9-11-2016

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

Michael G. Kahn
University of Colorado Anschutz Medical Campus, michael.kahn@ucdenver.edu

Tiffany J. Callahan
University of Colorado Anschutz Medical Campus, tiffany.callahan@ucdenver.edu

Juliana Barnard
University of Colorado Anschutz Medical Campus, juliana.barnard@ucdenver.edu

Alan E. Bauck
Kaiser Permanente Northwest, alan.bauck@kpchr.org

See next pages for additional authors

Recommended Citation

Kahn, Michael G.; Callahan, Tiffany J.; Barnard, Juliana; Bauck, Alan E.; Brown, Jeff; Davidson, Bruce N.; Estiri, Hossein; Goerg, Carsten; Holve, Erin; Johnson, Steven G.; Liaw, Siaw-Teng; Hamilton-Lopez, Marianne; Meeker, Daniella; Ong, Toan C.; Ryan, Patrick; Shang, Ning; Weiskopf, Nicole G.; Weng, Chunhua; Zozus, Meredith N.; and Schilling, Lisa (2016) "A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 4: Iss. 1, Article 18.
DOI: http://dx.doi.org/10.13063/2327-9214.1244
Available at: http://repository.edm-forum.org/egems/vol4/iss1/18

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

Abstract

Objective: Harmonized data quality (DQ) assessment terms, methods, and reporting practices can establish a common understanding of the strengths and limitations of electronic health record (EHR) data for operational analytics, quality improvement, and research. Existing published DQ terms were harmonized to a comprehensive unified terminology with definitions and examples and organized into a conceptual framework to support a common approach to defining whether EHR data is 'fit' for specific uses.

Materials and Methods: DQ publications, informatics and analytics experts, managers of established DQ programs, and operational manuals from several mature EHR-based research networks were reviewed to identify potential DQ terms and categories. Two face-to-face stakeholder meetings were used to vet an initial set of DQ terms and definitions that were grouped into an overall conceptual framework. Feedback received from data producers and users was used to construct a draft set of harmonized DQ terms and categories. Multiple rounds of iterative refinement resulted in a set of terms and organizing framework consisting of DQ categories, subcategories, terms, definitions, and examples. The harmonized terminology and logical framework's inclusiveness was evaluated against ten published DQ terminologies.

Results: Existing DQ terms were harmonized and organized into a framework by defining three DQ categories: (1) Conformance (2) Completeness and (3) Plausibility and two DQ assessment contexts: (1) Verification and (2) Validation. Conformance and Plausibility categories were further divided into subcategories. Each category and subcategory was defined with respect to whether the data may be verified with organizational data, or validated against an accepted gold standard, depending on proposed context and uses. The coverage of the harmonized DQ terminology was validated by successfully aligning to multiple published DQ terminologies.

Discussion: Existing DQ concepts, community input, and expert review informed the development of a distinct set of terms, organized into categories and subcategories. The resulting DQ terms successfully encompassed a wide range of disparate DQ terminologies. Operational definitions were developed to provide guidance for implementing DQ assessment procedures. The resulting structure is an inclusive DQ framework for standardizing DQ assessment and reporting. While our analysis focused on the DQ issues often found in EHR data, the new terminology may be applicable to a wide range of electronic health data such as administrative, research, and patient-reported data.

Conclusion: A consistent, common DQ terminology, organized into a logical framework, is an initial step in enabling data owners and users, patients, and policy makers to evaluate and communicate data quality findings in a well-defined manner with a shared vocabulary. Future work will leverage the framework and terminology to develop reusable data quality assessment and reporting methods.

Acknowledgements

FUNDING STATEMENT Research reported in this publication was funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1303-5581). The statements in this publication are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee.

ACKNOWLEDGEMENTS Members of the EDM Forum staff provided logics support for the face-to-face meetings and webinars illustrated in Figure 1 and in establishing a community web site and wiki for gathering

community input. These efforts were critical to outreach efforts. Members of the monthly Data Quality Collaborative call provided contacts with individuals and organizations to engage in additional outreach efforts. We thank this large group of dedicated individuals who have made data quality a topic of scientific inquiry, relevance, and importance.

Keywords
electronic health records; data use & quality; data completeness

Disciplines
Health Information Technology

Authors
Michael G Kahn, *University of Colorado Anschutz Medical Campus*; Tiffany J Callahan, *University of Colorado Anschutz Medical Campus*; Juliana Barnard, *University of Colorado Anschutz Medical Campus*; Alan E Bauck, *Kaiser Permanente Northwest*; Jeff Brown, *Harvard Pilgrim Health Care Institute*; Bruce N Davidson, *Hoag Memorial Hospital Presbyterian*; Hossein Estiri, *University of Washington, Institute of Translational Health Sciences*; Carsten Goerg, *University of Colorado Anschutz Medical Campus*; Erin Holve, *AcademyHealth*; Steven G Johnson, *University of Minnesota*; Siaw-Teng Liaw, *UNSW School of Public Health & Community Medicine*; Marianne Hamilton-Lopez, *National Academy of Sciences*; Daniella Meeker, *University of Southern California*; Toan C Ong, *University of Colorado Denver*; Patrick Ryan, *Janssen Research and Development*; Ning Shang, *Columbia University*; Nicole G Weiskopf, *Oregon Health & Science University*; Chunhua Weng, *Columbia University*; Meredith N Zozus, *University of Arkansas for Medical Sciences*; Lisa Schilling, *University of Colorado, Denver*.

eGEMs
Generating Evidence & Methods to improve patient outcomes

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

Michael G. Kahn, MD, PhD;[i] Tiffany J. Callahan, MPH;[i] Juliana Barnard, MA;[i] Alan E. Bauck;[ii] Jeff Brown, PhD;[iii] Bruce N. Davidson, PhD;[iv] Hossein Estiri, PhD;[v] Carsten Goerg, PhD;[i] Erin Holve, PhD, MPH, MPP;[vi] Steven G. Johnson, MS;[vii] Siaw-Teng Liaw, MBBS, PhD, FRACGP, FACHI;[viii] Marianne Hamilton-Lopez, PhD, MPA;[ix] Daniella Meeker, PhD;[x] Toan C. Ong, PhD;[xi] Patrick Ryan, PhD;[xii] Ning Shang, PhD;[xiii] Nicole G. Weiskopf, PhD;[xiv] Chunhua Weng, PhD, FACMI;[xiii] Meredith N. Zozus, PhD;[xv] Lisa Schilling, MD[xi]

ABSTRACT

Objective: Harmonized data quality (DQ) assessment terms, methods, and reporting practices can establish a common understanding of the strengths and limitations of electronic health record (EHR) data for operational analytics, quality improvement, and research. Existing published DQ terms were harmonized to a comprehensive unified terminology with definitions and examples and organized into a conceptual framework to support a common approach to defining whether EHR data is 'fit' for specific uses.

Materials and Methods: DQ publications, informatics and analytics experts, managers of established DQ programs, and operational manuals from several mature EHR-based research networks were reviewed to identify potential DQ terms and categories. Two face-to-face stakeholder meetings were used to vet an initial set of DQ terms and definitions that were grouped into an overall conceptual framework. Feedback received from data producers and users was used to construct a draft set of harmonized DQ terms and categories. Multiple rounds of iterative refinement resulted in a set of terms and organizing framework consisting of DQ categories, subcategories, terms, definitions, and examples. The harmonized terminology and logical framework's inclusiveness was evaluated against ten published DQ terminologies.

Results: Existing DQ terms were harmonized and organized into a framework by defining three DQ categories: (1) Conformance (2) Completeness and (3) Plausibility and two DQ assessment contexts: (1) Verification and (2) Validation. Conformance and Plausibility categories were further divided into subcategories. Each category and subcategory was defined with respect to whether the data may be verified with organizational data, or validated against an accepted gold standard, depending on proposed context and uses. The coverage of the harmonized DQ terminology was validated by successfully aligning to multiple published DQ terminologies.

