A CONSENSUS-BASED DATA QUALITY ASSESSMENT MODEL FOR PATIENT REPORTED OUTCOME INFORMATION IN DIGITAL QUALITY MEASUREMENT PROGRAMS

by
Ben Hamlin

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Public Health

Baltimore, Maryland
January 2022

Abstract

Quality measurement has been evolving to become more patient-focused and more meaningful in supporting quality improvement. Recent advancements in digital data and measurement standards have made this evolution possible, but this move to digital measurement presents several challenges despite its many benefits. Digital quality measures (dQMs) substantially reduce the computational burden of generating "quality" knowledge and improve the reliability of the measure scores they generate, however they rely on a very specific presentation of the electronic data to achieve the aforementioned benefits. Newer dQMs based on patient-reported outcomes (PROs) measured using patient-reported outcome measures (PROMs) have been gaining attention as they generate valuable insight into a person's perception of their own health status. Reliably capturing these insights is challenging however, as the information does not often exist in a format that can be processed by a dQM. This lack of standardization has resulted in the formation of clinical data repositories (CDRs) for the explicit purpose of extracting, transforming, and loading (ETL) PROM data from patients' medical records into a format that can support digital quality measurement.

These ETL processes are subject to rigorous evaluation to ensure that, as the information is being transformed, the integrity of the original information is being preserved. These evaluations inform decisions regarding data fitness for the specific purpose of using the data to measure quality of care. These "fit for purpose" decisions are not guided by a uniform set of expectations or requirements to assure consistency in decision-making, rather they frequently rely upon a variety of statistical and operational test results that can often present seemingly inconsistent information that requires substantial expertise to interpret and reconcile. A

uniform, well-defined list of data quality concepts pertinent to using patient-reported outcome measures for the purpose of quality measurement would provide much needed guidance and enhance the consistency and reliability of data fitness decision-making.

This research confirmed the scarcity of access to effective guidance for assessing fitness of PROM data and that there is a desire for a standard PROM-based data quality assessment (DQA) model to support decision making.

Thesis Readers:
Thesis Advisor: Sydney Dy, Professor – Health Policy and Management
Thesis Chair: Joseph Gallo, Professor – Mental Health
Ramin Mojtabai, Professor – Mental Health
Hadi Kharrazi, Associate Professor– Health Policy and Management
Claire Snyder, Professor – School of Medicine

Table of Contents

List of Tables

List of Figures

Chapter 1: Assessing Data Fitness Parameters for Patient Reported Outcome Performance Measurement (PRO-PM) Data.

Abstract

Introduction

Newer digital quality measures (dQMs) based on patient-reported outcomes (PROs) measured using patient-reported outcome measures (PROMs) have been gaining attention as they generate valuable insight into a person's perception of their own health status. Reliably capturing these insights is challenging however, as the information is often not captured or stored in a standardized format compatible with the dQM technical specification. This lack of standardization has led to the formation of clinical data repositories established for the explicit purpose of extracting, transforming, and loading (ETL) PROM data into a format that can support digital quality measurement.

Data fitness is a specific construct of data quality where the determination of appropriateness for a dataset's use is ultimately declared by the person (data consumer) responsible for making decisions about whether a dataset is adequate for its intended purpose.[1] The importance of the data consumer's perspective and their specific expectations for patient-reported outcome measure (PROM) data quality have not yet been well delineated, resulting in sluggish adoption of PROM based dQMs. This study's objectives were to identify relevant existing data quality assessment (DQA) frameworks and describe data consumer expectations when assessing the sufficiency of PROM data for performance measurement.

Methods

Observations from informed stakeholders were integrated with information derived from a scan of the available literature on DQA frameworks. First, an environmental scan was

conducted to identify potentially relevant data quality frameworks and then 24 semi-structured key informant interviews were conducted to document perspectives of those responsible for making data quality decisions in the field.

Results

The environmental scan yielded a total of 631 references of which 134 were found to be relevant to both data quality and PROMs. Key informants highlighted the importance of context, specifically around the mode of data collection, the environment (or scenario) in which it was collected and the persistence of all this contextual information through to the person evaluating the data's fitness for purpose. The interviews also highlighted the notion of "trust" as a central factor in assessing data quality and reinforced the need to define context-specific parameters for evaluators such that they could uniformly feel confident about a dataset's use in performance measurement. When questioned about objectives for determining data quality, respondents expressed common experiences with inconsistencies in data completeness, frequent missingness of key variables and implausibility of variables outside normative clinical expectations, although when pressed on the existence of authoritative sources for data fitness evaluation that might enhance their decision making; none were able to point to a specific external source of information that supported their own intuitive data quality assessment practices.

