHIR, and USCDI

Results of the analysis of the concepts represented in common data models (CDMs), FHIR, and related resources is tabulated in Appendix 2. The resources evaluated included:

- Sentinel Common Data Model v 6.0.2
- PCORnet Common Data Model v4.1

- Observational Health Data Sciences and Informatics (OHDSI) Common Data Model v5.3.1
- HL7 FHIR v3.0.1
- HL7 Common Clinical Registry Framework – Domain Analysis Model
- ONC USCDI – Jan 2018 draft

The CDMs (Sentinel, OHDSI, and PCORnet), FHIR, the HL7 Common Clinical Registry Framework (CCRF) initiative, and the USDCI are national to international resources that aim to facilitate the interchange of healthcare data, albeit from different perspectives. The CDMs define database structures for sets of healthcare concepts and create the opportunity for interchange of those clinical concepts as standardized data; FHIR is more focused on the exchange of standardized data per se. The HL7 Common Clinical Registry Framework initiative is specifying a set of core common clinical data elements (aligned with the Project) to populate a domain analysis model for registries. Common to all of these is the leveraging of standard terminologies and the coding systems thereof.

The concordances among the approaches was found to depend on the design and management of the resources themselves. Because the PCORnet CDM is based on the Sentinel CDM, a high level of concordance between the two models was both anticipated and observed. For many of the quantitative concepts, like vital signs, there was a high level of concordance in how the concepts are captured as data; however, the PCORnet CDM captures greater granularity than Sentinel for several concepts (sex, race, ethnicity). The CCRF Domain Analysis Model borrows heavily from the ONC CCDS with respect to the data elements in the model; concordance with the ONC USCDI was thus anticipated. Finally, the Observational Health Data Sciences and Informatics (OHDSI) CDM, a common platform for analysis of disparate observational databases, takes a different path from the other national models in its structure, use of concepts, concept IDs and linkages to terminologies, and was found to be highly discordant compared with the other approaches.

Of the 18 core clinical concepts identified as being shared across registries, 6 of the 18 (33%) were represented in all of the resources (Appendix 2). The concepts appearing universally were demographic information (date of birth, sex, race, ethnicity), procedure identifier, and laboratory data. Notably, there were differences among these resources in terms of: a) data element names, b) allowed values in the allowed value sets, and c) application of SDO standards.

Another 6 of the 18 were concepts were included in 5 of the 6 resources: vital signs (height, weight, systolic blood pressure, diastolic blood pressure), smoking status, and care team members. Again, among these resources, concepts differed in data element names, degree of granularity in the allowed value sets and use of SDO standards (e.g., allowed values for smoking status). Device data was represented in 4 of the 6 (67%), specifically via the unique device identifier.

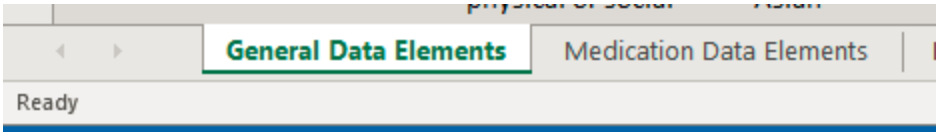
Finally, the core clinical concepts of this project that were least frequently represented fell into 2 buckets: a) patient vital status (e.g., death) and 2) use of illicit substances other than cigarette smoking (e.g., other forms of tobacco, alcohol consumption, and non-medicinal substance use). These appeared in less than half of the resources. None of the resources defined alcohol consumption or illicit substance use concepts in a consistent, succinct, interoperable manner.

D. Implementation Guidance

Singularly, the key artifact of this project is the list of core common clinical data elements and associated metadata for building in registry databases and other EHI systems. The core common clinical data elements and metadata crosswalk is included in Excel spreadsheet format and is available on the project website at https://dcri.org/registry-data-standards. The contents are intended to serve as an implementation guide and reference source for database developers who build or manage registry databases.

Spreadsheet Tabs

The following provides an orientation to the spreadsheet. The spreadsheet has 3 tabs, for single prompt:value concepts (termed General Data Elements), medication data elements, and laboratory test data elements. Within each tab, rows list the specifications for data elements and (where appropriate) allowed values; columns list the associated metadata for the data elements included in the tab. This screen shot illustrates two of the three tabs:



The first tab (**General Data Elements**) contains data elements that characterize registry participants (patients, subjects). These include first & last name, date of birth, & data of death as well as physical attributes such as race, ethnicity, sex; vital signs height, weight, blood pressure and heart rate/pulse; lifestyle factors such as smoking, drug and alcohol exposure. Also included are data elements that describe clinicians/care providers, medical procedures and implantable devices.

The second tab (**Medication Data Elements**) contains a core set of data elements that characterize medications. Included are medication name, dose, dose units, class, & code. These represent a core/minimum set of information about medications.

The third tab (**Lab Data Elements**) contain a core set of data elements that characterize laboratory tests and results. Included are test name, sample date/time, result (quantitative), result units, and result (qualitative).

Data Elements

On each tab of the spreadsheet, data elements and respective allowed values are listed in each row (one row for the data element name; one row for each allowed value where a value set applies).

