ADAPTABLE Supplement Report: Patient-Reported Health Data and Metadata Standards in the ADAPTABLE Study

Author: Lexie (Zidanyue) Yang zidanyue.yang@duke.edu
Advisor: Jessie Tenenbaum, PhD jessie.tenenbaum@duke.edu

Date: 7/19/18

This work was supported by a cooperative agreement (U54 AT007748) from the NIH Common Fund for the NIH Health Care Systems Research Collaboratory. The views presented here are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Table of Contents

Introduction

This document summarizes the activity and findings around data standards and metadata standards for the ADAPTABLE supplement project ("the Supplement") exploring patient-reported health data in the ADAPTABLE study .

One aspect of the Supplement includes storing patient-reported data collected during the trial and comparing that data to electronic health record (EHR) data mapped to the PCORnet Common Data Model (CDM). To better understand the patient-reported health (PRH) data elements that the ADAPTABLE Supplement aimed to collect, various sources were explored to:

1. Review previous work related to the data standards and metadata standards for PRH data elements.
2. Compare findings from review of previous work to the data standards recommended in the Interoperability Standards Advisory (ISA).
3. Compare the data elements in the ADAPTABLE Supplement data dictionary to existing LOINC terms in order to evaluate the coverage of ADAPTABLE concepts in LOINC.
4. Submit request for new LOINC codes for the data elements in the ADAPTABLE Supplement data dictionary.

The results and detailed reports for each aim are presented below.

Definitions

Data standards: The rules by which data are described and recorded in order to share, exchange, and understand data.

Metadata standards: The rules by which information *about* data is recorded in order to facilitate understanding of the origin, derivation, and/or provenance of the data.

For both types of standards, there are rules and heuristics related to the format, meaning, and/or minimum set of data elements to include. These are reviewed in the next section along with our findings.

1. Data Standards Literature Review Summary

The purpose of this informal literature review was to identify prior work related to data standards for variables of interest for the ADAPTABLE Supplement, and which data standards and variables were used. The variables of interest are the data elements in the primary, safety, and secondary endpoint components listed below.

Primary Endpoint Components

- Hospitalization for nonfatal MI
- Hospitalization for nonfatal stroke

Safety

- Major bleeding with an associated blood product transfusion

Secondary

- Coronary revascularization procedures (PCI and CABG)
- Race
- Ethnicity
- Smoking status

We searched PubMed and google scholar for two article types: clinical trials and review articles. Although we did not explicitly limit the time frame, most articles identified were published between 2010 and 2016, with one exception published in 1995. We searched various combinations of the following terms: “data standard”, “patient-reported”, “loinc”, “phenx”, “health events”, “questionnaire”, “stroke”, “hospitalization”, “rehospitalization”, “heart failure”, “self-reported”, “bleed”, “blood transfusion”, “CABG”, “coronary artery bypass graft”, “PCI”, “MI”, “myocardial infarction”, “percutaneous coronary intervention”.

Summary of findings

The initial search terms yielded a predictably large number of publications. The results are sorted by relevance using the function of the website. Then first 50 publications under each search term were scanned by titles and contents for relevance. After eliminating non-pertinent hits (e.g. The association between patient-reported incidents in hospitals and estimated rates of patient harm; The Rationale for Collecting Patient-Reported Symptoms during Routine Chemotherapy), we reviewed 29 publications related to data standards for patient-reported data elements, especially for data elements in the primary, safety and secondary endpoints listed above. For primary, safety and secondary endpoints, in 11 publications found to be relevant to patient-reported outcomes data standards were not explicitly addressed. In the remaining 18 publications, we were able to identify relevant data standards for race, ethnicity and smoking status, primarily through the PhenX (Phenotypes and eXposures) Toolkit.

Report

The following is a detailed list of findings followed by the list of articles.

Primary Endpoint Components

- Hospitalization for nonfatal MI and Hospitalization for nonfatal stroke

  2 articles were found relevant to patient-reported outcomes about nonfatal stroke. 7 articles were found relevant to self-reported disease history, hospitalization or re-hospitalization that mentioned both nonfatal MI and nonfatal stroke. However, none of the

articles mentioned which data standards or variables were used for capture of these patient-reported elements.