Conclusion

PROM data fitness assessment, from the perspective of a data fitness evaluation for quality measurement, must include a thorough appraisal of contextual and provenance information. At the contextual level, the types of information that govern the plausibility of PROM data are

contained wholly within the evaluator's intuitive understanding of what is most explanatory for each intended use. Evaluators of data fitness frequently do not have the requisite contextual basis to enable a judgement on a particular dataset, rendering the data useless despite its potential to inform care quality.

Background:

A standardized data quality assessment (DQA) model that assures the consistency and completeness of patient reported outcome measure (PROM) data is indispensable for assessing healthcare quality. The importance of the evaluator's perspective and their specific expectations for PROM data quality has not yet been well delineated, resulting in sluggish adoption of PROM-based dQMs. To standardize data quality, those intrinsic values considered by decision makers to be the "gold standard" must first be empirically defined. A pragmatic model providing information needed by evaluators assessing PROM data fitness for performance measurement has yet to be constructed. This research describes decision makers' expectations for assessing sufficiency of electronic PROM data, specifically those parameters required to achieve an understanding that it is satisfactory for use in performance measurement programs. Answering the question: Within the context of PROMs, what attributes and degree of conformance are required to decide that PROM data are satisfactory for quality measurement purposes?

Evidence suggests that PROM tools used in the primary care setting may accurately identify depressed individuals whose timely intervention can improve depression-related outcomes.[1,2] Regular use of standard PROM instruments to collect patient-reported outcomes could enable care teams to assess patient symptoms, functioning, and well-being. Electronic clinical datasets offer many opportunities for assessing a variety of patient outcomes, however measuring quality of care using PROMs is particularly reliant on a series of processes that transform the data to a format useable by a digital quality measure (dQM). The process of extracting, transforming and loading (ETL) PROM data from its original form to a standard format that can

support digital quality measurement requires a pragmatic model to guide the evaluation of the data's accuracy and relevance throughout the entire transformation process. Standardized clinical data are those data which have undergone a process by which it is transformed into a format that can be processed or shared electronically between systems of care.[3,4] Standard clinical data is that which is stored using common interoperability definitions using preferred terminologies (codes) that represent the stored knowledge in a manner that can be used for clinical decision making.[5] A patient reported outcome (PRO) is a report of a person's health that is documented directly as it was reported by them without any interpretation of the person's response by a member of the care team.[6] Patient reported outcome measures (PROMs) are the standardized instruments used to collect PROs generally through the administration of self-report questionnaires.[6] If outcomes such as symptoms, functional status, or well-being are documented based on the report of a caregiver or health professional, they are not considered PROs, but observer-reported outcomes or clinician-reported outcomes, respectively. There are many modalities for administering PROMs including paper questionnaires, web-based applications (web apps) or by the patient directly entering responses into their electronic medical record form. Variation in PROM administration practices lead to the data being documented in the patient's record in multiple formats and locations. The typically non-standard nature of PROM data in electronic health records requiring multiple transformations presents a considerable barrier to broad implementation of PROM-based performance measures.

A major focus of many DQA models resides at the structural, or data element level. These models primarily address the process of defining data quality through a comprehensive process

of flagging inconsistencies in the structure at each data point then dichotomously categorizing as meeting/not meeting expected parameters. Data fitness evaluators are left to try to both interpret the prespecified parameters of the structural analysis, and to evaluate whether the errors seen are consistent with their own knowledge- a process reliant on their having substantial prior experience evaluating PROM datasets. This reliance on an individual's ability to draw from their own personal expertise presents particular challenges in the adoption of dQMs* where an intuitive approach to data quality is heavily relied upon.

Wang and Strong define data fitness as a specific construct of data quality where the fitness determination is decreed by the end user.[1] Their model separates intrinsic from contextual data quality where the former is independent from intended use and the latter is specific to a particular use case. Contextual data quality is explicit in its focus on the importance of the end user's perspective. It is important to distinguish that contextual data quality is pertinent to the aggregate, or summary level information that would be utilized by an evaluator in making decisions about its appropriateness. Therefore, the importance of the end user of the information, the "data consumer", in their construct cannot be ignored in any PROM data quality assessment model. Wang and Strong's principles were explicit on the data consumer being able to: 1) interpret the information, 2) find the information relevant, and 3) find the information to be accurate.[1] It is with these principles in mind that the current framework for PROM data assessment was studied. The knowledge artifact presented to the evaluator must

*Digital Quality Measures (dQMs) are technical specifications expressed in a machine interpretable format using highly standardized measure logic-Clinical Quality Language (CQL)- and data interoperability definitions -Fast Healthcare Interoperability Resources (FHIR).

have a high level of face validity to that person—which assumes the knowledge presents a complete and accurate picture of their expectations for the data's use. Data fitness evaluation is contextually dependent, meaning a dataset might be deemed sufficient for observational research while being deemed completely inadequate for quality measurement [2-6] therefore it was important to discern what aspects of PROM data structure and provenance must be adherent for the dataset to be considered adequate for performance measurement purposes.