2 Data Elements; one row each)

Given Name [First Name]	The given or first name of the patient.	GIVEN_NAME, Patient.name.given	Char(140)	Allow any number of given names per patient.			LOINC 45392-8 (First Name)		FHIR Profile https://www.hl7.org/fhir/us/core/Structur		
Family Name [Last Name]	The family or last name of the patient.	FAMILY_NAME, Patient.na	Char(70)	Allow only one family name per patient.			LOINC 45394-4 (Last Name)		FHIR Profile: https://www.hl7.org/fhir/us/core/Structur		

1 Data Element / Value Set Pair; multiple rows

Sex [Birth Sex] [Sex (Birth Sex)]	The biological sex of the patient, assigned at birth, not to be confused with the social construct of gender.	F M UNK [Female, Male, Unknown]	SEX, birthsex	Value Set - Char(3)				LOINC: LL3324-2 (Sex assigned at birth)		FHIR references: https://www.hl7.org/fhir/us/core/StructureDefinition-us-core-patient.html; FHIR Resource: https://www.hl7.org/fhir/us/core/StructureDefinition-us-core-	2015 CCDS and USCD Office of the Nation Coordinator for Hea IT. Aligned with the CDA Birth Sex observation
					F	Female		LOINC: LA3-6			
					M	Male		LOINC: LA2-8			
					UNK	Unknown - a proper value is applicable, but not known.		LOINC: LA4489-6			

Metadata

Metadata for the registry concepts developed as core clinical data elements were identified to provide database developers with the basic information needed to implement the recommended data elements in a database. They are intended for use as a starting point; these recommendations are intentionally not inclusive of all metadata required for detailed domain, logical or clinical modeling or representation in a formal data standard such as ISO 11179-3. Related metadata are listed in a single column, rather than devoting separate columns to each specific type of metadata (e.g., FHIR Resource and FHIR Profile URLs are listed together in the FHIR column). Recommended metadata associated with each data element include (where appropriate) the concepts below. NOTE: Labels of each type of metadata were NOT copied from any existing terminology repository, but instead the text of each of the labels of the types of metadata were explicitly and purposefully crafted for maximal clarity and to be self-explanatory.

- Clinical concept label: this is the recommended text to print on a case report form / data capture form, or to display on an electronic data entry screen. Suggested alternatives or aliases are shown in brackets (e.g., Sex, [Birth Sex], [Sex (Birth Sex)]). The intent is for this to be the human readable "question" or "prompt" of the clinical concept.
- Clinical definition: this is the clinical definition of the concept. The clinical definition is intended for a clinical audience, in language and terminology generally understood and interpretable by clinicians. NOTE: This is the clinically useful definition of the concept – NOT instructions about how to respond to the prompt (e.g., if the clinical concept label is "History of hypertension", the clinical definition should NOT be "check the box if the patient has a history of hypertension", but instead be the definition of hypertension to be used to respond to the prompt).
- Clinical allowed values: for clinical concepts with a constrained set (pick list) of responses, these are the allowed or permissible values that comprise the list of responses. The intent is for this to be the human readable list of choices (e.g., labels for a series of radio buttons or checkboxes, or the content of a dropdown list on an electronic data collection form).
- Clinical allowed values definitions: analogous to the clinical definition, these are the individual clinical definitions of each of the allowed value responses.

The first 4 items in this list (clinical concept label, clinical definition, clinical allowed values, and clinical allowed values definitions) purposefully include the word "clinical" in the name of each metadata item, to denote that these are the items typically relevant to the clinician perspective. These 4 items also typically comprise the extent of items described and published in compilations of "clinical data standards' by professional societies and other academic groups.

- Database field label: this is the recommended label for a database developer to use for the field in a database system. Generally, this is the "address" used by a database system to identify the data content in the corresponding data field.
- Database field data type / format: this is the recommend format for the field in the database system (e.g., value set, text (Char), number, date, etc.)
- Database field business rules: where applicable, this is a minimum recommended set of rules to assure data consistency and quality (e.g., cardinality [where not explicitly obvious], edit checks, range checks)
- Database allowed values: for clinical concepts with a constrained list of responses, this is the recommend value to store in the database field. The database allowed values may or may not vary from the corresponding clinical allowed values. When different, the intent is to increase the likelihood that the stored value will be natively interoperable.

The second 4 items in this list (database field label, database field data type / format, database field rules, database allowed values) purposefully include the word "database" in the name of each metadata item, to denote that these are the items typically relevant to the perspective of the database programmer. In addition to the first 4 (clinical) metadata, there are the items needed by a database programmer to understand and capture the data for each clinical concept in a database system.

- Reference ontology concept bindings – this is the recommended reference code set binding for the clinical concept. Preference is given to bindings associated with the concept as represented in FHIR.
- Reference ontology allowed value bindings – for clinical concepts with a constrained list of responses, these are the recommended reference code set bindings for the allowed values. Preference is given to bindings associated with the allowed values as represented in FHIR.
- OID: this is the authoritative object identifier corresponding to the clinical concept and (where applicable) value sets as assigned by the International Organization for Standardization (ISO). OID values OIDs link standard concepts to unchanging reference terminologies such that consistent semantic meaning is preserved and maintained.
- FHIR references – FHIR profiles and resources that reference the clinical concept and (where applicable) allowed values.
- Other sources / references / notes – reference comments that describe source documents and other notes about the clinical concept and the metadata thereof.