Table 1. Publications relevant to data standards for primary endpoints

Instructive	Title	Journal	First author	Year published
No	Development and validation of a patient-reported outcome measure for stroke patients	Health Qual Life Outcomes	Yanhong Luo	2015
No	Functional status and patient-reported outcome 10 years after stroke: the Lund Stroke Register.	Stroke	Jönsson AC	2014
No	How Reliable are Patient-Reported Rehospitalizations? Implications for the Design of Future Practical Clinical Studies	J Am Heart Assoc	Arun Krishnamoorthy	2016
No	Reduced functionality in everyday activities of patients with self-reported heart failure hospitalization — Population-based study results	Int J Cardiol	Skalska A	2014
No	How does self-reported history of stroke compare to hospitalization data in a population-based survey in New Zealand?	Stroke	Carter K	2010
No	Long-term association between self-reported signs and symptoms and heart failure hospitalizations: the Atherosclerosis Risk In Communities (ARIC) Study.	Eur J Heart Fail	Avery CL	2010
No	Seizure-related injuries and hospitalizations: self-report data from the 2010 Australian Epilepsy Longitudinal Survey.	Epilepsy Behav	Bellon M	2013
No	Validity of self-reported hospital admissions in clinical trials depends on recall period length and individual characteristics.	J Eval Clin Pract	Seidl H	2016
No	Patient-Reported Outcomes in Heart Failure: Existing Measures and Future Uses	Curr Heart Fail Rep	Thompson LE	2015

Safety Endpoints Components

- Major bleeding with an associated blood product transfusion

  2 articles were found relevant to bleeding and blood transfusion. However their focus on haemophilia make them somewhat less relevant for our purposes. No names of data standards are mentioned.

Table 2. Publications relevant to data standards for safety-related data elements

Useful	Title	Journal	First author	Year published
No	Patient-reported experience of bleeding events in haemophilia	Eur J Haematol	Emuella Flood	2014
No	Patient-reported outcomes of 182 adults with severe haemophilia in Germany comparing prophylactic vs. on-demand replacement therapy	Haemophilia	Mondorf W	2013

Secondary Endpoints Components

- Coronary revascularization procedures (PCI and CABG)

  No relevant articles were identified. Searching combination of "patient-reported" and "PCI/CABG" (and the full names) in PubMed and Google Scholar returns publications regarding patient-reported outcomes or patient-reported quality of life **after** PCI/CABG, which are not the variables of interest for our purposes. For example, the article "Comparison of patient-reported outcomes after elective coronary artery bypass grafting in patients aged ≥ and <65 years[1]".

- Race, ethnicity, and smoking status

  We found 18 articles relevant to data standards used for patient-reported outcomes involving race, ethnicity and smoking status included. Several articles (marked with ***) used PhenX/LOINC for patient-reported race, ethnicity and smoking status. The variable names they used are also listed in these articles as follows:

Table 3. Standardized elements identified for demographics and smoking status

Data elements	Description of variable	Phenx toolkit ID
Race	Race	010601
Ethnicity	Ethnicity	010501
Smoking status	Tobacco – smoking status (adult protocol)	030602
	Tobacco - smoking status	030601

Below are the data elements as they appear in PhenX (https://www.phenxtoolkit.org/):

Race-010601

SELECT ONE OR MORE CATEGORIES

[ ] 10 WHITE

[ ] 11 BLACK/AFRICAN AMERICAN

[ ] 12 INDIAN (AMERICAN)

[ ] 13 ALASKA NATIVE

[ ] 14 NATIVE HAWAIIAN

[ ] 15 GUAMANIAN

[ ] 16 SAMOAN

[ ] 17 OTHER PACIFIC ISLANDER (SPECIFY)

[ ] 18 ASIAN INDIAN

[ ] 19 CHINESE

[ ] 20 FILIPINO

[ ] 21 JAPANESE

[ ] 22 KOREAN

[ ] 23 VIETNAMESE

[ ] 24 OTHER ASIAN (SPECIFY)

[ ] 25 SOME OTHER RACE (SPECIFY)___

[ ] 77 REFUSED

[ ] 99 DON'T KNOW

Ethnicity-010501

Do you consider yourself Hispanic/Latino? [Where did your ancestors come from?]

[ ] 1 YES [ask follow-up question]
[ ] 2 NO
[ ] 7 REFUSED
[ ] 9 DON'T KNOW

Please give me the number of the group that represents your Hispanic origin or ancestry. Please select 1 or more of these categories.

[ ] 10 PUERTO RICAN
[ ] 12 DOMINICAN (REPUBLIC)
[ ] 13 MEXICAN/MEXICANO
[ ] 14 MEXICAN AMERICAN
[ ] 15 CHICANO
[ ] 18 CUBAN
[ ] 19 CUBAN AMERICAN
[ ] 20 CENTRAL OR SOUTH AMERICAN
[ ] 40 OTHER LATIN AMERICAN
[ ] 41 OTHER HISPANIC
[ ] 77 REFUSED
[ ] 99 DON'T KNOW

Table 4. Smoking status data elements in PhenX

Variable Name	Variable ID	Variable Description	Version
PX030601_Cigarette_Smoking	PX030601010000	Have you ever smoked part or all of a cigarette?	4
PX030601_Cigarette_Smoking_100	PX030601020000	Have you smoked at least 100 cigarettes in your entire life?	4
PX030601_Cigarette_Smoking_Current	PX030601030000	Do you now smoke cigarettes every day, some days, or not at all?	4
PX030602_Cigarette_Smoking_100	PX030602010000	Have you smoked at least 100 cigarettes in your entire life? (Note to interviewer: 100 CIGARETTES = APPROXIMATELY 5 PACKS)	4
PX030602_Cigarette_Smoking_Current	PX030602020000	Do you now smoke cigarettes every day, some days, or not at all?	4
PX030602_Cigarette_Smoking_Everyday_6Month	PX030602030000	Has there ever been a period in your life when you smoked cigarettes every day for at least 6 months?	4

Table 5. Publications identified relevant to data standards for race, ethnicity, and smoking status.