Methods:

This project used a stepwise approach comprised of an environmental scan of published and grey literature and a series of 24 semi structured key informant interviews. An Endnote™ database was created for the purposes of documenting and classifying the environmental scan results. Google and Medline searches were conducted using terms including "Data quality assessment", "Data quality framework", "Data completeness", and "Data quality + electronic health records" to identify relevant published literature, grey literature and web material. Search parameters were set to ascertain knowledge published primarily within the last ten years, however older studies, publications, and posts were not automatically excluded from the review. Reference material was automatically imported from PubMed and WorldCat, whenever possible, for classification and grouping by domain. Materials identified in the search that were not available as a direct import were manually entered into the database.

The findings from the environmental scan informed the development of a semi structured interview guide intended to inform the refinement of the concept of PROM data fitness into a structured list of data quality assessment objectives (Appendix A-1). Stakeholder expertise was

sought to identify those key factors in contextual DQA indicating satisfactory fitness and PROM data content plausibility.

A purposive sample of key informants were selected, including perspectives from broad array of disciplines including clinical informaticists, public health informaticists, data intermediaries, data scientists, clinical outcome researchers, Chief Executive Officers (CEOs), Chief Medical Innovation Officers (CMIOs), Chief Medical Officers (CMOs), Quality Directors, epidemiologists, healthcare consultants and health policy experts. The study was primarily targeting decision makers; however, this does not necessarily equate to a data evaluator, rather recruitment of key stakeholders focused on identifying end users of DQA information for 1) policy decisions, 2) payment decisions, or 3) to inform quality improvement activity. A cohort of key informants was sought based on their active industry participation, although their expertise did not need to be restricted to specific experience with PROM data, just familiarity with patient reported outcomes in general. Key informants were recruited for the interviews using several listservs relevant to quality measurement including, but not limited to, attendees of the Digital Quality Summit, members of the National Committee for Quality Assurance (NCQA) Digital Measurement Community, the Health Effectiveness Data and Information Set (HEDIS) Audit Practice Leaders group, and from the investigator's LinkedIn contacts. It was important to study as wide an array of viewpoints possible to account for the variety of possible perspectives of DQA information utility within this specific context.

Requests to participate in the research were sent containing a summary of the study objectives, brief information on the interview process and a request for referral to other colleagues as deemed appropriate. Twenty-four (24) key informant interviews were conducted via

conference call, Zoom, or MS Teams depending on participant preferences. Key informants were asked to self-identify their role as a data owner, data consumer, data scientist, data requestor or other for the purposes of classifying their feedback. Conversations were recorded and the MP3 recordings were converted into a tagged JSON file using AWS Transcribe which was then edited into cleanly organized text for identification of key concepts. The files were reviewed and tagged by the study author to identify common elements that were mentioned as important to key informants in fitness determinations of PROM data.

Results

The environmental scan yielded a total of 631 references including presentations, white papers and other online materials whose abstracts or summaries were reviewed for relevancy to the specific topic of data quality and electronic clinical data (Table 1). Of these, 134 were found to be relevant to data quality and PROMs and were reviewed in greater depth to identify any reference to best practices for assessing data quality or data dimensionality, or any mention of a data quality assessment framework. 48 documents referenced a data quality framework or discussed best practices for data management or parameters for data fitness.

Published DQA models that harmonize terminology and frameworks for data quality assessment cover a wide range of descriptive data quality dimensions.[2-5, 7, 10-29] These models are ambiguous in many cases due to the large number of contextual uses for data quality assessment, each use case having its own unique requirements. These theoretical models primarily address the process of defining fitness through a comprehensive process of flagging inconsistencies in the structure of each data point, then categorizing those into a generic guideline for fitness assessment. Each evaluator is left to interpret the prespecified parameters

of the structural analysis, specifically for those data that present outside anticipated clinical normative ranges but still could have meaning in assessment of clinical performance.[7]

Kahn et al[2] and Weiskopf et al[7] both published data quality frameworks that were deemed to have applicability to this research. Their publications defined parameters for a data quality assessment framework (DQAF) for assessing electronic clinical data's secondary use in research which is very similar to the data's use for performance measurement.

Figure 1.1: Weiskopf's 3x3 DQA Guideline

	A: COMPLETE	**B: CORRECT**	**C: CURRENT**
1: PATIENTS	**1A** There are sufficient data points for each patient	**1B** The distribution of values is plausible across patients	**1C** All data were recorded during the timeframe of interest
2: VARIABLES	**2A** There are sufficient data points for each variable	**2B** There is concordance between variables	**2C** Variables were recorded in the desired order
3: TIME	**3A** There are sufficient data points for each time frame	**3B** The progression of data over time is plausible	**3C** Data were recorded with the desired regularity over time

*from Weiskopf et al[7]

Through the process of constructing her framework, Weiskopf formulated a *Desiderata for a Data Quality Assessment Guideline* outlining the ideals for any data quality assessment guide.[7]

According to Weiskopf's desiderata an ideal DQA guideline must be:

- Systematic and evidence- or expert-knowledge based.
- Flexible enough to accommodate the task-dependent nature of data quality.
- Engaging of users in assessment and decision making, rather than a black-box process.
- Independent from the availability of gold-standard data.