The third set of 4 metadata items (Reference ontology concept bindings, reference ontology allowed value bindings, OID, and FHIR references) are included to enhance the potential for interoperability of the concepts, particularly when content must be transformed to move the data from one database system to another. The reference ontology entries are the recommended bindings to the clinical concepts and (where applicable) allowed value responses of those clinical concepts, along with the corresponding object identifier in the VSAC. Preference is given to the NLM VSAC as the reference source given its responsibility for long-term maintenance of reference mappings. FHIR profiles and resources that served as reference source information are listed as FHIR references.

E. Feedback

Socialization of the Project was extensive across the stakeholder community throughout the duration of the project, with specific engagement of the registry community, selected informaticians focused on clinical vocabularies, and representative federal agencies. The Project solicited feedback by direct communication (e.g., email, conference calls) and formally during via 3 venues:

- During the in-person meeting of stakeholders held on August 21, 2018 at The Pew Charitable Trusts facility, Washington, DC.
- During the Q & A of the stakeholder webinar held on September 24, 2018 (recording and transcripts are available at https://dcri.org/registry-data-standards/).
- Via Qualtrics survey distributed via email on September 26, 2018 (distribution list included over 300 recipients).

Following the in-person meeting (August 21, 2018), two independent reviewers analyzed 85 pages of transcripts of the meeting and break-out session per mixed-methods research methodologies using NVIVO 11 software to identify ideas or concerns not previously noted. There were 149 unique, unstructured comments categorized into the structured themes of Administrative Contributions/Burdens, Advocacy Opportunities, Data Harmonization, Quality & Patient Safety, and Usability. The key finding is that there were no notable divergences between the quantitative product and qualitative contributions. The full analysis is available at the Project website https://dcri.org/registry-data-standards/.

Select questions representing themes and topics of respondents to the Qualtrics survey are summarized as follows.

Q: *The project identified a key set of 12 metadata fields necessary to both clinically define concepts and build those terms in registry databases (and other EHI systems). What other metadata are required by clinicians and database developers?*

Responses (n = 13) for this question ranged from 'metadata fields proposed will cover the common challenges across registries' to suggestions for metadata fields that are 'not necessary but might be helpful' (the most common suggestion was to set context, i.e., 'intent of data collection'). Cardinality was mentioned by several respondents to describe whether a response is required or optional and/or the number of times a data element could be valued. (We believe that cardinality is best defined by the registry use case and thus did not include an explicit metadata field for cardinality.) The method used to obtain the data element response was suggested (e.g., measured vs reported, auscultated vs palpated, or location used to obtain a measurement). Versioning of data elements and allowed values/code sets was also recognized as critical metadata. It was recommended that the OID (object identifier) has permanence across versions; the OID was therefore added by the Project team to become the 13th metadata element for characterizing CDEs .Other responses for this question identified additional data elements to be recommend in future versions of the work product. Examples include patient identifiers such as Social Security Number, middle name/initial and other concepts to facilitate linkage of records across data sets. A few suggestions were felt to be clearly out of scope for a core set of common registry data elements (e.g., indicator of correctness of sex assigned at birth, 'adult' or 'child' characterization, pregnancy status).

There were also 13 responses characterizing issues not addressed by the proposed metadata. These include validation of consistency of semantic meaning across systems, number of expected values in response to a prompt, changes in meaning and updates to terminology as new knowledge is generated, compliance with privacy and / or data use agreements, degree of certainty or confidence in the value reported, and enhancing the ability to link records across datasets from different sources. These operational and governance aspects were felt to be outside the scope of the Project.

Q. The project identified 15 primary concepts frequently present in registries, which are being developed as approximately 30 data elements (rows of the spreadsheet). Please identify content errors, suggestions, and edits.

There were many insightful responses to this question (n = 18). Additional patient identifiers were suggested along with metadata describing provenance and governance of data elements, value sets and the data collected therein. Recommendations for format / data structure for dates (ISO 11179 vs other) and text responses (e.g., "char" vs "varchar") as well as alternate LOINC codes, categories for race, ethnicity, and smoking status, and multiple flavors of null were suggested. Most suggestions were tagged for consideration in future versions of this work, however, a few were actionable in this phase of work and the necessary updates were made to the recommendation spreadsheet. Specific examples include the following. Numerous suggestions for alternate LOINC codes resulted in consultation with a terminologist and updates to the content. Smoking status, alcohol use, and illicit substance use were widely recognized as problematic – these concepts were adjusted and updated, particularly from the perspective of clinical use. Business rules were adjusted to specify conditions in which other data elements are required (e.g., when x is valued, y is required) or when multiple response options may be selected (e.g., race) versus when only one response is allowed (e.g., ethnicity). Efforts to provide links to FHIR resources, FHIR profiles and other references were included. The levels of detail represented by data elements and value sets were re-evaluated. Collectively, an effort was made to align with federal guidelines, FHIR and other

widely adopted standards. Suggestions for improvement in the OMB/federal value sets and the ONC/USCDI value sets will be triaged to the respective authoritative sources.