Useful	Title	Journal	First author	Year published
No	Standardized Cardiovascular Data for Clinical Research, Registries, and Patient Care	J Am Coll Cardiol	H. Vernon Anderson	2013
Yes	C3-PRO: Connecting ResearchKit to the Health System Using i2b2 and FHIR	PLoS One	Pascal B. Pfiffner	2016
Yes	A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research	J Am Med Inform Assoc	Rémy Choquet	2015
No	Health care delivery reorganization innovative outcome: universal computerized patient identification.	Medinfo	Murray CL	1995
Yes	On the uses of routine patient-reported health outcome data.	Health Econ	Smith PC	2013

Useful	Title	Journal	First author	Year published
Yes	Integrating the Use of Patient-Reported Outcomes for Both Clinical Practice and Performance Measurement: Views of Experts from 3 Countries	Milbank Q	Van Der Wees PJ	2014
Yes	Standards for Patient-Reported Outcome–Based Performance Measures	JAMA	Basch E	2013
Yes	Using PhenX toolkit measures and other tools to assess urban/rural differences in health behaviors: recruitment methods and outcomes	BMC Res Notes	Michael M Hitz	2014
Yes	How to assess common somatic symptoms in large-scale studies: A systematic review of questionnaires	J Psychosom Res	Zijlema WL	2013
No	Approach for Classification and Severity Grading of Long-term and Late-Onset Health Events among Childhood Cancer Survivors in the St. Jude Lifetime Cohort	Cancer Epidemiol Biomarkers Prev	Hudson MM	2017
No	Development of the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE)	Journal of the National Cancer Institute	Ethan Basch	2014
No	The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review	Supportive Care in Cancer	Thomas M. Atkinson	2016
Yes	Psychological Assessment Instruments: A Coverage Analysis Using SNOMED CT, LOINC and QS Terminology	AMIA Annu Symp Proc	Piper A. Ranallo	2013
Yes	Semantic Interoperability of Health Risk Assessments	AMIA Annu Symp Proc	Jay Rajda	2011
Yes	PhenX RISING: real world implementation and sharing of PhenX measures	BMC Med Genomics	McCarty CA	2014
Yes	Validation of PhenX measures in the personalized medicine research project for use in gene/environment studies	BMC Med Genomics	Catherine A McCarty	2014
Yes	Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield Personalized Medicine Research Project Biobank.	Pac Symp Biocomput	MOLLY A. HALL	2015

Useful	Title	Journal	First author	Year published
No	Physical activity and physical fitness: standardizing assessment with the PhenX Toolkit.	Am J Prev Med	William L. Haskell	2013

2. Metadata Standard Literature Review Summary

The purpose of this informal metadata standards literature review is to identify prior work related to metadata standards related to patient-reported data. In this literature review we aimed to identify previous publications about metadata standards that are being used for common metadata elements such as: title, proxy vs. self-completion, PRO version, location of administration, mode of administration, free text vs. multiple choice, date of measure, number of items, topic, language, etc.

We searched PubMed and PMC with combinations of keywords: “metadata”, “metadata standards”, “metadata standards AND PRO”, “metadata standards AND patient-reported”, etc. Though results were sparse, we identified 11 publications related to our goal. The most frequently used metadata standard in these articles is ISO11179 (http://metadata-standards.org/11179/). Additionally, the Dublin Core metadata standard (http://dublincore.org/documents/dces/), FDGC (Federal Geographic Data Committee- https://www.fgdc.gov/metadata/geospatial-metadata-standards), EPHT (Environmental Public Health Tracking)- https://nmtracking.org/epht-view/dataportal/metadata/Introduction.html) are also mentioned. In a paper about metadata schema in China, standards of the Clinical Data Interchange Standards Consortium (CDISC) is used.