Ultimately, Weiskopf's summarized 3x3 DQA Guideline[7] was selected as the foundation for the key informant interview guide as it provided a framework for the very specific type of guidance that was being sought from the key stakeholders being interviewed.

Table 1: Data Quality Frameworks and Data Dimensionality

Search parameter	***# identified***
Data Quality	248
Data Quality Framework	48
Patient Reported Outcome	95
PROM	118
Data Consumer	5
Accuracy	42
Plausibility	5
Conformance	5
Validity	33
Trust	32

Key informants highlighted the importance of context, specifically around the mode of data collection, the environment (or scenario) in which it was collected and the persistence of all this contextual information through to the person evaluating the data's fitness for purpose. The level of detail of contextual provenance information necessary depended upon the individual informant's profession and/or use case for the dataset, but overall importance of persisting data provenance details alongside the information regarding data collection were generally said to be a priority in fitness evaluations. Maintaining the veracity and completeness of the provenance data at every ETL stage was also considered a chief DQA consideration by multiple stakeholders (Appendix B-1).

The interviews also highlighted the notion of end user "trust" as a central factor in assessing data quality and reinforced the need to define context-specific parameters specific to the end

user such that they would uniformly feel confident about a dataset's fitness for use in performance measurement. The prominent perspective regarding the contextual information was whether the data was collected by a clinician during an encounter (whether physical or virtual), completed immediately before and encounter (e.g., on a tablet in the waiting room), or at home prior to encounter. Reasons given for the importance of this information included the necessity of understanding whether a standard PROM instrument was used and how/where the observations were documented in the patient's medical record.

Several interviewees expressed a preference for clinician or observer-reported outcome data collection instead of PROMs as it enables clinicians to associate other interpretive findings related to the patient that could potentially be important for care planning. Key informants reported it is common for depression screening and other behavioral health questions and responses to be documented in the social history section of a patent's record which may confound electronic extraction of the information whose search parameters are restricted to diagnoses and observations from a problem list or primary medical history section. Interviewees expressed this as an oft-missed opportunity since information documented in different locations, while potentially very informative, could not be readily parsed or shared with other members of the care team.

Some of the important contextual metadata that key stakeholders thought to be important for "High" fitness assertion included:

1) Inclusion of patient directives (Patient's preferred language, end of life preferences, cultural accommodations)
2) Mode of PROM data collection (face to face, asynchronous via online form, other)
3) Use of standard, validated instrument to collect PROs

a) PROM type (behavioral, pain, Functional Status Assessment (FSA), other)

b) Instrument used, score and acceptable score parameters (valid range, threshold above which indicates F/U required)
c) Use of standard (preferred) terminology to document PROM results
4) Frequency of data points
a) how many times has the patient contributed data on each specific outcome directly to record.
b) prevalence of scoring or observations and verify that this does/does not have an effect on the representation and interpretation of scores)
5) Role\licensure of person documenting in record (patient, Care Manager, MD, RN, etc.)

This level of information was not identified by any key informant to be currently available in a consistent and easily retrievable fashion, however these were the most expressed preferences for the information that those interviewed would like to see persisting in the metadata of PROM datasets. When questioned about objectives for determining data quality, respondents expressed common experiences with specific inconsistencies in data completeness, frequent missingness of key variables and implausibility of variables outside normative clinical expectations, although when pressed on the existence of authoritative sources for data quality evaluation that might reinforce their decision making; none were able to point to a specific external source of information that informed their own intuitive data quality assessment practices.

Discussion

It is important to recognize that, from a data quality assessment perspective, PROM data fitness does not rely on merely the presence or absence of a questionnaire response or the instrument's score, rather it must also include the relevant log containing critical contextual and provenance information alongside it. In the context of PROM data collected during face-to-face encounters, the main considerations expressed by key stakeholders included 1) the use of

a standardized instrument 2) the location of the information documented in the patient's record and 3) the log of the encounter containing the metadata for the transaction. When standardized instruments are administered during face-to-face encounters, the issue of shareability of the information is subjugated by the accessibility of the information. In the cases of a clinician collecting this information as part of the clinical encounter, a concern was expressed that this information may end up included in the summary of the visit alongside all other information discussed in the form of free text clinical notes and therefore be unavailable for later retrieval. In these scenarios, a dQM query would fail to find the necessary information despite it being available in the electronic medical record, leading to misrepresented performance results.