One comment in particular was noteworthy as it clearly articulates the value of and support for the Project: '… simply capturing LOINC, SNOMED-CT and UDI in association with the other data elements will dramatically improve the interoperability of the healthcare data…." Also insightful and interesting was the comment regarding the need for data elements that characterize clinically significant attributes about implanted medical devices; of note, efforts in the medical device community to develop a repository for listing of data about device characteristics is in progress. (47)

Q. There was little concordance among registries about how to document smoking status, alcohol use, and illicit drug use. What approach would you recommend to address this lack of consistency?

Responses (n = 18) to this question ranged from "not applicable to our registry" to recommending an iterative, evidence based approach targeted at specific use cases. Already employed as part of this effort are the environmental scan – an assessment of current practices, search for existing standards (in particular, the work of the National Institute on Drug Abuse), and recognition of the differences that state laws, impact of patient-reported responses, the necessity for reproducibility, and recognition of the value for crosswalks between several instruments. Quantitative or integer responses were preferred over narrative or qualitative responses. There was recognition that judicious use of the data is important as well as the clinical significance and utility of the responses, coupled with sensitivity to the level of detail balanced with data collection burden.

Q. Is your registry database system compatible with the suggested Simple Database Field Label? Is your database system compatible with the FHIR Database Field Label?

There were 8 who responded in the affirmative, that their system was compatible; 4 responded that compatibility was 'not applicable'; 3 replied with uncertainty (not sure); and 5 replied as 'No/Not compatible". One organization responded that their registry was compliant with the simple database field labels but they did not think their system was compatible with the FHIR database field labels. This has obvious implications in terms of native interoperability if FHIR database field labels cannot be programmed into databases.

Q. Does your registry currently use or plan to use standards development organization (SDO) / reference ontology bindings (i.e., LOINC, SNOMed CT) to support the interoperable transfer of data from EHR to registry?

There were 7 affirmative responses, 6 not applicable, 1 not sure, and 2 "no" responses (1 was ambiguous). In terms of why not – the comments ranged from 'existing reference ontology bindings are woefully inadequate' to 'we have mapped most of our terms to SNOMED-CT and have no plans to apply for additional terms". One response expressed hope that use of LOINC and SNOMED-CT would be supported / adopted by the registry community. Others described plans to integrate SNOMED-CT, LOINC, RxNorm through SOLOR. Overall, SNOMED-CT and LOINC were mentioned most often by those who mapped their terms to an ontology. Cancer registry standards were described as being available prior to contemporary SDO's, specifically use of the NAACCR Data Standards for consistency. Another reason for not mapping to standard terminologies was that their institution provided mappings to ICD and CPT codes, but not others; one organization noted that lack of resources to do the mapping was a contributing factor.

Q. Other comments or suggestions?

Several referenced the lack of identifiers and impact on patient linkages; one recommended PhenX as a resource that should be considered. One advised using numeric vs character string values and cited the difficulty that comes with analyzing / searching for character strings vs finding fields using a numeric value. There were suggestions to be more precise and unambiguous with code sets and to reconcile the Sample Completed CRF pdf with the recommendations in the spreadsheet. These latter suggestions were incorporated into the work product of the Project.
One remark was particularly encouraging and supports continued efforts to advance healthcare data interoperability: "This work is invaluable to helping advance healthcare data interoperability efforts in ways that will both improve patient care and reduce the overall financial and labor burdens to the healthcare ecosystem. It will be critical to make a thorough evaluation on the return on investment from this work so it can be used to engage institutions that have yet to take the leap towards investing in their own interoperable infrastructure. If this can serve as an example, it can begin to break down data silos and expand healthcare data utility."

The Project team greatly appreciated the thought, effort and support from the stakeholder community. Suggestions were carefully considered and effort expended to incorporate these where possible.

DISCUSSION and RECOMMENDATIONS

The call for interoperability of data in healthcare is neither new nor novel, with work to accomplishing this goal extending across multiple decades. (48, 49) CMS has authored an explicit roadmap to accomplish interoperability nationwide (3), and Congress has specifically addressed the need to accelerate and accomplish interoperability in the 21st Century Cures Act (4). Non-government organizations such as the Health Information Management Systems Society have collated literature, authored white papers, and created sets of tools critical to the operational implementation of standards-based, semantically interoperable health information exchange. (50) Despite this leadership, the Project demonstrates that the registry community has not benefitted from, is not aligned with, and does not contribute to interoperability efforts. At least from the perspective of clinical registries, semantic interoperability remains a distant goal. This suggests that we are at crisis in achieving semantic interoperability.