Summary of findings

1. **Title**: ISO 11179 has a variable “title” of a reference document. A Reference document is a document that provides pertinent details for consultation about a subject. The Dublin Core has a variable “title,” which is a name given to the resource.
2. **Proxy vs self-completion**: Not included in any metadata standards from literature review.
3. **PRO Version**: ISO 11179 has a variable “version”.
4. **Location of administration and mode of administration:** ISO 11179 does not have location variable, but has mail address/postal address, not specifically for administration. The Dublin Core has variable “coverage”, which can include a spatial location. ISO 11179 has variable “administration information”, which is information about an administrated item in a metadata registry and can include creation date, last change date, origin, change description, explanatory comment, etc.
5. **Free text vs multiple choices:** Not identified from literature review.
6. **Date of measure:** Various kinds of date variables in ISO 11179, including date, datetime, creation date, last change date, etc. The Dublin Core also has date variable.

7. **Number of items:** Not identified from literature review. But ISO 11179 has a concept called representation class which categorizes data elements such as “quantity”: A continuous number such as the linear dimensions, capacity/amount (non-monetary) of an object.
8. **Topic:** The Dublin Core has variable “subject,” which describes a topic.
9. **Language:** Various kinds of variables in ISO 11179, including language, definition language, designation language, etc. The Dublin Core also has variable “language”.

In addition to the metadata elements in the original list, the Dublin Core also has variable “contributor” and “creator” (e.g. who made the ADAPTABLE questionnaires), and “format”. These might be appropriate elements to add to the metadata list.

Report

The data elements of each data standard mentioned in the literatures are listed below.

Table 6. ISO 11179 metadata elements

Name	Permissible values	Permissible values
Identifier	Related data reference	Maximum size
Version	Type of relationship	Minimum size
Context	Data type	Keywords
Classification scheme		

Table. ISO 11179 Metadata Elements

Table 7. Dublin Core metadata elements

Coverage	Subject	Rights
Description	Title	Date
Type	Contributor	Format
Relation	Creator	Identifier
Source	Publisher	Language

Table. Dublin Core Metadata Elements

Table 8. FGDC metadata elements

Data resource title	Contact information	Status
Attributes	Purpose	Citation
Spatial domain	Distribution	Access constraints
Time period of data	Resource	Keywords
Process steps		

Table. An example of the FGDC Metadata Elements

Table 9. EPHT metadata profile elements

Element	Definition
Identification elements	
Citation: Information to be used to reference the dataset	
Originator	The name of an organization or an individual who developed the dataset. If the names of editors or compilers are provided, the name must be followed by "(ed.)" or "(comp.)," respectively
Publication date	The date when the dataset is published or otherwise made available for release
Title	The name by which the dataset is known
URL	The name of an on-line computer resource that contains the dataset. Entries should follow the uniform resource locator convention of the Internet (complete if applicable)
Description: A characterization of the dataset, including its intended use and limitations	
Abstract	A brief narrative summary of the dataset
Purpose	A summary of the intentions with which the dataset was developed
Element	**Definition**
Metadata access constraints	Restrictions and legal prerequisites for accessing the metadata. These include any access constraints applied to ensure the protection of privacy or intellectual property and any special restrictions or limitations on obtaining the metadata
Metadata use constraints	Restrictions and legal prerequisites for using the metadata after access are granted. These include any metadata use constraints applied to ensure the protection of privacy or intellectual property and any special restrictions or limitations on using the metadata
Time period information	
Single date: Means of encoding a single date and time	
Calendar date	The year (optionally month or month and day)
Time of day	The hour (and optionally minute, or minute and second) of day
Range of dates: Means of encoding a range of dates	
Beginning date	The first year (optionally month or month and day) of the event
Beginning time	The first hour (and optionally minute, or minute and second) of the event
Ending date	The last year (and optionally month or month and day) for the event
Ending time	The first hour (and optionally minute, or minute and second) of the event
Contact information	
Contact information: This section provides a means of identifying individuals and organizations and is used by other sections of the metadata standard	
Contact organization	The name of the organization
Contact position	Title of the individual (complete if applicable)
Contact address: The address for the organization	
Address type	Address type
Address	Contact address for organization
City	Contact address city
State or province	Contact address state or province
Postal code	Contact address, ZIP or postal code
Country	Contact address country
Contact telephone number	The telephone number by which individuals can speak to the organization.
Contact TDD/TTY telephone	The telephone number by which hearing-impaired individuals can contact the organization (complete if applicable)
Contact fax number	The telephone number of a facsimile machine of the organization (complete if applicable)
Contact e-mail address	The address of the electronic mailbox of the organization (complete if applicable)
Hours of service	Time period when individuals can speak to the organization. (complete if applicable)
Contact instructions	Supplemental instructions on how or when to contact the organization (complete if applicable)