This study focused on isolating those specific parameters of a quality assessment process (individual elements and numeric representation of acceptable limits) that would consistently characterize the fitness of an electronic dataset intended to be used for PROM performance measurement. For example: a scenario in which every single person in a large cohort of unsorted patients being the same height or age, would most likely rule out a dataset from any use regardless of the researcher's knowledge of a specific patient cohort. Other perhaps less obvious scenarios, such as the presence of over 2,000 spacecraft related incidents occurring within a cohort during the year 2020, rely on much more nuanced reference information to characterize its fitness for use. In the context of a worker safety monitoring program at Cape Canaveral, these may be plausible findings, however in the context of a general population's medical records, it signals to the end user that anomalies are present within the data. Finally, the most challenging plausibility scenarios concern the inclusion of 'outlier' data; or data that

are sufficiently within range of technically plausible values as not to be immediately ruled out, but their presence, particularly with high frequency, often creates uncertainty as to the fitness of the dataset.

Study Strengths and Limitations

This research did not address the task-level or data element-level (structural) determinations for data completeness or correctness, rather it examined the possible commonalities in the summary results generated from the execution of structural processes. The study focused on the ability to define a semi-universal set of criteria that could be used when examining PROM datasets for use in performance measurement. These perspectives were considered in the context of a single patient, which may or may not be scalable when extrapolated to population performance assessment. On a population-based performance measurement scale, this level of contextual information could be overwhelming due to the many meta variables considered important and the variation in the completeness of each on the context of the whole. Future research is needed to determine how one might harmonize or standardize these common intuitive assessments into a model that could apply to patient-reported outcomes, such that a bespoke construct need not be created for each performance measurement context.

Conclusion:

PROM data fitness assessment, from the perspective of a data fitness evaluation for quality measurement, must include a thorough appraisal of contextual and provenance information. Key informants stressed the absolute importance of specific contextual information to evaluate a dataset's fitness for use by dQMs. At the contextual level, the types of information that

govern the plausibility of PROM data are contained wholly within the evaluator's intuitive understanding of what is most explanatory for each intended use. Evaluators of data fitness frequently do not have the requisite contextual basis to enable a judgement on a particular dataset, rendering the data useless despite its potential to inform care quality. When questioned about objectives for determining data quality, key informants indicated that their own intuition about data quality issues played a major role in their determinations of a dataset's fitness. The main authoritative source for their evaluation was their own personal experience with these types of data.

References:

1. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of management information systems : JMIS*. 1996;12(4):5.
2. Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Washington, DC)*. 2016;4(1):1244. doi:10.13063/2327-9214.1244
3. Lee K, Weiskopf N, Pathak J. A Framework for Data Quality Assessment in Clinical Research Datasets. *AMIA Annual Symposium proceedings AMIA Symposium*. 2017;2017:1080-1089.
4. Feder SL. Data Quality in Electronic Health Records Research: Quality Domains and Assessment Methods. *Western journal of nursing research*. May 2018;40(5):753-766. doi:10.1177/0193945916689084
5. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. A Data Quality Ontology for the Secondary Use of EHR Data. *AMIA Annual Symposium proceedings AMIA Symposium*. 2015;2015:1937-46.
6. Jain ST, B. Data Lineage: An Important First Step for Data Governance. *BeyeNetwork* blog. December 18, 2013. http://www.b-eye-network.com/view/17023
7. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A Data Quality Assessment Guideline for Electronic Health Record Data Reuse. *EGEMS (Washington, DC)*. Sep 4 2017;5(1):14. doi:10.5334/egems.218
8. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med*. Dec 1 2009;151(11):793-803. doi:10.7326/0003-4819-151-11-200912010-00007
9. Williams SB, O'Connor EA, Eder M, Whitlock EP. Screening for child and adolescent depression in primary care settings: a systematic evidence review for the US Preventive Services Task Force. *Pediatrics*. Apr 2009;123(4):e716-35. doi:10.1542/peds.2008-2415
10. Brown JS, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. *Med Care*. Aug 2013;51(8 Suppl 3):S22-9. doi:10.1097/MLR.0b013e31829b1e2c
11. Callahan TJ, Bauck AE, Bertoch D, et al. A Comparison of Data Quality Assessment Checks in Six Data Sharing Networks. *EGEMS (Washington, DC)*. Jun 12 2017;5(1):8. doi:10.5334/egems.223
12. Dziadkowiec O, Callahan T, Ozkaynak M, Reeder B, Welton J. Using a Data Quality Framework to Clean Data Extracted from the Electronic Health Record: A Case Study. *EGEMS (Washington, DC)*. 2016;4(1):1201. doi:10.13063/2327-9214.1201
13. Estiri H, Stephens K. DQ(e)-v: A Database-Agnostic Framework for Exploring Variability in Electronic Health Record Data Across Time and Site Location. *EGEMS (Washington, DC)*. May 10 2017;5(1):3. doi:10.13063/2327-9214.1277
14. Hart R, Kuo MH. Better Data Quality for Better Healthcare Research Results - A Case Study. *Studies in health technology and informatics*. 2017;234:161-166.
15. Holve E, Kahn M, Nahm M, Ryan P, Weiskopf N. A comprehensive framework for data quality assessment in CER. *AMIA Joint Summits on Translational Science proceedings AMIA Joint Summits on Translational Science*. 2013;2013:86-8.
16. Huser V, DeFalco FJ, Schuemie M, et al. Multisite Evaluation of a Data Quality Tool for Patient Level Clinical Data Sets. *EGEMS (Washington, DC)*. 2016;4(1):1239. doi:10.13063/2327-9214.1239 17.