The absence of native data standardization as an enabler of semantic interoperability is not unique to the registry community. Across the clinical care and research domains, the generation of most biomedical data is highly distributed and is accomplished mainly by individuals, whether clinicians or scientists. At the enterprise level, few systems have succeeded in capturing standardized data even locally. (1) Moreover, data are captured and stored in a wide variety of formats, compounding the inability of analysts and researchers to identify and use biomedical data generated by others and creating the requirement for repetitive and extensive data "cleaning." A 2016 survey found that data scientists across a wide array of fields spend the majority of their work effort (upwards of 80 percent) in data "janitorial" services: identifying and collating existing datasets, and cleaning, preparing, and organizing the data in those datasets for analysis purposes. (51)

What would semantic interoperability look like from a registry perspective? The ability to transfer data between computers is only a component of a larger framework. Other elements of the framework include the capture of clinical concepts as data (via CDEs) at the point of care in source HIT / EHR documentation systems, storage of data (not just documents) that are natively interoperable in those EHR systems, the compilation and packaging of data specific to a registry from source systems, the electronic transmission of data to the registry, and acceptance of that data in an electronic format. And all components must work seamlessly together for the ideal state to be achieved. Given the multiple processes involved, it should come as no surprise that this project demonstrated a lack of consistent adherence to interoperability standards. Furthermore, this lack of adoption of standards appears to extend beyond the registry community to EHR and other HIT systems, collectively hampering efforts to efficiently capture and use clinical data. In part, the poor adoption may be related to the lack of an authority with overarching command of the interoperability framework and the capacity to guide individual efforts – it was surprisingly difficult for our expert team to correctly identify standards for the programmer's implementation guidance, even though we explicitly approached the project intending to identify the "best" existing standard (rather than create yet another standard).

The project did not set out to explicitly understand the reasons for the gap between desired state and the reality of registries. Nonetheless, we did glean multiple insights that could contribute to understanding the pathway forward. Admittedly, the paradigm of "capture once, use many times" (with the focus of the Project being the registry use case) is complex and multidimensional. Alignment and harmonization across the other stakeholders involved in healthcare data interoperability – clinicians, researchers / academia, healthcare systems, healthcare product manufacturers, HIT vendors, standards development organizations, payers, government agencies, and patients – must be coordinated for the registry community to benefit (and vice versa). Unfortunately, failure of any individual component compromises the whole. While we identified core content (e.g., FHIR profiles) that will work for registries, that alone appears insufficient. Instead, the package of obligations that the registry community should embrace appear to include the following.

- Prioritize work that anticipates the interoperability of healthcare data
    - Establish the goal of interoperable, electronic data transfer from source documentation to registry
    - Build the core clinical common data elements delineated in the Project into all registry databases, either at the physical level or at least as a representational layer
    - Discuss with registry data contributors (and their respective EHR and HIT vendors) how to capture and transmit data electronically from source documentation system to registry (rather than via an intermediary system, or otherwise requiring human interdiction)
    - Encourage use of electronic specifications (e.g., .xml) for the receipt of data
    - Extend data dictionaries to more completely qualify the concepts captured in a registry system, inclusive of the 13 metadata types of this Project
- Recognize and contribute to those components of the interoperability ecosystem that depend upon clinical subject matter expertise possessed by registry owners
    - Identify and define domain-specific, core clinical concepts, defining the concepts as clinical data elements inclusive of the 13 metadata types of the Project
    - Deeply understand how clinical data elements must serve clinical documentation needs, assessment of processes, and outcomes determinations as the first priority – aiming for parsimony (i.e., a minimum core set) rather than all-inclusiveness
    - Align and coordinate across disciplines where there are areas of overlap (e.g., Registry Assessment of Peripheral Interventional Devices (RAPID) – a project that developed a peripheral artery disease vocabulary now shared across cardiology, interventional radiology, and vascular surgery) (52)
- Develop in-house expertise in informatics, and particularly the informatics of interoperability, and require the use of interoperability standards for the capture and submission of data to registries
    - Employ informaticians to model domain-specific clinical concepts per processes such as the Clinical Information Modeling Initiative (53)
    - Re-imagine and re-create registry instruments (e.g., data capture forms) based upon data being captured at the point of care – rather than as derived, summative concepts
    - Disseminate best-practice approaches for the alignment of data capture integrated into clinical workflows, emphasizing the capture of core clinical data at the point of care with the explicit intention of reuse of that data (8)
    - Include informatics expertise in organizational management and governance structures

Our research identified 2 additional frameworks that further articulate and argue for the case for data liquidity while providing guidance regarding data interoperability. The FAIR Guiding Principles were authored by a stakeholder group including academia, industry, funding agencies, and scholarly publishers to address the "urgent need to improve the infrastructure supporting the reuse of scholarly data". (54) The 4 key principles identified by this group are that biomedical data should be findable, accessible, interoperable, and reusable, with a "… specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals." To be interoperable, the FAIR Guiding Principles articulate that "… data should 'use and speak the same language' via use of standardized vocabularies". The work of Drolet and Johnson to disambiguate and categorize healthcare registries into specific classes includes a framework of five characteristics to evaluate in the categorization process. (10) This framework, given the acronym MDR-OK, reflects basic requirements of registries within the interoperability framework described above: mergeable data, dataset standardized, rules for data collection, observations associated over time, and knowledge of outcomes.