[i]University of Colorado Anschutz Medical Campus, [ii]Kaiser Permanente Northwest, [iii]Harvard Pilgrim Health Care Institute, [iv]Hoag Memorial Hospital Presbyterian, [v]University of Washington, Institute of Translational Health Sciences, [vi]AcademyHealth, [vii]University of Minnesota, [viii]UNSW School of Public Health & Community Medicine, [ix]National Academy of Sciences, [x]University of Southern California, [xi]University of Colorado, Denver, [xii]Janssen Research and Development, [xiii]Columbia University, [xiv]Oregon Health & Science University, [xv]University of Arkansas for Medical Sciences

CONTINUED

Discussion: Existing DQ concepts, community input, and expert review informed the development of a distinct set of terms, organized into categories and subcategories. The resulting DQ terms successfully encompassed a wide range of disparate DQ terminologies. Operational definitions were developed to provide guidance for implementing DQ assessment procedures. The resulting structure is an inclusive DQ framework for standardizing DQ assessment and reporting. While our analysis focused on the DQ issues often found in EHR data, the new terminology may be applicable to a wide range of electronic health data such as administrative, research, and patient-reported data.

Conclusion: A consistent, common DQ terminology, organized into a logical framework, is an initial step in enabling data owners and users, patients, and policy makers to evaluate and communicate data quality findings in a well-defined manner with a shared vocabulary. Future work will leverage the framework and terminology to develop reusable data quality assessment and reporting methods.

Background and Significance

The era of large-scale health data research networks has arrived, along with high expectations that new electronic data sources and analytic methods will answer questions that cannot be examined using traditional controlled clinical trials.[1–5] While large administrative claims databases have long been used for retrospective observational studies, limitations have led to heightened interest in health data from other electronic sources, such as electronic health records (EHRs).[6–11] Studies using EHR data have enabled investigators to examine the impact of diagnostic and therapeutic interventions in diverse real-world clinical settings.[12–18]

Comparative effectiveness studies, patient-centered outcomes research, and pragmatic trials using EHR data captured during routine clinical care from one or more practice settings are becoming an important complement to prospective randomized trials for generating new insights and knowledge. National and international EHR-based clinical research networks are expanding the scope and depth of available data to answer critical questions about care decisions and outcomes important to patients and families.[19–23]

Detailed clinical data are mostly generated by electronic transactions in operational systems that are not primarily intended for research and secondary analysis. Secondary data use refers to the use of data for purposes other than those for which it originally was collected, such as operational, quality improvement, and research analytics. Access to large quantities of clinical data from operational EHRs holds much promise. However, a major concern is that data not collected systematically for research will be poorer quality, which could have negative impacts on findings generated from these data.[24–27] Transaction-oriented systems rarely include prespecified, unambiguous data definitions, and uniform (unbiased) data collection procedures. A substantial body of research suggests that data

Volume 4 (2016) Issue Number 1

collected in EHRs and other operational systems may not be of sufficient quality for research.[28–37] EHRs typically are optimized for efficient patient care and nonclinical administrative requirements, resulting in great variation in clinical documentation practices, even among users of the same systems.[38–40] As clinical data warehouses and large-scale EHR-based data networks become established repositories of electronic health data, consistent methods for describing, assessing, and reporting data quality (DQ) findings could be one way to help secondary data users and consumers understand the potential impact of DQ on reusing data and interpreting findings.[36,41]

The Need for a Harmonized Data Quality (DQ) Terminology

The current DQ literature is inconsistent in the use of terms that describe the complex multidimensional aspects of DQ.[42–51] Inconsistent use of terms to describe DQ features makes it difficult to understand when similar or different DQ features are being discussed. The lack of consistent DQ definitions also makes it difficult to compare DQ results across multiple data-sharing partners. The underlying premise of this work is that standardizing the terms and definitions of DQ concepts, the methods used to evaluate these concepts, and the metrics and formats used to report DQ findings could improve understanding and transparency about the limitations of the data and the results based on these data. The current project focuses on the first challenge—developing a unified DQ terminology and definitions. Our DQ terminology was explicitly scoped to focus on the broad-based evaluation of a large data set typically found in clinical data sharing networks—features that would allow a network to determine the acceptability of data from a contributing partner.

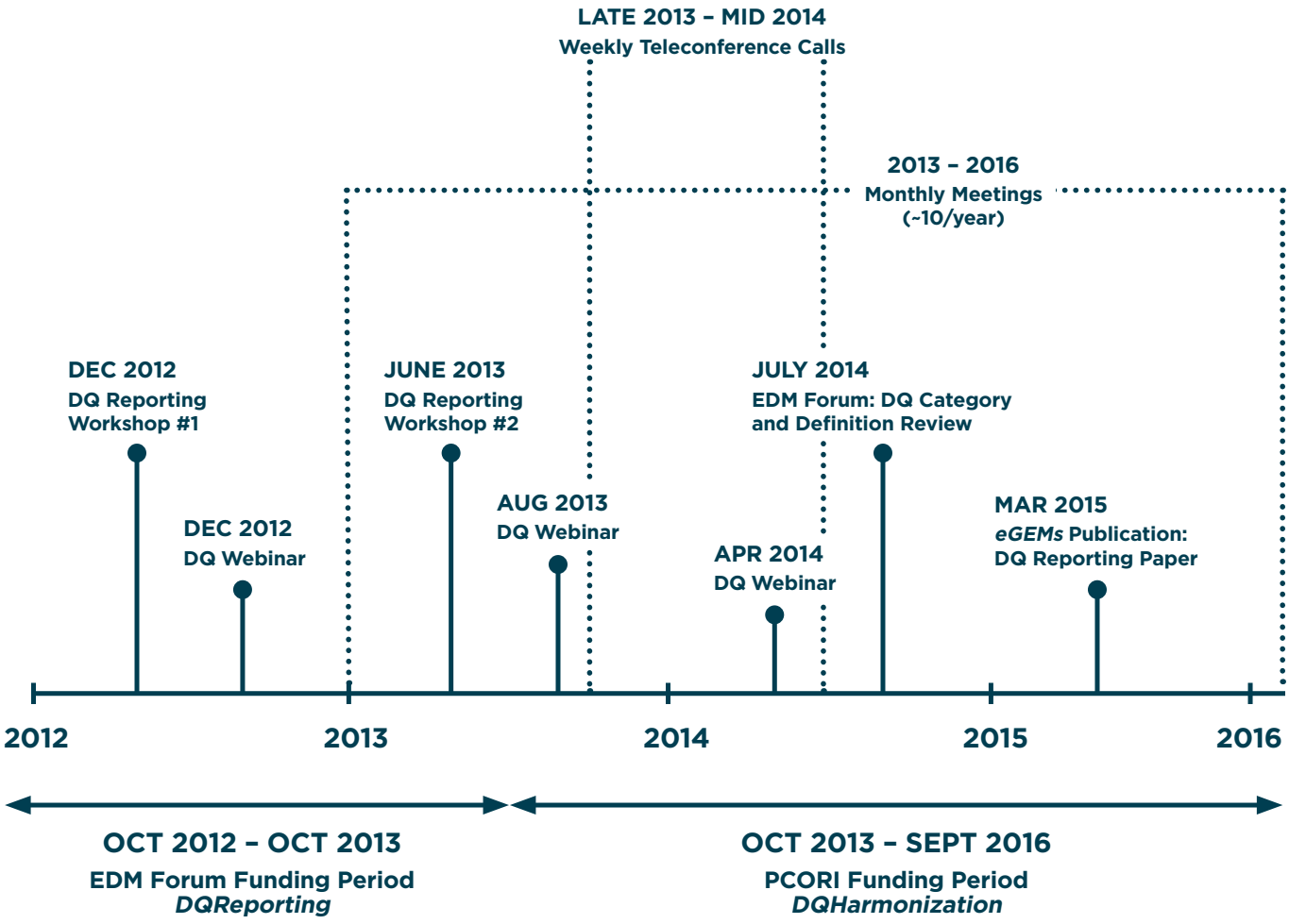
Materials and Methods

Figure 1 depicts a timeline documenting the events that took place as a part of a community-based DQ terminology harmonization effort. Investigators currently engaged in DQ-related work as well as all members of a panel of experts who participated in an earlier effort on DQ reporting sponsored by the Electronic Data Methods (EDM) Forum[52] were enlisted to provide expertise on identifying existing DQ models, terms, and assessment methods.