17. Huser V, Kahn MG, Brown JS, Gouripeddi R. Methods for examining data quality in healthcare integrated data repositories. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*. 2018;23:628-633.

18. IMF. *IMF Data Quality Assessment Framework*. Department PbtSDatPDaR; 2001. *Fourth Reveiw of the Fund's Data Standards' Initiatives*. https://www.imf.org/external/np/sta/dsbb/2001/070201.pdf

19. Kahn MG, Brown JS, Chun AT, et al. Transparent reporting of data quality in distributed data networks. *EGEMS (Washington, DC)*. 2015;3(1):1052. doi:10.13063/2327-9214.1052

20. 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. *Med Care*. Jul 2012;50 Suppl:S21-9. doi:10.1097/MLR.0b013e318257dd67

21. Knowlton J, Belnap T, Patelesio B, Priest EL, von Recklinghausen F, Taenzer AH. A Framework for Aligning Data from Multiple Institutions to Conduct Meaningful Analytics. *EGEMS (Washington, DC)*. Dec 15 2017;5(3):2. doi:10.5334/egems.195

22. Maxwell J. Conceptual Framework: What do you think is going on? *Qualitative Research Design: An interactive approach*. Sage Publications; 2012.

23. McClimans L. A theoretical framework for patient-reported outcome measures. *Theor Med Bioeth*. Jun 2010;31(3):225-40. doi:10.1007/s11017-010-9142-0

24. Naresh Sundar Rajan RG, Julio C. Facelli. A Service Oriented Framework to Assess the Quality of Electronic Health Data for Clinical Research. presented at: Healthcare Informatics (ICHI), 2013 IEEE International Conference; 2013;

25. Reimer AP, Milinovich A, Madigan EA. Data quality assessment framework to assess electronic medical record data for use in research. *International journal of medical informatics*. Jun 2016;90:40-7. doi:10.1016/j.ijmedinf.2016.03.006

26. Rogers JR, Callahan TJ, Kang T, et al. A Data Element-Function Conceptual Model for Data Quality Checks. *EGEMS (Washington, DC)*. Apr 23 2019;7(1):17. doi:10.5334/egems.289

27. Sebastian-Coleman L. Data Quality Assessment Framework. presented at: The Fourth MIT Information Quality Industry Symposium; 2010; Accessed March 2019. http://mitiq.mit.edu/IQIS/Documents/CDOIQS_201077/Papers/03_08_4B-1.pdf

28. Wang Z, Dagtas S, Talburt J, Baghal A, Zozus M. Rule-Based Data Quality Assessment and Monitoring System in Healthcare Facilities. *Studies in health technology and informatics*. 2019;257:460467.

29. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform*. Oct 2013;46(5):830-6. doi:10.1016/j.jbi.2013.06.010

Appendix A-1: Key Informant Semi-Structured Interview Guide

Thank you for agreeing to participate in my research and first I want to request your permission to record this interview to assist me with my notetaking. The recording will not be released publicly, nor will any statement you make today be attributed back to you unless you expressly desire it.

My questions for today are chiefly organized around two major themes:

First, patient reported outcomes and those "common" inconsistencies encountered during data capture and documentation, and

Second, the notion of what is important to "trust" the data for secondary use, such as performance measurement.

1. Could you please briefly describe your experience with patient reported outcomes? Describe your role in terms of data owner, data consumer, data scientist, data requestor, etc.…
2. How does your organization collect patient reported outcome data (specifically PHQ)? If so, in what format(s)?
3. Could you walk me through a typical workflow for the collection of PHQ data?
   [prompt: Is there guidance provided for data recorders/documenters?]
4. What are the primary reasons PHQ data might be used in a secondary context such as QM or QI in your org?
5. In terms of data plausibility, what do you look for to assure yourself that the data is adequate?
6. What are some of your main objectives when assessing the quality or fitness of PRO data?
   [Prompt: gap analysis, quality reporting,]
7. What specific steps or processes do you perform to assess for the presence of discrepancies in patient PRO data?
8. What are some of the most common types of errors you experience when working with PRO data? a. Data Completeness?
   a. Data Consistency?
   b. Data Plausibility?
9. When assessing data completeness, what do you do to assess missingness of critical data elements?
10. In your experience, what are some of the most common missing elements in patient-reported data?
11. How do you verify the consistency or the # of discrepancies within your data?
    [prompt: What tools are you using to quantify data discrepancies?]
12. Have you ever used open-source data quality tools such as Achilles, OMOP Data Quality Dashboard or FHIR Validator?
    [prompt: What tools do you generally use?]
13. In What measures must be satisfactorily completed for you to "trust" a dataset for use in PROMs?
14. the case of using this data for quality, what is the tolerance level for "data imperfections"? Is this a quantifiable concept or is it relative to the intent for its use?
15. Is there anything else that I haven't covered today that you can share with me about data quality and PRO data?