Table 9. Continued

Element	Definition
Supplemental info	Other descriptive information about the dataset (complete if applicable)
Access constraints	Restrictions and legal prerequisites for accessing the dataset. These include any access constraints applied to ensure the protection of privacy or intellectual property and any special restrictions or limitations on obtaining the dataset.
Use constraints	Restrictions and legal prerequisites for using the dataset after access are granted. These include any use constraints applied to ensure the protection of privacy or intellectual property and any special restrictions or limitations on using the dataset.
Native dataset environment	A description of the dataset, including the name of the software, computer operating system, file name, and dataset size
Time period of content: Time period for which the dataset corresponds to the currentness reference	
Currentness	The basis on which the time period of content information is determined
Status: The state of and maintenance information for the dataset	
Progress	The state of a dataset
Maintenance and update frequency	The frequency that changes are made to the dataset after the initial dataset is completed
Spatial domain: The geographic area covered by the dataset	
West bounding coordinate	Western-most coordinate of the limit of coverage expressed in longitude
East bounding coordinate	Eastern-most coordinate of the limit of coverage expressed in longitude
North bounding coordinate	Northern-most coordinate of the limit of coverage expressed in latitude
South bounding coordinate	Southern-most coordinate of the limit of coverage expressed in latitude
Keywords: Words or phrases summarizing an aspect of the dataset	
Theme keyword thesaurus	Reference to a formally registered thesaurus or a similar authoritative source of theme keywords
Theme keyword	Common use word or phrase used to describe the subject of the dataset
Place keyword thesaurus	Reference to a formally registered thesaurus or a similar authoritative source of place keywords
Place keyword	The geographic name of a location covered by a dataset (includes city, county, state, state acronym, regional descriptions and references)
Security information: Handling restrictions imposed on the dataset because of national security, privacy, or other concerns	
Security classification system	Name of the classification system
Security classification	Name of the handling restrictions on the dataset
Security handling description	Additional information about the restrictions on handling
Data quality information	
Logical consistency report	An explanation of the fidelity of relationships in the dataset and tests used
Completeness report	Information about omissions, selection criteria, generalization, definitions used, and other rules used to derive the dataset
Process step: Information about a single event	
Process description	An explanation of the event and related parameters or tolerances
Process date	The date when the event was completed
Entity and attribute	
Overview description: Description of the entities, attributes, attribute values, and related characteristics encoded	
Entity and attribute overview	Detailed summary of the information contained in a dataset
Entity and attribute detail citation	Reference used to the complete description of the entity types, attributes, and attribute values for the dataset
Distribution information	
Distributor: The party from whom the dataset may be obtained	
Resource description	The identifier by which the distributor knows the dataset
Distribution liability	Statement of the liability assumed by the distributor
Custom order process	Description of custom distribution services available and the terms and conditions for obtaining these services
Metadata reference	
Metadata date	The date that the metadata were created or last updated
Metadata standard name	The name of the metadata standard used to document the dataset

Figure 1. An open metadata schema for prospective clinical research (openPCR) in China