To obtain a broad representation of DQ terms in current use, we used materials from the previous EDM Forum project,[41] such as the Mini-Sentinel (MS) data characterization routines,[53] and DQ rules embedded in the Observational Medical Outcomes (OMOP) and the Observational Health Data Sciences and Informatics (OHDSI) open-source DQ tools.[54,55] We also included standard operating procedures for DQ assessment used in past or current projects, published "best practices," and DQ publications from both the clinical research and information sciences literature.[36,43,47,56–60] In total, representatives from approximately 20 of the largest United States distributed research networks or large data owners, along with international engagement, either participated directly or were represented in the assessment materials. Collectively, participants represent clinical data networks that contain data on over 540 million patient records. Based on iterative discussions, an initial draft set of DQ terms, categories, and definitions was developed over nine months and was revised as detailed below.

In July 2014, the EDM Forum hosted two one-day workshops to review and critique the draft DQ terms and definitions. One workshop enlisted participants representing patients, patient advocacy groups, and policymakers interested in DQ. The second workshop enlisted members of the informatics and comparative effectiveness research (CER)

Figure 1. Timeline of Significant Events in Developing the Harmonized DQ Terminology



community. Both workshops explored the need for clarity and transparency in DQ concepts to the specific community and participants. The patients and policymaker workshop also focused on preferred methods for communicating DQ findings and engagement methods to ensure that DQ efforts include stakeholder perspectives. The informatics and CER community workshop participants reviewed and critiqued the draft DQ terms and definitions. Significant portions of the discussions during each workshop were recorded and were later transcribed, imported into ATLAS.ti (qualitative data analysis software), and reviewed to identify themes and subthemes expressed by participants during each workshop. Recommended changes to the DQ terms, categories, and definitions identified by workshop participants were incorporated into the next version of the terminology. A second round of iterative, expert consensus development was performed by incorporating comments received from a wiki page dedicated to the DQ terminology, as well as introducing the revised harmonized draft terminology presented by webinars to national audiences.

Volume 4 (2016) Issue Number 1

It is estimated that through all outreach efforts, approximately 100 unique individuals from diverse disciplines, as well as United States and international networks and projects, contributed to the development and review of the harmonized DQ terminology. Recommendations that had strong consensus were integrated throughout the process. DQ terms that addressed similar issues were merged and grouped into categories and subcategories. Terms that focused on DQ features intrinsic to data elements—such as their presence or absence, format, values, and distributions—were included in the terminology. Terms that focused on DQ features extrinsic to data elements, such as data access, system availability, security and privacy concerns, and features for determining fitness for a specific analysis (fitness for use) were not included. We examine this scoping decision in the Discussion section. The final set of DQ terms developed by the primary authors were organized into a logical framework with three major categories, which were further separated into two evaluation contexts, Verification and Validation.

Results

Table 1 presents the harmonized set of DQ terms, definitions, and examples organized by categories and subcategories. Each DQ category conveys a DQ concept that needs to be interpreted within a certain context. Contexts distinguish different strategies for assessing DQ. Therefore, all DQ categories are divided into Verification and Validation (Table 1, top row). The key difference between these two DQ assessment contexts is *framing expectations*. Verification focuses on how data values match expectations with respect to metadata constraints, system assumptions, and local knowledge. This DQ context does not rely on an external reference or benchmark. Expectations for DQ measures are internally derived, based on expert judgment, relevant heuristics, and knowledge of impossibilities or contradictions. DQ assessment measures in this context can be created using information available within the existing data environment. While the scope of "relevant heuristics and local knowledge" may vary across environments (e.g., one group may have access to domain experts not available to another group), the key feature with Verification is the ability to determine expected values and distributions using resources within the local environment.

Validation focuses on the alignment of data values with respect to relevant external benchmarks. In Validation, expectations are derived from comparisons to known true or relative gold standards and external knowledge that exists as resources independent of the data source being evaluated. Declaring a resource or benchmark to be a gold standard requires special knowledge about the degree of confidence or trust that a community of users has in an external data source.[26] Another common method of creating a relative gold standard can be to combine results across multiple data providers, such as all data partners in a data network, to create a merged data set that can be used as an "external" comparator.

Verification and validation can be equally appropriate to assess EHR DQ, depending on the data context and intended use. EHR data is a novel source of information in many organizations and communities. As a result, verification of local data may be the *only* appropriate reference against which to frame expectations and outcomes. For example, EHR data may be used to generate estimates on a denominated population for which a near-census is available. In such cases, validation against national survey data is likely to be a poor reference due to fundamental differences in the goals of data collection and subsequent impacts on survey design decisions such as the sampling frame and the granularity of data elements collected. Context and

intended use of EHR data motivate incorporating a verification or validation assessment strategy (or both).

DQ Category 1: Conformance

Conformance focuses on DQ features that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions. Conformance DQ measures do not attempt to assess the completeness or plausibility of the values that are recorded but merely whether the values that are present meet syntactic or structural constraints. Expected conformance features are often described in a document called a "data dictionary," which lists the intended format and allowed values for every data element. Conformance DQ features are divided into three subcategories: value conformance, relational conformance, and computational conformance.

Value Conformance

Value conformance seeks to determine if recorded data elements are in agreement with a prespecified, constraint-driven data architecture. Internal data constraints are typically imposed by a formal data model, which specifies expectations for data types, data domains and allowed values, and data formats. These constraints typically are documented in a data dictionary. Validation adds constraints imposed by conforming to external standards for data representation and values, such as value set constraints imposed by external terminology standards.

Relational Conformance

Relational conformance seeks to determine if the recorded data elements are in agreement with additional structural constraints imposed by the physical database structures that store data values. In this category are conformance to nullability constraints (data fields that are allowed to null or must always have a value) and to primary key and foreign key relationships. Both Verification and Validation relational constraints express how a data model or standards body represents reality; the structures that define these constraints are usually represented in metadata descriptions or implemented as database integrity rules or attribute domain sets.

Computational Conformance

Computational conformance seeks to determine if computations used to create derived values from existing variables yield the intended results either *within* a data set (Verification) or *between* data sets (Validation), when programs are based on identical specifications. Computational conformance focuses on the correctness of the output value of calculations against technical functional specifications. Other DQ categories, such as Plausibility, focus on the tenability of the calculated results. That is, a calculation may be correct according to its formal specification (Computational Conformance) yet not be an accurate representation of the intended concept (Plausibility) because the specification is incorrect or incomplete. Areas of concern in this category are the use of correct logic and formulas, including conditions with unusual or unexpected inputs, which yield the intended output under all circumstances.