Appendix B-1: Parameters for Stakeholder Trust of PRO-PM Datasets

Parameter	Definition	Primary Considerations
Context	Context for use of information	Level of data quality required for patient administration should not define base standard as it is not necessarily worthless for other uses. Mismatch in expectations and perspectives between policymakers and the on the ground workflow considerations.
Measurement	Create measures to assess systems' ability to monitor contextual information	Discernable patterns: look for systematic errors and enforcement of mandatory fields Reliability of patient matching algorithm – patient corrected mismatched information
Use of standard instrument	Completion of instrument protocol: Yes/No	Data collected virtually and asynchronously, or in person • face-to-face w/clinician • face-to-face with MA • in waiting room during intake
Terminology	Data quality as codable concept	Standardize commonly used clauses to reduce ambiguity
Reusable knowledge	Uniform definitions	Creating shareable knowledge artifacts and repurposing existing protocols is essential, esp. in resource-poor systems
Face to Face collection	Associate other clinically relevant findings	If collected face-to-face, what components of the clinical reasoning process were documented along with the score and other observations made by the clinician.
Documentation access	Location determines findability and useability	Location – useable, locatable, feasible to access and process as part of the clinical decision process?
Alerts	Identifying inconsistencies during in-line monitoring of data quality.	Practical Specificity: by ignoring alerts you create a safety problem. Use best practices (accepted) objective info from past experience to define level of notification to prevent alert fatigue.
Information (data) log	Workflow and provenance	Presence/persistence and ability to review information persistence that is true to original form in the end user interface.
Context specificity	Use case for information retrieval	Status of DQA and the use cases Population level vs patient level decision making

Chapter 2: A Scoping Review to Identify Relevant Standards Defining Use of Patient Reported Outcomes in Quality Indicators.

Abstract

Introduction

Decisions regarding data fitness require explicit information about the context of use and those intrinsic values considered by evaluators to be necessary for data fitness decisions. Quality programs that utilize patient reported outcome measure (PROM) data must provide specific guidance and/or benchmarks relevant to its validation and fitness to implementors, ensuring reliability of the results. Rules setting the parameters for making effective fitness decisions are generally derived from experience in field and, in the case of national quality reporting programs, often subject to extensive public input, therefore they best represent the existing expectations that must met for PROM data to be considered fit for use in quality. These rules serve as the basis for audit of the datasets, ensuring transparency of the evaluator's expectations for both the quality and usability of the information represented by the dataset.

Methods

To identify existing requirements and benchmarks provided for quality programs, as well as any gaps in the existing guidance, a scoping study was conducted to define the overall existence for those resources that might be available to data fitness evaluators.[1,2] The objective was to identify existing rules and standards for PROMs which could be synthesized into a database that allows cross referencing of concepts as they are identified into thematic categories.

Results

A lack of any directly relevant information in programs specific to PROMs forced an expansion of the scoping definitions used to include patient generated health data (PGHD). Four resources stood out as being most relevant to addressing the research question driving this scoping review: The US Food and Drug Administration (FDA) *Guidance for Industry*[3]; the Office of the National Coordinator (ONC) *United States Core Data for Interoperability (USCDI)*[4]; the National Committee for Quality Assurance (NCQA) *HEDIS VOLUME 5: HEDIS Compliance Audit™: Standards, Policies and Procedures (HEDIS Vol. 5)*[5]; and the Observational Medical Outcomes Partnership's *Data Quality Dashboard*[6] (OMOP DQD).

Conclusion

There is a body of data quality research that could provide a foundation for a standardized PROM data quality assessment (DQA) model, however few national quality programs incorporate much, if any, of this in their program guidance. Clarifying data consumer expectations for PROM data consistency, completeness and plausibility through unambiguous requirements in quality program rules and regulations is needed to increase acceptance of PROM data's use for digital quality measurement.