A key work product of the project was the authoring of an "easy button" (55) implementation guidance specifically for database programmers (https://dcri.org/registry-data-standards). This was founded on the

observation by the investigators of a lack of consistency in the representation of common clinical concepts (e.g., sex, vital signs, tobacco use, laboratory data, medications) across clinical registries to which the investigators (and their respective hospitals) contribute data. The purpose is to provide database programmers an authoritative, "just the facts" reference resource for the concepts we identified as being routinely captured across registries. In other words, the implementation guidance is intended to provide database programmers the responses needed to build the core clinical data elements into database systems. The style of the implementation guidance is purposefully minimalist and practical, rather than a formal instantiation of the suggested CDEs following a data element development standard such as ISO/IEC 11179. Data element examples where this approach – favoring quality over completeness – are numerous. For example, we intentionally elected to develop only last (family) name and first (given) name, ignoring middle name, salutation, and other name-related components. For blood pressure, we purposefully did not develop the associated units of measure, since blood pressure is always measured in millimeters of mercury. Another example is date – while the representation of date and time is an international standard in ISO 8601, we elected not to specify a date format, since date ETL processes are already built into database systems and are thus already largely interoperable. The pragmatic approach we took is NOT intended to suggest that middle name, blood pressure units of measure, or laboratory methodology could or should not be included in a database should the need or desire to include same exist; in fact, there is nothing about the implementation guidance that should be interpreted as mandatory, nor should the implementation guidance be considered all-inclusive. Indeed, formal modeling (e.g., via the CIMI approach (53)) still remains to be accomplished.

Findings about several specific concepts deserve further reflection. First, we did not find a single instance where laboratory data was transferred electronically from source system to registry using LOINC bindings. In part, we believe this is because many laboratory values have multiple potential LOINC code bindings. The SHIELD project, led by FDA, includes the creation of single high-level LOINC codes for common laboratory data, a promising solution to the multiple bindings issue. (27) For this project, we elected not to specify LOINC codes for specific laboratory tests, anticipating the availability of a unifying approach in the very near future. Second was the discordance between representations of tobacco abuse captured in registries versus the recommended structure for same in the USCDI. While the concept of tobacco abuse was only captured in 11 registries, the degree of discordance among those registries suggests that the USCDI structure and value sets could be improved to better reflect the application of granular data for purposes beyond smoking cessation. To this end, a framework for the capture of tobacco abuse inclusive of types of tobacco used (i.e., not limited to only cigarette smoking), quantitation of the amount of tobacco consumed, duration of tobacco exposure, duration of cessation of tobacco exposure, and other concepts may need to be developed to supplant or replace the current USCDI cigarette smoking concept. Third, despite the emphasis of recent programs on the misuse and abuse of narcotics and other illicit substances, there was no consistency among the registries, and no clear-cut standards, for the capture of granular data regarding illicit substance abuse and alcohol consumption.

There are a number of limitations of this project. As noted above, the project purposefully delineated CDEs using only 13 meta-data elements, rather than complete and formal informatics modeling. This was a considered design choice specific to the primary work product, the implementation guidance. Longer term we do intend to utilize the work product to support formal modeling as a project of the CIMI modeling process. Another limitation is that we did not survey the universe of all healthcare registries – neither time nor resources allowed expansion of the project. Nonetheless, even with a convenience sample, we believe the registry artifacts reviewed in this project are representative, given the approach to solicitation for participation (open call at both the Council of Medical Specialty Societies and National Quality Registry Network meetings, groups which include membership of essentially all professional society registry owners and many non-society registry owners) along with the depth and breadth of the subject matter covered across the registry participants. Finally, one meta-data item was explicitly not developed – the context of the CDE (e.g., history, pre-procedure, follow-up period). Within registries,

context is generally self-apparent. For example, patient name, race, ethnicity and sex are considered demographics, while medications are captured across a multiplicity of contexts (e.g., at baseline, pre-procedure, intra-procedure, during follow-up). As this project was an analysis of registry artifacts, developing a context concept as meta-data was not necessary. However, external to the registry use case, specification of context will likely be needed, along with an ontology that accomplishes same.

In conclusion, the specification and implementation of well-defined common clinical data elements by and within registries has tremendous potential to advance the interoperability agenda. Standardized data collected at the point of care and used across the healthcare continuum has the potential to save millions to billions of dollars annually, improve approaches to assessing healthcare quality and delivery, and result in the betterment of patient care and outcomes.

Acknowledgements

The principal investigator (James Tcheng) and the Project team wish to express thanks to the many contributors to this project, particularly the registry owners and the stakeholders who responded to our request for participation. We wish to acknowledge the leadership of the Pew team of Anqi Lu, Ben Moscovitch, and Joshua Rising; the convening function of the registry community by the Council of Medical Specialty Societies (Helen Burstin) and the National Quality Registry Network (Seth Blumenthal, Chrystal Price, Marjorie Rallins); the clinician leadership provided by Joseph Drozda and John Windle, and the research team including Davera Gabriel, Anne Heath, Julia Skapik, Asba Tasneem, Michael Waters, Rebecca Wilgus, Mary Williams, and Thomas Windle for their review and contributions to the work and this report.

The project was supported by a grant from the Pew Charitable Trusts.

References:

1. http://digitaldukemed.mc.duke.edu/databank/. Accessed November 1, 2018.

2. HIT Standards Committee - Clinical Operations Workgroup Report. Department of Health and Human Services, Office of the National Coordinator for Health Information Technology. August 20, 2009.

3. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap, Final Version 1.0. Department of Health and Human Services, Office of the National Coordinator for Health Information Technology. 2015.

4. https://www.congress.gov/bill/114th-congress/house-bill/34/. Accessed November 1, 2018.

5. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHrIncentivePrograms/. Accessed November 1, 2018.

6. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems, implications for outcomes research. Ann Intern Med. 1993 Oct 15;119(8):844-50.

7. Messenger JC, Ho KK, Young CH, et al. The National Cardiovascular Data Registry (NCDR) Data Quality Brief: the NCDR Data Quality Program in 2012. J Am Coll Cardiol. 2012 Oct 16;60(16):1484-8. doi: 10.1016/j.jacc.2012.07.020. Epub 2012 Sep 19.

8. Sanborn TA, Tcheng JE, Anderson HV, et al. ACC/AHA/SCAI 2014 health policy statement on structured reporting for the cardiac catheterization laboratory: a report of the American College of Cardiology Clinical Quality Committee. J Am Coll Cardiol. 2014 Jun 17;63(23):2591-2623. doi: 10.1016/j.jacc.2014.03.020. Epub 2014 Mar 28.

9. Gliklich R., Dreyer N. Registries for Evaluating Patient Outcomes: A User's Guide: 2. Agency for Healthcare Research and Quality Publication No. 10-EHC049. Agency for Healthcare Research and Quality, Rockville, MD. 2010.

10. Drolet BC, Johnson KB. Categorizing the world of registries. J Biomed Inform. 41:1009-20, 2008. doi: 10.1016/j.jbi.2008.01.009.

11. Terry S, Runkle D, Wicks P. Patient- or Participant-Generated Registries: Addendum to Registries for Evaluating Patient Outcomes: A User's Guide, Third Edition. Agency for Healthcare Research and Quality, Rockville, MD. 2018. PMID:29671988.

12. Blumenthal S. The use of clinical registries in the United States: a landscape survey. EGEMS (Wash DC). 2017 Dec 7;5(1):26. doi: 10.5334/egems.248.

13. Gliklich R, Dreyer NA, Leavy MB, Velentgas P, Khurana L. Standards in the conduct of registry studies for patient-centered outcomes research. Patient-Centered Outcomes Research Institute. March 15, 2012.

14. Anderson HV, Weintraub WS, Radford MJ, et al. Standardized cardiovascular data for clinical research, registries, and patient care; A report from the data standards workgroup of the National Cardiovascular Research Infrastructure Project. J Am Coll Cardiol. 2013 May 7;61(18):1835-46. doi: 10.1016/j.jacc.2012.12.047. Epub 2013 Mar 6.

15. Weintraub WS, Karlsberg RP, Tcheng JE, et al., 2011 ACCF/AHA key data elements and definitions of a base cardiovascular vocabulary for electronic heath records. J Am Coll Cardiol. 2011 Jul 5;58(2):202-22. doi: 10.1016/j.jacc.2011.05.001. Epub 2011 Jun 7

16. Hicks KA, Tcheng JE, Bozkurt B, et al. 2014 ACC/AHA Key Data Elements and Definitions for Cardiovascular Endpoint Events in Clinical Trials. J Am Coll Cardiol. 2015 Jul 28;66(4):403-69. doi: 10.1016/j.jacc.2014.12.018. Epub 2014 Dec 29.

17. Califf RM, Robb MA, Bindman A, et al. Transforming Evidence Generation to Support Health and Health Care Decisions. N Engl J Med. 2016 Dec 15;375(24):2395-2400. DOI: 10.1056/NEJMsb1610128.

18. Sedrakyan A, Cronenwett JL, Venermo M, Kraiss L, Marinac-Dabic D, Björck M. An international vascular registry infrastructure for medical device evaluation and surveillance. J Vasc Surg. 2017 Apr;65(4):1220-1222. doi: 10.1016/j.jvs.2017.01.020. Epub 2017 Feb 17.

19. Grover FL, Vemulapalli S, Carroll JD, et al., for the STS/ACC TVT Registry. 2016 annual report of the Society of Thoracic Surgeons/American College of Cardiology Transcatheter Valve Therapy Registry. J Am Coll Cardiol. 2017 Mar 14;69(10):1215-1230. doi: 10.1016/j.jacc.2016.11.033. Epub 2016 Dec 9.

20. Kumar A, Matheny ME, Ho KK, et al. The data extraction and longitudinal trend analysis network study of distributed automated postmarket cardiovascular device safety surveillance. Circ Cardiovasc Qual Outcomes. 2015 Jan;8(1):38-46. doi: 10.1161/CIRCOUTCOMES.114.001123. Epub 2014 Dec 9.

21. Pappas G, Berlin JA, Avila-Tang E, et al. Determining the value of Coordinated Registry Networks (CRNs): a case of transcatheter valve therapies. *Personal communication / submitted for publication.*

22. https://qpp.cms.gov/mips/overview. Accessed November 1, 2018.

23. Tonner C1, Schmajuk G, Yazdany J. A new era of quality measurement in rheumatology: electronic clinical quality measures and national registries. Curr Opin Rheumatol. 2017 Mar;29(2):131-137. doi: 10.1097/BOR.0000000000000364.