DQ Category 2: Completeness

Completeness focuses on features that describe the frequencies of data attributes present in a data set *without reference to data values*. Completeness measures assess the absence of data at a single moment over time or when measured at multiple moments over time, without reference to its structure or plausibility, which are assessed in the Conformance and Plausibility DQ categories

Volume 4 (2016) Issue Number 1

Table 1. Harmonized DQ Terms, Definitions, and Examples: Organized by Verification and Validation Contexts Within Categories and Subcategories

VERIFICATION		VALIDATION	
DEFINITION	**EXAMPLE**	**DEFINITION**	**EXAMPLE**
CONFORMANCE: DO DATA VALUES ADHERE TO SPECIFIED STANDARDS AND FORMATS?			
VALUE CONFORMANCE			
a. Data values conform to internal formatting constraints. b. Data values conform to allowable values or ranges.	a. Sex is only one ASCII character. b. Sex only has values "M," "F," or "U."	a. Data values conform to representational constraints based on external standards.	a. Values for primary language conform to ISO standards.
RELATIONAL CONFORMANCE			
a. Data values conform to relational constraints. b. Unique (key) data values are not duplicated. c. Changes to the data model or data model versioning.	a. Patient medical record number links to other tables as required. b. A medical record number is assigned to a single patient. c. Version 1 data does not include medical discharge hour.	a. Data values conform to relational constraints based on external standards.	a. Data values conform to all not-NULL requirements in a common multi-institutional data exchange format.
COMPUTATIONAL CONFORMANCE			
a. Computed values conform to computational or programming specifications.	a. Database- and hard-calculated Body Mass Index (BMI) values are identical.	a. Computed results based on published algorithms yield values that match validation values provided by external source.	a. Computed BMI percentiles yield identical values compared to test results and values provided by the CDC.
COMPLETENESS: ARE DATA VALUES PRESENT?			
a. The absence of data values at a single moment in time agrees with local or common expectations. b. The absence of data values measured over time agrees with local or common expectations.	a. The encounter ID variable has missing values. b. Gender should not be null. c. Medical discharge time is missing for three consecutive days.	a. The absence of data values at a single moment in time agrees with trusted reference standards or external knowledge. b. The absence of data values measured over time agrees with trusted reference standards or external knowledge.	a. The current encounter ID variable is missing twice as many values as the institutionally validated database. b. A drop in ICD-9CM codes matches implementation of ICD-10CM
PLAUSIBILITY: ARE DATA VALUES BELIEVABLE?			
UNIQUENESS PLAUSIBILITY			
a. Data values that identify a single object are not duplicated.	a. Patients from a single institution do not have multiple medical record numbers.	a. Data values that identify a single object in an external source are not duplicated.	a. An institution's CMS facility identifier does not refer to a multiple institutions.

Table 1. Harmonized DQ Terms, Definitions, and Examples: Organized by Verification and Validation Contexts Within Categories and Subcategories (Cont'd)

VERIFICATION		VALIDATION	
DEFINITION	**EXAMPLE**	**DEFINITION**	**EXAMPLE**
ATEMPORAL PLAUSIBILITY			
a. Data values and distributions agree with an internal measurement or local knowledge. b. Data values and distributions for independent measurements of the same fact are in agreement. c. Logical constraints between values agree with local or common knowledge (includes "expected" missingness). d. Values of repeated measurement of the same fact show expected variability.	a. Height and weight values are positive. a. Counts of unique patients by diagnoses are as expected a. Distribution of encounters per patient or medications per encounter distributions are as expected b. Serum glucose measurement is similar to finger stick glucose measurement. b. Oral and axillary temperatures are similar. c. Sex values agree with sex-specific contexts (pregnancy, prostate cancer). d. Height values are similar when taken by two separate nurses within the same facility using the same equipment.	a. Data values and distributions (including subgroup distributions) agree with trusted reference standards or external knowledge. b. Similar values for identical measurements are obtained from two independent databases representing the same observations with equal credibility. c. Two dependent databases (e.g., database 1 abstracted from database 2) yield similar values for identical variables.	a. HbA1c values from hospital and national reference lab are statistically similar under the same conditions. a. Distribution of patients with cardiovascular disease diagnoses are similar to CDC rates for the same age and sex groups a. Readmission rates by age groups for Medicare patients agree with CMS values b. Diabetes ICD-9CM and CPT codes are similar between two independent claims databases serving similar populations. c. Recorded date of birth is consistent between EHR data and registry data for the same patient.
TEMPORAL PLAUSIBILITY			
a. Observed or derived values conform to expected temporal properties. b. Sequences of values that represent state transitions conform to expected properties. c. Measures of data value density against a time-oriented denominator are expected *based on internal knowledge*.	a. Admission date occurs before discharge date. b. Date of an initial immunization precedes date of a booster immunization. c. Similar counts of patient observations between extraction-transformation-load cycles. c. Counts of emergency room visits by month shows expected spike during flu season. c. Medications per patient-day are as expected	a. Observed or derived values have similar temporal properties across one or more external comparators or gold standards. b. Sequences of values that represent state transitions are similar to external comparators or gold standards. c. Measures of data value density against a time-oriented denominator are expected *based on external knowledge*.	a. Length of stay by outpatient procedure types conforms to Medicare data for similar populations. b. Immunization sequences match the CDC recommendations. c. Counts of emergency room visits by month shows spike during flu season that are similar to local health department reports. c. Medications per patient-day matches claims data.

Notes: The lettering in each column can be used to map each definition to its corresponding example. Not every definition has a corresponding example.

Extract, Transform, Load ETL (ETL); International Organization for Standardization (ISO); Electronic Health Record (EHR) Data; International Classification of Diseases, Ninth and Tenth Revisions (ICD-9CM and ICD-10CM); Current Procedural Terminology (CPT); Centers for Medicare & Medicaid Services (CMS); Centers for Disease Control and Prevention (CDC).

Volume 4 (2016) Issue Number 1

respectively. Not explicitly represented by a separate Completeness subcategory is the critical concept of *why* data are complete or incomplete (i.e., the underlying mechanism leading to incompleteness), commonly known as "missingness" in the statistics literature.[61,62] While there are important differences between types of missingness (i.e., missing at random, missing completely at random, and missing not at random), these differences do not require different conceptualizations within the DQ category of Completeness, but do suggest different assessment methods, analytic approaches, and underlying root causes that lead to incompleteness.[62,63]

The restriction of completeness to the presences of data "without reference to data values" is an important feature that distinguishes counts that are considered measures of completeness from counts that are considered measures of plausibility.

DQ Category 3: Plausibility

Plausibility focuses on features that describe the believability or truthfulness of data values. For this category, plausibility is determined by a variable's value, when a value is placed within the context of another variable (i.e., two independent variables assessing the same construct), or a temporal sequence or state transition (i.e., patient follow-up treatment for a disease must be preceded by a corresponding diagnosis). Unlike Conformance and Completeness, which focus only on the structure and presence of values respectively, Plausibility focuses on actual values as a representation of a real-world object or conceptual construct by examining the distribution and density of values or by comparing multiple values that have an expected relationship to each other.

The term "plausibility" was carefully selected; it implies the existence of an acceptable variable value range and distribution rather than requiring the existence of a single absolute truth. Alternative words, such as "accuracy," "correctness," and "validity" were considered and set aside because these words have a wide range of competing and inconsistent definitions in the literature and are applied differently in the psychometric and DQ communities. Plausibility has uniqueness, atemporal, and temporal subcategories.

Uniqueness Plausibility

The Uniqueness subcategory seeks to determine if objects (entities, observations, facts) appear multiple times in settings where they should not be duplicated or cannot be distinguished within a database (Verification) or when compared with an external reference (Validation). Duplication frequently occurs when disparate data streams that contain overlapping objects are combined. Data extraction errors, such as incomplete relational join conditions, can also generate duplicate records.