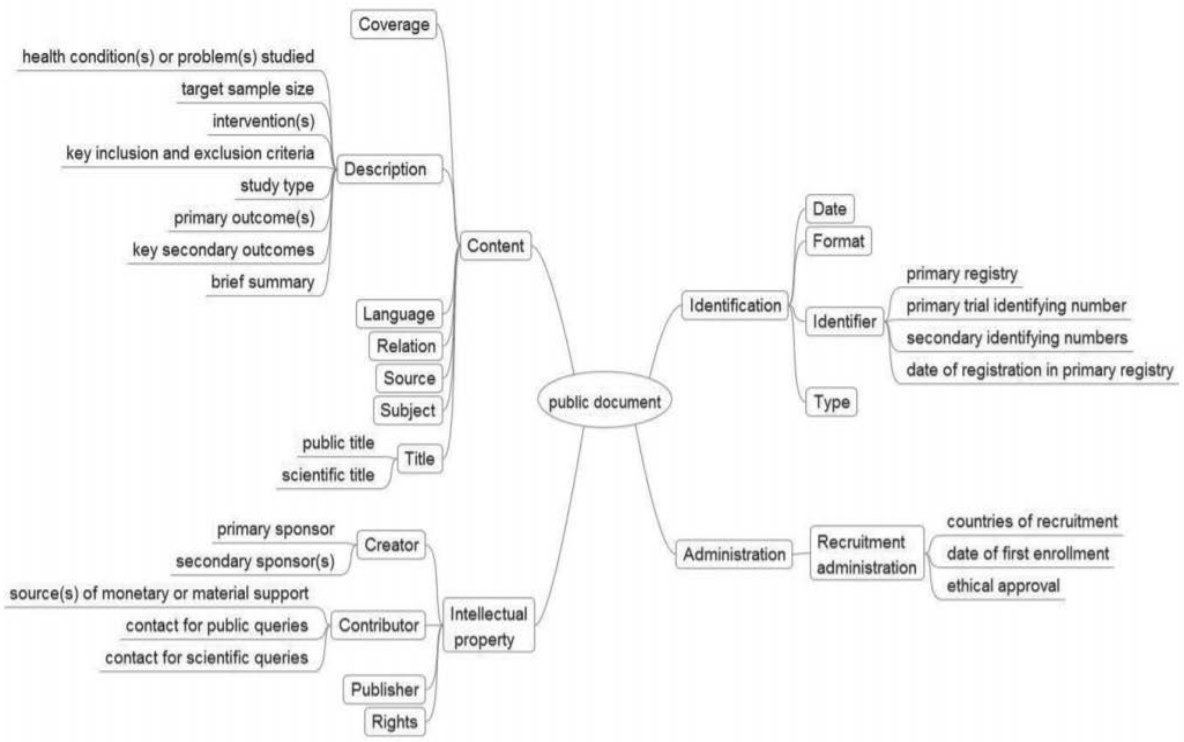

Table 10. Publications identified relevant to metadata standards

Title	Author	Year published	Metadata standard mentioned
CCR+: Metadata Based Extended Personal Health Record Data Model Interoperable with the ASTM CCR Standard	Yu Rang Park, PhD	2014	International Organization for Standardization. Information technology: metadata registries (MDR). Part 3. Registry metamodel and basic attributes.
Assessing Metadata Quality of a Federally Sponsored Health Data Repository	David T. Marc, PhD	2016	Dublin Core Metadata Initiative
Establishment of Kawasaki disease database based on metadata standard	Yu Rang Park	2016	ISO/IEC 11179
A metadata schema for data objects in clinical research	Steve Canham	2016	ISO 8601, CDISC Foundational Standards
Metadata for HIM: ISO standards for global interoperability.	Lisa Spellman	2012	ISO/IEC 11179, ISO 15000-3
caCORE: A common infrastructure for cancer informatics	Peter A. Covitz	2003	ISO11179

Title	Author	Year published	Metadata standard mentioned
Achieving interoperability for metadata registries using comparative object modeling	Yu Rang Park	2010	
Establishing Semantic Interoperability of Biomedical Metadata Registries using Extended Semantic Relationships	Yu Rang Park	2013	
Describing Environmental Public Health Data: Implementing a Descriptive Metadata Standard on the Environmental Public Health Tracking Network	Patridge, Jeff	2008	Dublin Core, ISO11179, FGDC, EPHT
EBM Metadata Based on Dublin Core Better Presenting Validity of Clinical Trials	Wei Xu	2007	Dublin Core
Development of an Open Metadata Schema for Prospective Clinical Research (openPCR) in China	W Xu	2014	Standards of the Clinical Data Interchange Standards Consortium (CDISC)

3. The Interoperability Standards Advisory (ISA) Findings

The Interoperability Standards Advisory (ISA) provides the healthcare industry with a single list of data standards and their implementation specifications to address specific interoperability needs of healthcare information. ISA has broad categories for different types of healthcare information:

1. Section I: Vocabulary/Code Set/Terminology Standards and Implementation Specifications
2. Section II: Content/Structure Standards and Implementation Specifications
3. Section III: Standards and Implementation Specifications for Services (i.e., the infrastructure components deployed and used to address specific interoperability needs)
4. Section IV: Models and Profiles
5. Section V: Administrative Standards and Implementation Specifications (i.e., payment, operations and other "non-clinical" interoperability needs)

As a supplement to the previous data standard literature review, data elements of our interest for the ADAPTABLE Supplement were explored in the ISA and compared to what was summarized in the literature review.

Summary of findings

1. The coronary revascularization procedures concepts of CABG and PCI may be represented using standardized terms from SNOMED CT.
2. ISA suggests OMB standard or the CDC code set for race in EHRs, however PhenX is a well-accepted and NIH-supported standard and we believe it is more appropriate for these research purposes.
3. For smoking status, ISA recommends LOINC and SNOMED CT vs. PhenX, which was observed in the literature. For purposes of this study with the goal of calculating concordance, the variable D_SMOKE from the CDM was mapped to a binary variable of "Current smoker: Yes/No."
4. No useful standardized concepts were found to represent "Major bleeding with an associated blood product transfusion" nor "Hospitalization for nonfatal MI and Hospitalization for nonfatal stroke." We do not believe there is much to be gained for the purposes of this study by combining existing atomic concepts to represent these more complex items.

Report

Primary Endpoint Components

- Hospitalization for nonfatal MI and Hospitalization for nonfatal stroke

https://www.healthit.gov/isa/sending-a-notification-a-patients-admission-discharge-andor-transfer-status-other-providers

Under Section II is found "Patients' Admission, Discharge and/or Transfer Status to Other Providers". HL7 ADT (admission, discharge, transfer) messages are recommended. However, reasons for admission is not required in the ADT messages. This standard is therefore not appropriate to represent the concepts of "Hospitalization for nonfatal MI" and "Hospitalization for nonfatal stroke."