24. Toussaint JS, Correia K. Why process is healthcare's biggest problem. Harvard Business Review, 2018 Mar 19.

25. FDA/CDC/NLM Workshop on Promoting Semantic Interoperability of Laboratory Data (2015). https://wayback.archive-it.org/7993/20170111193824/http:/www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm453897.htm. Accessed November 1, 2018.

26. CDC/FDA/NLM/ONC/CMS Workshop on Promoting Semantic Interoperability of Laboratory Data (2016). https://wayback.archive-it.org/7993/20170111091048/http://www.fda.gov/MedicalDevices/NewsEvents/WorkshopsConferences/ucm523316.htm. Accessed November 1, 2018.

27. National Evaluation System for healthcare Technology (NEST) Demonstration Projects: Systemic Harmonization and Interoperability Enhancement for Laboratory Data (SHIELD). https://nestcc.org/demonstration-projects/. Accessed November 1, 2018.

28. https://www.healthit.gov/sites/default/files/ccds_reference_document_v1_1.pdf; https://www.healthit.gov/sites/default/files/commonclinicaldataset_ml_11-4-15.pdf. Accessed November 1, 2018.

29. https://www.healthit.gov/sites/default/files/draft-uscdi.pdf. Accessed November 1, 2018.

30. http://www.hl7.org/fhir/overview.html. Accessed November 1, 2018.

31. http://www.hl7.org/fhir/profilelist.html. Accessed November 1, 2018.

32. http://www.hl7.org/fhir/us/core/history.html. Accessed November 1, 2018.

33. Bailey C, Peddie D, Wickham ME, et al. Adverse drug event reporting systems: a systematic review. Br J Clin Pharmacol. 2016 Jul;82(1):17-29. doi: 10.1111/bcp.12944. Epub 2016 May 9.

34. Ghitza UE, Gore-Langton RE, Lindblad R, Tai B. NIDA Clinical Trials Network Common Data Elements Initiative: Advancing Big-Data Addictive-Disorders Research. Front Psychiatry. 2015 Mar 3;6:33. doi: 10.3389/fpsyt.2015.00033.

35. Krucoff MW, Sedrakyan A, Normand SL. Bridging Unmet Medical Device Ecosystem Needs With Strategically Coordinated Registries Networks. JAMA. 2015 Oct 27;314(16):1691-2. doi: 10.1001/jama.2015.11036.

36. Leavy MB. Multinational Registries: Challenges and Opportunities. Addendum to Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition. Agency for Healthcare Research and Quality, Rockville, MD. 2017.

37. Sharp J. Registries in Accountable Care. Addendum to Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition. Agency for Healthcare Research and Quality, Rockville, MD. 2017.

38. https://cdn.ymaws.com/www.thepcpi.org/resource/resmgr/2017maturational-framework-p.pdf. Accessed November 1, 2018.

39. https://www.healthit.gov/sites/default/files/tefca2017oncannualmeetingfinal.pdf. Accessed November 1, 2018.

40. Mack C, Su ZH, Westreich D. Managing Missing Data in Patient Registries. Addendum to Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition. Agency for Healthcare Research and Quality, Rockville, MD. 2017.

41. Lin M, Vreeman DJ, Huff SM. Investigating the semantic interoperability of laboratory data exchanged using LOINC codes in three large institutions. AMIA Annu Symp Proc. 2011;2011:805-14. Epub 2011 Oct 22.

42. Logical Observations Identifiers Names and Codes (LOINC) Recognition (2015): https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfstandards/detail.cfm?id=32889. Accessed November 1, 2018.

41. Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) Recognition (2015): https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfstandards/detail.cfm?id=33224. Accessed November 1, 2018.

44. LIVD – Digital Format for Publication of LOINC to Vendor IVD Test Results (2017): http://ivdconnectivity.org/livd/. Accessed November 1, 2018.

45. FHIR LIVD Implementation Guide v0.1.0 (2018): http://hl7.org/fhir/uv/livd/2018Sep/general.html. Accessed November 1, 2018.

46. Logical Observation Identifiers Names and Codes for In Vitro Diagnostic Tests (2018): https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM610636.pdf. Accessed November 1, 2018.

47. http://www.ahrmm.org/resources/learning-udi-community/index.shtml. Accessed November 1, 2018.

48. Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. Health Aff (Millwood). 2005 Jan-Jun; Suppl Web Exclusives: W5-10-W5-18.

49. Brailer DJ. Interoperability: the key to the future health care system. Health Aff (Millwood). 2005 Jan-Jun; Suppl Web Exclusives:W5-19-W5-21.

50. HIMSS Interoperability and Health Information Exchange. https://www.himss.org/library/interoperability-health-information-exchange. Accessed November 1, 2018.

51. CrowdFlower. 2016 Data Science Report. https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. Accessed November 1, 2018.

52. Jones WS, Krucoff MW, Morales P, et al. Registry Assessment of Peripheral Interventional Devices (RAPID): Registry assessment of peripheral interventional devices core data elements. J Vasc Surg. 2018 Feb;67(2):637-644.e30. doi: 10.1016/j.jvs.2017.07.141.

53. HL7 Clinical Information Modeling Initiative. http://www.hl7.org/Special/Committees/cimi/index.cfm

54. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18.

55. Staples® Easy Button™. https://www.staples.com/Staples-Easy-Button/product_606396. Accessed November 1, 2018.