Atemporal Plausibility

Atemporal Plausibility seeks to determine if observed data values, distributions, or densities agree with local or "common" knowledge (Verification) or from comparisons with external sources that are deemed to be trusted or relative gold standards (Validation).[26] Examples include physical quantities that cannot be negative, exceed possible boundaries, or (in health care) represent physiologically impossible states. Expected values, distributions, or densities may vary based on context; variables may have different expected distributions or densities when stratified by age, gender, and socioeconomic values.

In settings where logic or knowledge do not provide clear guidance on expectations, external gold standards created by organizations thought to implement strict DQ standards can represent an external reference source of comparisons for variable values and distributions. Atemporal plausibility also

focuses on the observed relationships between independent variables that have expected or known relationships, either due to local knowledge about how concepts in the data set should be related or due to physical or conceptual constraints that should tie variables together in a specific manner. Included in this subcategory is the examining of whether repeated values of the same observation or fact align as expected. The expected variability of repeated measurements of the same fact (observation, event) comes either from local knowledge or from comparisons with measurements drawn from similar external data sources.

Temporal Plausibility

Temporal plausibility seeks to determine if time-varying variables change values as expected based on known temporal properties or across one or more external comparators or gold standards. Temporal properties that establish expectations in this subcategory include temporal stability (do values vary over time as expected), temporal continuity (do values persist over time as expected), state transitions (do sequences of events occur as expected), and temporal dependencies between time-varying variables. In health care data, additional observed temporal properties may be temporal cycles, such as diurnal variations or cyclical treatment regimens, and temporal reoccurrences, such as recurring disease flares. Computations that derive temporal abstractions, such as condition durations or medication exposure intervals based on medication ordering events, would not be included in this category but would be included in the computational conformance DQ category, even though time is involved in calculating the derived temporal concepts.

Table 2 illustrates the alignment of the harmonized DQ terms, categories, and subcategories with 10 previously published terminologies. When comparing the subset of terms and categories within these terminologies that apply to intrinsic DQ features and not to external operational features such as data access and security, the harmonized terminology had good coverage. A majority of the existing terminologies adequately mapped to DQ concepts within the data verification context, and fewer mapped to DQ concepts within the data validation context.

Discussion

The DQ framework and terminology presented in Table 1 aligns a wide array of existing DQ concepts that have had different and sometimes inconsistent definitions. In a nonmedical context, Wang and Strong enumerated 179 different DQ terms (called "DQ attributes") collected as part of a survey with business data consumers.[42] From a review of the electronic medical record literature, Weiskopf and Weng similarly categorized 27 unique terms for DQ dimensions.[45] Other approaches, such as the ontological methods by Liaw[46] and Johnson,[51] have been used to develop alternative DQ terms. In the current work, community-based consensus, iterative refinement, and continuous alignment across existing DQ terminologies were used to arrive at a common set of consensus terms and definitions organized into a three-category framework.

Our framework highlights two distinct data-quality assessment contexts that distinguish two distinct strategies for the source of expectations or comparisons of EHR data based on internal characteristics (Verification) or external resources (Validation). We make no claim that one strategy or set of expectations is "better than" another source. In fact, we argue that a comprehensive assessment program will include both types of assessments to address these equally important but distinct aspects of DQ.

Volume 4 (2016) Issue Number 1

Table 2. Crosswalk Between Harmonization Terminology, Categories, and Subcategories Versus Pre-Existing Categories and Frameworks

PROPOSED CATEGORIES	JOHNSON 2015 [51a]	ZOZUS 2014 [50b]	LIAW 2013 [46c]	WEISKOPF 2013a [48d]	WEISKOPF 2013b [45e]
CONFORMANCE					
Value	• Representation-Integrity • Coding-Consistency • Representation-Consistency	• Consistency	• Internal Consistency • External Consistency		• Plausibility
Relational	• Domain-Consistency • Domain-Metadata	• Data Element Completeness • Information Loss and Degradation			
Computational			• Correctness (Accuracy Elements)		• Concordance
COMPLETENESS					
Completeness	• Representation-Complete • Domain-Complete • Relative-Completeness	• "Column" Data Value • Completeness	• Completeness (Elements of Correctness)	• Documentation Completeness • Density Completeness	• Completeness
PLAUSIBILITY					
Uniqueness		• Ascertainment Completeness	• No Duplication		
Atemporal	• Domain-Consistency • Relative-Correctness • Relative-Completeness	• Representational Inaccuracy • Information Loss and Degradation • Consistency	• Correctness (Reliability Elements) • Consistency (Reliability Elements) • External Consistency	• Density Completeness	• Correctness • Concordance • Plausibility
Temporal	• Representation-Correctness	• Consistency			

Notes: Existing DQ approaches were organized chronologically; approaches with the same year of publication were ordered alphabetically. Only the first author and publication date are provided in the table.

[a]Johnson[51] Correctness (RepresentationIntegrity, RelativeCorrectness, RepresentationCorrectness, Reliability); Consistency (RepresentationConsistency, DomainConsistency, CodingConsistency, DomainMetadata); Completeness (RepresentationComplete, DomainComplete, RelativeCompleteness, Sufficiency, DomainCoverage, TaskCoverage, Flexibility, Relevance); Currency (RepresentationCurrent, Dataset Current, TaskCurrency). The proposed terminology does not capture Reliability, Sufficiency, DomainCoverage, TaskCoverage, Flexibility, Relevance, RepresentationCurrent, DatasetCurrent, or TaskCurrency.

[b]Zozus[50] Completeness (Data Element Completeness, "Column" Data Value Completeness, "Row" Data Value Completeness, and Ascertainment Completeness); Accuracy (Representational Inadequacy, Information Loss and Degradation); Consistency has no lower-level terminology. The proposed terminology does not capture "Row" Data Value Completeness.

[c]Liaw[46] The proposed terminology does not capture : Timeliness, Relevance, Usability, or Security.

[d]Weiskopf[48] Completeness (Documentation, Breadth, Density, and Prediction). The proposed terminology does not capture Breadth or Prediction.

[e]Weiskopf[45] Both Plausibility and Concordance are proxies of Correctness. The proposed terminology does not capture Currency.

Table 2. Crosswalk Between Harmonization Terminology, Categories, and Subcategories Versus Pre-Existing Categories and Frameworks (Cont'd)

PROPOSED CATEGORIES	KAHN 2012 [44f]	NAHM 2012 [57g]	MCGILVRAY 2008 [64h]	EPPLER 2006 [65i]	WANG 1996 [42j]
CONFORMANCE					
Value	• Attribute Domain Constraints • Historical Data Rules • State-Dependent Object Rules	• Granularity • Precision	• Data Integrity Fundamentals		• Representational Consistency
Relational	• Relational Integrity Rules	• Attribution	• Data Specifications		
Computational	• Attribute Dependency Rules				
COMPLETENESS					
Completeness	• Attribute Domain Constraints	• Completeness	• Data Integrity Fundamentals		• Completeness
PLAUSIBILITY					
Uniqueness	• Relational Integrity Rules		• Duplication		
Atemporal	• Attribute Domain Constraints • Relational Integrity Rules • Attribute Dependency Rules	• Consistency (Internal) • Granularity	• Data Integrity Fundamentals • Accuracy • Consistency and Synchronization	• Consistency • Correctness • Accuracy	• Believability • Accuracy • Representational Consistency
Temporal	• Historical Data Rules • Attribute Dependency Rules • State-Dependent Object Rules	• Accuracy	• Data Integrity Fundamentals		

Notes: Existing DQ approaches were organized chronologically; approaches with the same year of publication were ordered alphabetically. Only the first author and publication date are provided in the table.