Safety Endpoint Components

- Major bleeding with an associated blood product transfusion

  The closest relevant concept to "major bleeding with an associated blood project transfusion" in ISA is "Representing Patient Medical Encounter Diagnosis" under section I-B. The recommended standard for this type is SNOMED CT.

  SNOMED includes the terms for "Transfusion" (SNOMED ID: 5447007) and "Bleeding" (SNOMED ID 131148009). The latter is a child of the top level concept "Clinical Finding" and has dozens of child concepts.

  Neither of these individual terms appropriately represents the original intended meaning.

Secondary Endpoint Components

- Coronary revascularization procedures (PCI and CABG)

Representing Medical Procedures Performed

Type	Standard Implementation/Specification	Standards Process Maturity	Implementation Maturity	Adoption Level	Federally required	Cost	Test Tool Availability
Standard	SNOMED CT®	Final	Production	●●●●●	Yes	Free	N/A
Standard	CPT-4	Final	Production	●●●●●	Yes	$	N/A
Standard	HCPCS	Final	Production	●●●●●	Yes	Free	N/A
Standard	ICD-10-PCS	Final	Production	●●●●○	Yes	Free	N/A

https://www.healthit.gov/isa/representing-medical-procedures-performed

Elements in safety and secondary endpoints belong to medical procedures. Under "Medical Procedures Performed", four data standards are recommended by the ISA: SNOMED CT, CPT-4, HCPCS, ICD-10-PCS. From ISA:

- ICD-10-PCS is primarily a billing code used only in inpatient settings.
- CPT and HCPCS are codes used to report procedures and services in outpatient procedures.
- ICD-10-PCS is named in the 2015 Edition certification rules as an optional code set for procedures.
- SNOMED CT procedure codes can be used to describe treatment in any clinical setting and is not tied to billing, but can be cross-mapped to corresponding ICD-10-PCS and CPT/HCPCS codes.

ISA does not provide any detailed information about variables in each data standard, but further exploration in SNOMED CT showed the following concepts of interest:

- Percutaneous coronary intervention (SNOMED ID: 415070008, CUI: C1532338, Synonym: PCI)
- Coronary artery bypass graft (SNOMED ID: 232717009, CUI: C0010055, Synonym: CABG)

The latter has a number of child concepts that describe more specific cases; e.g., method or numbers of grafts.

- Race, ethnicity, and smoking status

Representing Patient Race and Ethnicity

Type	Standard Implementation/Specification	Standards Process Maturity	Implementation Maturity	Adoption Level	Federally required	Cost	Test Tool Availability
Standard	OMB standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity, Statistical Policy Directive No. 15, Oct 30, 1997	Final	Production	●●●●○	Yes	Free	N/A
Standard	CDC Race and Ethnicity Code Set Version 1.0	Final	Production	Feedback Requested	Yes	Free	N/A

https://www.healthit.gov/isa/Representing_Patient_Race_and_Ethnicity

For race and ethnicity, ISA recommends OMB standard and CDC code set, while PhenX/LOINC was found in the literature review. From ISA:

- LOINC® provides observation codes for use in the observation / observation value pattern for communicating race and ethnicity.
- The LOINC answers for Race look similar to CDC/HL70005, but don't match; this may be confusing to implementers.

Representing Patient Tobacco Use (Smoking Status)

Type	Standard Implementation/Specification	Standards Process Maturity	Implementation Maturity	Adoption Level	Federally required	Cost	Test Tool Availability
Standard for observations	LOINC®	Final	Production	●●●●●	No	Free	N/A
Standard for observation values	SNOMED CT®	Final	Production	●●●●●	Yes	Free	N/A

https://www.healthit.gov/isa/representing-patient-tobacco-use-smoking-status

For smoking status, ISA recommends LOINC and SNOMED CT, while PhenX was found in the literature review. Although LOINC and PhenX sometimes share common variables, ISA uses 'Tobacco smoking status NHIS' LOINC 72166 which is different from the PhenX variable in the literature review.

ISA recommendations:

LOINC: 72166-2 Tobacco smoking status NHIS:
https://s.details.loinc.org/LOINC/72166-2.html?sections=Comprehensive

From ISA:
ONC's 2015 Edition certification requirements reference the following value set for smoking status. Codes from SNOMED CT®:

1. Current every day smoker. 449868002
2. Current some day smoker. 428041000124106
3. Former smoker. 8517006
4. Never smoker. 266919005
5. Smoker, current status unknown. 77176002
6. Unknown if ever smoked. 266927001
7. Heavy tobacco smoker. 428071000124103
8. Light tobacco smoker. 428061000124105

4. Existing LOINC Terms

There are 48 data elements in total in the ADAPTABLE Supplement data dictionary. We searched for existing LOINC terms that could be used for all 48 data elements through the Search LOINC website (http://search.loinc.org/searchLOINC/search.zul).