[f]Kahn[44] Attribute Domain Constraints (Attribute Profiling, Optionality, Format, Valid Values, Precision); Relational Integrity Rules (Identity, Reference, Cardinality, Inheritance); Historical Data Rules (Currency, Retention, Granularity, Continuity, Timeline Patterns, Value Patterns, Event Dependencies, Event Conditions, Event Attributes); State-Dependent Object Rules (State-Transition Profiling, State Domain, Action Domain, Terminator Domain, State-Actions); Attribute Dependency Rules (Continuity, Duration, Redundant Attributes, Derived Attributes, Partially Dependent Attributes, Conditional Optionality, Correlated Attributes). The proposed terminology does not capture Historical Data Rules: Retention.

[g]Nahm[57] Inherent (Accuracy, Currency, Completeness, Consistency (internal), Specificity, Attribution); Context Dependent (Timeliness, Relevance, Granularity, Precision). The proposed terminology does not capture Currency, Timeliness, Relevance, or Specificity.

[h]McGilvray[64] The proposed terminology does not capture Timeliness And Availability, Ease Of Use And Maintainability, Data Coverage, Presentation Quality, Perception, Relevance and Trust, Data Decay, or Transactability.

[i]Eppler[65] Community Level (Comprehensiveness, Accuracy, Clarity, Applicability); Product Level (Conciseness, Consistency, Correctness, Currency); Process Level (Convenience, Timeliness, Traceability, Interactivity); Infrastructure Level (Accessibility, Security, Maintainability, Speed).The proposed terminology does not capture Comprehensiveness, Clarity, Applicability, Conciseness, Currency, Convenience, Timeliness, Interactivity, Accessibility, Security, Maintainability, or Speed.

[j]Wang[42] Intrinsic DQ (Believability, Accuracy, Objectivity, Reputation); Contextual DQ (Value-Added, Relevance, Timeliness, Completeness, Appropriate Amount of Data); Accessibility DQ (Accessibility, Access Security); Representational DQ (Interpretability, Ease of Understanding, Representational Consistency, Concise). The proposed terminology does not capture Value-Added, Cost-effectiveness, Relevancy, Interpretability, Ease of Understanding, Ease of Operations, Accessibility, Flexibility, Objectivity, Timeliness, Reputation, Concise, Access Security, Appropriate Amount of Data, Variety of Data, or Traceability.

Volume 4 (2016) Issue Number 1

The ordering of the DQ categories (Conformance, followed by Completeness, followed by Plausibility) in Table 1 is purposeful. The ordering represents the additive effect of a category on the next category. Conformance focuses strictly on the agreement of values against various technical specifications without regard to the amount or believability of those values. Completeness focuses on the absence of data of a variable, again without regard to the believability of those values. Finally, Plausibility focuses on the believability or correctness of data and counts of data that conform to technical specifications (Conformance) and are present (Completeness) in the data. The restriction of completeness to the presence of data "without reference to data values" is an important feature that distinguishes completeness measures from plausibility measures. Counts (or density) by values (or by grouping) is based on expected distributions by observed values. We placed all DQ features that are dependent on values (or groupings of values such as dates by quarter) into Plausibility. That is, measures that must be interpreted in the presence of data values are categorized as assessing the believability of the distribution across values.

Our harmonized terminology and framework focuses exclusively on DQ issues related to the *intrinsic* features of data values. Wang and Strong defined intrinsic DQ as DQ features that involved only the data values "in their own right" without reference to external requirements or tasks.[42] They differentiated intrinsic DQ features from contextual DQ features, which typically entail unique contextual or task-specific DQ requirements. Other terminologies in Table 2 include DQ issues related to operational features such as data access, availability, security, maintainability, and fitness-for-use features—such as timeliness, appropriate amount for intended use, and relevancy. These extrinsic DQ concepts are not represented in the harmonized terminology but will be the focus of future work.

The DQ literature uses the concept of fitness for use or fitness for purpose, where knowing the intended use of data determine when DQ assessment results are declared as fit or not fit.[43,66] Fitness for use does not change the underlying intrinsic DQ features of the elements in a data set; it does change the *acceptability* of measures of DQ based on the intended use. For example, a completeness measure of 70 percent may be acceptable for a variable that is known to be not relevant to an analysis but would be unacceptable in an analysis where the variable was deemed important.

Our DQ terminology does not include terms such as accuracy, validity, or correctness that are commonly found in other DQ terminologies (Table 2). These terms are used in widely differing ways across the DQ literature. In addition, they have very specific meanings in the psychometric methods development community that are different than their intended meaning in a DQ context. Given the large diversity of interpretations and the competing use of the same terms in a closely related field that caused significant confusion while constructing our harmonized terminology, we avoided using these terms—while selecting terms that captured our intended concept that represents the trueness of data values.[67,68] A secondary goal was to select terms that are easily assessable to a nontechnical audience and that have the least likelihood of being confused with alternative existing technical uses of the same word.

DQ issues can occur in data sets constructed from a single institution. More challenging is understanding DQ issues in data sets constructed by combining data sets from multiple separate institutions. Differences in how data are captured and stored

in the original source-data system—usually the institutional EHR or enterprise data warehouse, how data are extracted and transformed into the analytic data, and the impact of data workflows or provenance can cause significant challenges in ensuring common data formats (syntax) and meaning (semantics).[69] Currently, most data networks rely on manual processes for establishing common data definitions and reproducible data extraction conventions. Hence, DQ issues can appear in any data source, even those generated by highly skilled technical staff. The use of a harmonized DQ terminology will not avoid DQ issues but is the first step of a collective community-based effort to identify and harmonize their use in a consistent manner.

Limitations

While the broad and diverse group of collaborators who developed the harmonized DQ terminology represent diverse networks that include hundreds of millions of patients, feedback regarding the harmonization terminology was generated from a convenience sample of EDM Forum members. However, a wide range of outreach methods were used to ensure representative community stakeholder engagement.[70] Most of the researchers who developed DQ terminologies included in Table 2 were collaborators and co-authors. While over 30 individuals provided input during the development of the terminology, many other relevant constituencies were not included, such as individuals responsible for using clinical data for operational and quality improvement activities within health systems, technical staff developing enterprise data warehouses or data marts, and journal editors interested in ensuring DQ transparency in their publications.

As highlighted in Table 2, the scope of this harmonization effort was limited to intrinsic DQ issues that focus on the structure and presence of data and their values. Not included in this terminology are other widely identified dimensions of DQ that focus on operational or fitness for use features. These DQ concepts are important in the context of a specific data system and environment, and the unique criteria that determine if a data set can answer a specific question. In addition, we did not include DQ issues that arise from deficiencies in the data representation or data model used to store data values. For example, a data model that captures International Classification of Diseases, Ninth and Tenth Revisions (ICD-9CM and ICD-10CM) codes for diagnoses in a single table but does not allow for representing data provenance is unable to distinguish between admitting, preliminary, and discharge diagnoses. This limitation of the data model can have a significant impact on the fitness for use of a data set but is not considered part of the DQ framework presented here. We discuss data model issues and its impact on CER elsewhere.[71]

Similarly, our framework does not focus on the quality of the definition of a data variable other than how that definition has an impact on expectations about values associated with that variable. Inadequacies or inaccuracies in the definition of a variable may result in high-quality values that do not represent the intended concept and therefore are not useful.

This analysis focused on "traditional" data extracted from EHRs and administrative systems. In the era of big data, which brings together new data sources with widely varying data characteristics, new DQ concepts, measures, and computational methods could emerge, resulting in expansion or alteration of the current terminology. For example, biological "-omics" data may generate new DQ concepts based on unique DQ features seen in these data that are not present in clinical and EHR data. Similarly, real-time physiological signals, wearable medical devices, and patient-entered data may also have unique DQ

Volume 4 (2016) Issue Number 1

features that require additional terms. Methods are needed for ensuring that this terminology remains all-inclusive while being continually adapted to accommodate new uncovered DQ dimensions. A table has been posted on a wiki hosted by the EDM Forum (http://repository.academyhealth.org/dqc/) where new DQ terms and methods can be debated and added by the community.

Finally, the terminology in Table 1 represents only one part of a comprehensive unified DQ framework, which would include DQ assessment methods and DQ reporting standards. Proposed DQ reporting standards have been published previously.[52]

Future Work

Table 2 validates the inclusive breadth of our harmonized DQ concepts and framework against a wide range of published DQ terms. A next step in ensuring that the harmonized DQ concepts are comprehensive is to evaluate existing DQ assessment measures (also called "DQ checks") to ensure that these DQ measures have a corresponding DQ category and term. An assessment method that cannot be unambiguously assigned to a DQ concept implies that either the concepts are not clearly defined or may indicate a missing concept. One difficulty in performing this assessment is that many DQ activities are performed as study-specific "one-off" analyses that are not well documented and are therefore difficult to discover. Some large national networks have published their DQ checks or have made their DQ check code available as open source. MS[53] and OHDSI Achilles Heel are examples.[72]

The ultimate value of this work will be seen when data owners develop and report DQ measures based on the DQ categories in Table 1. To encourage adoption of these measures, we are developing a DQ assessment toolkit that will compute standardized DQ measures and data visualizations that will help identify DQ issues using open source tools. Critical to this work is the development of a data model for storing DQ measures that is not tied to the specific data formats used to store the original data set.

In a previous publication, we provide recommendations for reporting DQ results.[52] However, technical and nontechnical barriers to reporting on DQ issues may prevent investigators from sharing DQ results. For example, concerns about reputational or organizational impact from revealing DQ issues could lead to reluctance to make these findings widely available. We have initiated a community survey to explore these possibilities.

Conclusions

The arrival of the era of larger volumes of electronically available data has increased the availability and reuse of EHR data. While these data have great potential for significant advancement in clinical practice and research, the quality of these data sources ultimately determine their utility. To fully understand and accurately characterize the limitations of these data sources, establishing standardized, validated methodologies for assessing and reporting DQ is crucial. The current project developed a comprehensive DQ terminology organized into a three-tier conceptual DQ framework for EHR data. Future research on the DQ terminology should aim to verify its generalizability and utility, add fitness for use terms and concepts, and extend the application of the terminology to emerging forms of digital health data poised to provide new insights into health and well-being. Future efforts should also focus on expanding this work from a harmonized terminology to a harmonized operational framework that includes reusable DQ assessment, visualization, and reporting capabilities for understanding the strengths and limitations of EHR data for secondary use.

Acknowledgements

Research reported in this publication was funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1303-5581). The statements in this publication are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee. Members of the EDM Forum staff provided logics support for the face-to-face meetings and webinars illustrated in Figure 1 and in establishing a community web site and wiki for gathering community input. These efforts were critical to outreach efforts. Members of the monthly Data Quality Collaborative call provided contacts with individuals and organizations to engage in additional outreach efforts. We thank this large group of dedicated individuals who have made data quality a topic of scientific inquiry, relevance, and importance.

References

1. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. Am J Prev Med. 2007 Aug;33(2):155–61.
2. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007 Jan 1;14(1):1–9.
3. Weiner MG, Embi PJ. Toward Reuse of Clinical Data for Research and Quality Improvement: The End of the Beginning? Annals of Internal Medicine. 2009;151(5):359–60.
4. Lopez MH, Holve E, Sarkar IN, Segal C. Building the informatics infrastructure for Comparative Effectiveness Research (CER): A review of the literature. Medical care. 2012 Jul;50 Suppl:S38-48.
5. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. J Am Med Inform Assoc. 2014 May 12;
6. Pace WD, Cifuentes M, Valuck RJ, Staton EW, Brandt EC, West DR. An electronic practice-based network for observational comparative effectiveness research. Ann Intern Med. 2009 Sep 1;151(5):338–40.
7. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. Medical care. 2010;48(6):S45–S51.
8. Randhawa GS, Slutsky JR. Building sustainable multi-functional prospective electronic clinical data systems. Med Care. 2012 Jul;50 Suppl:S3-6.
9. Randhawa GS. Building electronic data infrastructure for comparative effectiveness research: accomplishments, lessons learned and future steps. J Comp Eff Res. 2014 Nov;3(6):567–72.
10. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc. 2014 May 12;
11. Curtis LH, Brown J, Platt R. Four Health Data Networks Illustrate The Potential For A Shared National Multipurpose Big-Data Network. Health Affairs. 2014 Jul 1;33(7):1178–86.
12. Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. Am J Manag Care. 2007 Jun;13(6 Part 1):277–8.
13. Immanuel V, Johnson K, Young B, Hart G. Testimony on secondary uses of health data to the National Committee on Vital and Health Statistics. 2007.
14. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc. 2010 Mar 1;17(2):124–30.
15. Chute CG, Pathak J, Savova GK, Bailey KR, Schor MI, Hart LA, et al. The SHARPn project on secondary use of Electronic Medical Record data: Progress, plans, and possibilities. AMIA . Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2011;2011:248–56.
16. Holzer K, Gall W. Utilizing IHE-based Electronic Health Record systems for secondary use. Methods of information in medicine. 2011;50:319–25.
17. Roth C, Shivade CP, Foraker RE, Embi PJ. Integrating population- and patient-level data for secondary use of electronic health records to study overweight and obesity. Stud Health Technol Inform. 2013;192:1100.
18. Johnson EK, Broder-Fingert S, Tanpowpong P, Bickel J, Lightdale JR, Nelson CP. Use of the i2b2 research query tool to conduct a matched case–control clinical research study: advantages, disadvantages and methodological considerations. BMC Med Res Methodol. 2014 Jan 30;14:16.
19. Helmer KG, Ambite JL, Ames J, Ananthakrishnan R, Burns G, Chervenak AL, et al. Enabling collaborative research using the Biomedical Informatics Research Network (BIRN). Journal of the American Medical Informatics Association : JAMIA. 2011 Jul;18:416–22.
20. McMurry AJ, Murphy SN, MacFadden D, Weber G, Simons WW, Orechia J, et al. SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies. PLoS ONE. 2013 Mar 7;8(3):e55811.