Summary of findings

1. 9 out of 48 data elements in the ADAPTABLE Supplement can be matched to existing LOINC terms with the same variable description.
2. 15 out of 48 data elements in the ADAPTABLE Supplement have variable descriptions similar to existing LOINC terms. However, the questions being asked or answer lists differ from the existing LOINC terms.
3. 24 out of 48 data elements in the ADAPTABLE Supplement cannot be matched to existing LOINC terms. These data elements have questions and answer lists that are unique to the ADAPTABLE study.

Report

Further details can be found in LOINC submission spreadsheet.

Table 11. ADAPTABLE data elements and LOINC terms

ADAPTABLE data elements	Existing LOINC terms and similarity to ADAPTABLE data elements
Patient's self-assessment on ability to run errands and shop.	61635-9, same
Patient's self-assessment on depression in the past week.	61967-6, same
Patient's self-assessment on fatigue in the past week.	61878-5, same
Patient's self-assessment on sleep disturbance in the past week.	61998-1, same
Patient's self-assessment on having trouble doing leisure activities with others in the past week.	75417-6, same
Patient's self-assessment on pain interference in the past week.	61758-9, same
Name	54503-8, same
Gender	46098-0, same
Date of Birth	21112-8, same
Patient's answer to the question: are you currently pregnant/nursing?	66174-4, only pregnant, no nursing
Patient's self-assessment on health condition.	64438-5, same question, different answers
Patient's home/mailing address.	63728-0, similar
Patient's best contact phone number.	65651-2, 71724-9, 71751-2, 71723-1, similar
Patient's last four digits of social security number.	45396-9, similar
Patient's secondary contact's name.	54503-8, similar
Patient's secondary contact's phone number.	65651-2, 71724-9, 71751-2, 71723-1, similar
Patient's secondary contact's email address.	76458-9, similar
Patient's race.	32624-9, same question, different answer
Patient's answer to the question: is your ethnicity Hispanic?	56051-6, similar
Patient's answer to the question: are you covered through Medicare?	69431-5, similar
Patient's answer to the question: are you currently a cigarette smoker (at least one cigarette a day)?	63582-1, 64234-8, similar
Patient's answer to the question: before joining ADAPTABLE were you regularly taking aspirin?	67450-7, similar
Patient's answer to the question: now that you are a part of ADAPTABLE are you regularly taking aspirin?	67450-7, similar
Patient's answer to the question: why did you stop taking aspirin?	63950-0, similar, different med

Recommendations

Primary Endpoint Components

- Hospitalization for nonfatal MI—no extant standard; submitted to LOINC
- Hospitalization for nonfatal stroke—no extant standard; submitted to LOINC

Safety

- Major bleeding with an associated blood product transfusion—no extant standard; submitted to LOINC

Secondary

- Coronary revascularization procedures (PCI and CABG): SNOMED CT
- Race: Phenx
- Ethnicity: Phenx

Metadata elements

1. Title: the Dublin Core
2. Proxy vs. self-completion—no extant standard
3. PRO version: ISO 11179
4. Location of Administration: ISO 11179 or the Dublin Core
5. Mode of Administration: ISO 11179
6. Free text vs. multiple choice—no extant standard
7. Date of Measure: ISO 11179 or the Dublin Core
8. Number of items: ISO 11179
9. Topic: the Dublin Core
10. Language: ISO 11179 or the Dublin Core
11. Contributor: the Dublin Core
12. Creator: the Dublin Core
13. Format: the Dublin Core

LOINC Submission

ADAPTABLE PRH data elements were submitted to LOINC on 19 March, 2018, using LOINC's spreadsheet-based submission template. A pre-release report was reviewed with LOINC staff on May 15 and will be included in the June release with "Trial" status. All 48 data elements were included in the submission spreadsheet regardless their similarity to the existing LOINC terms. For each data element, property, timing, system, scale, answer list, unit and formula were defined according to the LOINC guidelines. A new column was added to the submission template for "Questionnaire," with three categories: enrollment/early visit/early visit, follow-up visits, because in the ADAPTABLE study the data elements are collected at the three different phases. Further details can be found in the LOINC submission spreadsheet.

Conclusion

The literature review of data standards and metadata standards of patient-reported outcome data yielded limited results with the primary and safety endpoints elements, but we were able to identify several works that related to data standards of smoking status, race and ethnicity. We were also able to identify several metadata standards that are frequently used in the medical field. We further make recommendations of which data standards to use based on the results from the literature review and from the Interoperability Standards Advisory (ISA) recommendations. New LOINC terms have been submitted for all ADAPTABLE Supplement data elements to standardize ADAPTABLE study variables and to be stored in a central repository for future pragmatic trials to use.