Volume 4 (2016) Issue Number 1

21. Ross TR, Ng D, Brown JS, Pardee R, Hornbrook MC, Hart G, et al. The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. eGEMs (Generating Evidence & Methods to improve patient outcomes) [Internet]. 2014 Mar 24 [cited 2014 Apr 16];2(1). Available from: http://repository.academyhealth.org/egems/vol2/iss1/2
22. NIH Collaboratory Health Care Systems Research Collaboratory home page [Internet]. [cited 2016 Jun 8]. Available from: https://www.nihcollaboratory.org/about-us/Pages/default.aspx
23. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform. 2015 Feb;53:162–73.
24. van der Lei J. Use and abuse of computer-stored medical records. Methods Inf Med. 1991 Apr;30(2):79–80.
25. Brennan PF, Stead WW. Assessing data quality: from concordance, through correctness and completeness, to valid manipulatable representations. J Am Med Inform Assoc. 2000 Jan;7:106–7.
26. Kahn MG, Eliason BB, Bathurst J. Quantifying clinical data quality using relative gold standards. AMIA . Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2010;2010:356–60.
27. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care. 2013 Aug;51(8 Suppl 3):S30-37.
28. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997 Oct;4(5):342–55.
29. Aronsky D, Haug PJ. Assessing the quality of clinical data in a computer-based record for calculating the pneumonia severity index. J Am Med Inform Assoc. 2000 Feb;7(1):55–65.
30. Arts D, de Keizer N, Scheffer G-J, de Jonge E. Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. Intensive Care Med. 2002 May;28(5):656–9.
31. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ. 2003 May 15;326(7398):1070.
32. Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. AMIA Annu Symp Proc. 2006;324–8.
33. Cruz-Correia RJ, Rodrigues P, Freitas A, Almeida FC, Chen R, Costa-Pereira A. Data quality and integration issues in electronic health records. In: Hristidis V, editor. Information discovery on electronic health records. Chapman and Hall/CRC; 2009. p. 55–95.
34. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: Data quality issues and informatics opportunities. AMIA Summits on Translational Science proceedings AMIA Summit on Translational Science. 2010;2010:1–5.
35. Hripcsak G, Knirsch C, Zhou L, Wilcox A, Melton G. Bias associated with mining electronic health records. J Biomed Discov Collab. 2011;6:48–52.
36. Brown JS, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. Medical care. 2013;51:S22–S29.
37. Rusanov A, Weiskopf NG, Wang S, Weng C. Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. BMC Med Inform Decis Mak. 2014;14:51.
38. Kahn MG, Ranade D. The impact of electronic medical records data sources on an adverse drug event quality measure. J Am Med Inform Assoc. 2010 Mar 1;17(2):185–91.
39. Chan KS, Fowles JB, Weiner JP. Electronic Health Records and Reliability and Validity of Quality Measures: A Review of the Literature. Medical Care Research and Review [Internet]. 2010 Feb [cited 2010 Jul 6]; Available from: http://0-mcr.sagepub.com.impulse.ucdenver.edu/cgi/rapidpdfsidebar/1077558709359007v1?&frameset_url=http%3A%2F%2Fmcr.sagepub.com%2Fcgi%2Frapidpdf%2F1077558709359007v1
40. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. Journal of the American Medical Informatics Association. 2012 Jul 1;19(4):604–9.
41. Brown JS, Chun A, Davidson BN, Holve E, Kahn MG, Hamilton Lopez M, et al. Recommendations for transparent reporting of data quality assessment results for observational healthcare data. eGEMs (Generating Evidence & Methods to improve patient outcomes). 2015 accepted for publication;
42. Wang R, Strong D. Beyond accuracy: What data quality means to data consumers. J Management Information Systems. 1996;12:5–34.
43. Redman TC. Data Quality: The Field Guide. Boston: Digital Press; 2001. 241 p.
44. Kahn MG, Raebel MA, Glanz JM, Riedlinger K, Steiner JF. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Medical care. 2012 Jul;50 Suppl:S21-9.
45. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. Journal of the American Medical Informatics Association : JAMIA. 2013 Jan 1;20:144–51.
46. Liaw ST, Rahimi A, Ray P, Taggart J, Dennis S, de Lusignan S, et al. Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature. International Journal of Medical Informatics. 2013 Jan;82(1):10–24.
47. Sadiq S, editor. Handbook of Data Quality. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013.
48. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform. 2013 Oct;46(5):830–6.
49. Rahimi A, Liaw S-T, Ray P, Taggart J, Yu H. Ontological specification of quality of chronic disease data in EHRs to support decision analytics: a realist review. Decision Analytics. 2014 Feb 19;1(1):5.
50. Zozus MN, Hammond WE, Green BB, Kahn MG, Richesson RL, Rusincovitch SA, et al. Assessing Data Quality for Healthcare Systems Data Used in Clinical Research (Version 1.0) [Internet]. [cited 2014 Sep 13]. Available from: http://sites.duke.edu/rethinkingclinicaltrials/assessing-data-quality/

51. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. A Data Quality Ontology for the Secondary Use of EHR Data. In: Proceedings 2015 American Medical Informatics Association Fall Symposium. San Francisco, CA; p. accepted for publication.
52. Kahn MG, Brown J, Chun A, Davidson B, Meeker D, Ryan P, et al. Transparent Reporting of Data Quality in Distributed Data Networks. eGEMs (Generating Evidence & Methods to improve patient outcomes) [Internet]. 2015 Mar 23;3(1). Available from: http://repository.academyhealth.org/egems/vol3/iss1/7
53. Mini-Sentinel Coordinating Center. Mini-Sentinel Standard Operating Procedure: Data Quality Checking and Profiling. 20; Available from: http://www.mini-sentinel.org/work_products/About_Us/Mini-Sentinel_SOP_Data-Quality-Checking-and-Profiling.pdf
54. Observational Medical Outcomes Partnership. OSCAR - Observational Source Characteristics Analysis Report (OSCAR) Design Specification and Feasibility Assessment [Internet]. 2011 [cited 2013 Apr 1]. Available from: http://omop.fnih.org/OSCAR
55. Observational Medical Outcomes Partnership. Generalized Review of OSCAR Unified Checking [Internet]. 2011 [cited 2013 Apr 1]. Available from: http://omop.fnih.org/GROUCH
56. Canadian Institute for Health Information. The CIHI data quality framework [Internet]. Ottawa, Ont.: CIHI; 2009. Available from: http://www.cihi.ca/CIHI-ext-portal/pdf/internet/DATA_QUALITY_FRAMEWORK_2009_EN
57. Nahm M. Data quality in clinical research. In: Clinical Research Informatics. London: Springer-Verlag; 2012. p. 175–201. (Health Informatics).
58. Maydanchik A. Data quality assessment. Bradley Beach, NJ: Technics Publications; 2007. xiv, 321 . (Data quality for practitioners series).
59. Magnusson D, Bergman LR, European Network on Longitudinal Studies on Individual Development. Data quality in longitudinal research. Cambridge [England] ; New York: Cambridge University Press; 1990. xii, 285 .
60. Singh S. Evaluation of data quality. [London, England], International Statistical Institute by Oxford University Press,; 1987. p. 618-643.
61. Little RJ, Rubin DB. Statistical Analysis with Missing Data (2nd Ed.). New York, N.Y.: Wiley; 2002. (Series in probability and statistics; vol. 1).
62. Haziza D. Imputation and Inference in the Presence of Missing Data. In: Rao CR, editor. Handbook of Statistics [Internet]. New York, N.Y.: Elsevier; 2009 [cited 2013 May 6]. p. 215–46. Available from: http://www.sciencedirect.com/science/article/pii/S0169716108000102
63. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology. 2006 Oct;59(10):1087–91.
64. McGilvray D. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information. 1 edition. Amsterdam ; Boston: Morgan Kaufmann; 2008. 352 p.
65. Eppler M, Helfert M. A Classification and Analysis of Data Quality Costs. In Miami, Florida, USA; 2004.
66. Juran JM, Godfrey AB, editors. Juran's quality handbook. 5th ed. New York: McGraw Hill; 1999. 1 p. (McGraw-Hill handbooks).
67. International Organization for Standardization. Accuracy (trueness and precision) of measurement methods and results-Part 1: General principles and definitions. ISO; 1994. Report No.: ISO 5725-1.
68. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accreditation and Quality Assurance. 2007 Jan 9;12(1):45–7.
69. Johnson KE, Kamineni A, Fuller S, Olmstead D, Wernli KJ. How the Provenance of Electronic Health Record Data Matters for Research: A Case Example Using System Mapping. eGEMs (Generating Evidence & Methods to improve patient outcomes) [Internet]. 2014 Apr 16 [cited 2016 Mar 13];2(1). Available from: http://repository.edm-forum.org/egems/vol2/iss1/4
70. Creswell JW. Qualitative, Quantitative, and Mixed Methods Approaches (Crewell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches) 4th edition. Fourth Edition edition. Thousand Oaks, California: SAGE Publications, Inc; 2013. 273 p.
71. Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. Medical care. 2012 Jul;50 Suppl:S60-7.
72. Observational Health Data Sciences and Informatics (OHDSI). ACHILLES for data characterization | OHDSI [Internet]. [cited 2015 Jul 6]. Available from: http://www.ohdsi.org/analytic-tools/achilles-for-data-characterization/