Background

A standardized data quality assessment model that assures acceptable consistency and completeness of patient reported outcome measure (PROM) data is indispensable for accurately evaluating healthcare quality. For data to be useful when evaluating performance, it must be deemed "fit for purpose" by those end users seeking actionable information. Electronic clinical data quality assessment is contextually dependent, meaning the fitness of a dataset may be deemed sufficient for research while being completely inadequate for quality measurement.[7-11] PROM data is captured and stored in many different formats with varying degrees of accuracy and completeness. Kahn et al (2015) proposed that data quality assessment terms be organized into three major categories: 1) *Conformance*, representing the compliance of the data representation to formatting and relational definitions; 2) *Completeness*, describing the frequencies of attributes for data elements without contemplating their actual values; and, 3) *Plausibility*, which determines the level of trust in the element values' faithfulness in representing the intended construct.[7]

A major focus of existing data DQA models resides at the structural, or data element level, addressing the process of defining data quality through flagging inconsistencies in each data point then dichotomously categorizing as meeting/not meeting expected parameters. Evaluators are left to both interpret the pre-specified parameters of the structural analysis, and to evaluate whether the flagged issues are consistent with their own expectations- a process reliant on them having substantial prior experience in the field with the specific datatype. The necessity to draw upon personal experiences is particularly challenging when assessing PROM data because it frequently exists in many non-standard formats. While individual evaluators

may have varying objectives for understanding data fitness there is a common need to determine whether a dataset is suitable to sufficiently answer the questions they have when assessing care quality.

Decisions regarding data fitness require explicit information about the context of use and those intrinsic values considered by evaluators to be necessary to achieve a fit for purpose designation. Establishing clear expectations for data fitness provides effective guidance on how to preserve the necessary information as data is transformed into a dQM useable format. These rules serve as the basis for audit of the datasets, ensuring transparency of the evaluator's expectations for both the quality and usability of the information represented by the dataset. Quality programs that utilize PROM data must provide specific guidance and/or benchmarks relevant to the validation processes for the data, providing implementors with a consistent set of benchmarks to reference. Rules setting the parameters for making effective fitness decisions are generally derived from experience in field and, in the case of federal regulation, often subject to extensive public input-particularly in the case of quality reporting programs, therefore they best represent the existing expectations that must be met for PROM data to be used in quality programs.

Methods:

To identify existing requirements and benchmarks provided for quality programs, as well as any gaps in the existing guidance, a scoping study was conducted to define the overall existence for those resources that might be available to decision makers. Arksey and O'Malley (2005) define a scoping study framework in 5-stages:[1]

Stage 1. Identifying the research question
Stage 2. Identifying relevant studies
Stage 3. Study selection
Stage 4. Charting the data
Stage 5. Collating, summarizing and reporting the results

1. Identifying the research question

The research question this scoping study intended to answer was specifically designed to identify existing resources that are available to guide implementors intending to use PROM data in quality reporting programs. The objective was to identify existing rules and standards for PRO data use which could be synthesized into a database that allows cross referencing of the concepts as they are identified into the thematic categories.

2. Identifying relevant studies

The scan of existing rules and other regulatory guidance was performed to identify and document any existing information on the use of PROM data or requirements for the evaluation of its fitness. The scoping review was intentionally limited to a search for terms pertaining specifically to the use of PROMs and information natively published in English using Pubmed for published literature and Google for grey literature. A lack of any directly relevant information in programs specific to PROMs forced an expansion of the scoping definitions used to include patient generated health data (PGHD). Search parameters were set to ascertain knowledge

published primarily within the last four years; however older references were not automatically excluded from the review. A keyword search[12] was performed using the terms “person driven outcomes”; “patient reported information”; “patient reported outcome” or its acronym “PRO” and “patient reported outcome measure” or its acronym “PROM”[13]; “patient generated health data” or the acronym “PGHD”[14] and “data quality”, “quality assessment” or “quality measure” to identify any peer-reviewed journal articles that offered insights into patient reported outcome data use in quality programs. Citations identified as possessing at least one of the terms for PROMs and quality reporting were imported along with the pdf publication into a citation management application (EndNote 20). Reference material was automatically imported into Endnote whenever possible, and those materials identified in the search that were not available as a direct import were manually entered.

A Google search was subsequently performed to identify websites, blog posts, podcast transcripts, white papers or reviews that mentioned person reported health information or referenced published regulations pertaining to this type of data. Primary consideration was given to those documents that were intended as an educational resource and included information on PROMs and the rules pertaining to their use in quality programs. All web materials were converted into static documents amended with provenance information then imported into EndNote 20 for further evaluation.

3. Study selection

Documents were reviewed to elicit specific information about existing rules and concepts that might be relevant to the development of a trust framework for use of PROM for performance measurement. The documents were scanned to identify instances where relevant data quality

assessment issues were explicitly discussed. Those concepts identified from the documentation and deemed to potentially have sufficient relevance were added to the concept mapping table for further evaluation and refinement. As additional concepts were identified in the environmental scan, they were continuously added to the list, cross referenced against OMOP DQD for alignment of the concept definition and DQA Framework domain. The environmental scan was managed using Endnote 20 and imported into MS Excel for concept mapping. Documents were classified into two major categories: 1) *Research* and 2) *Requirements*. The former being articles, presentations, commentary and/or other educational materials that discussed specific aspects of the latter. The *Requirements* category was further divided into sub-categories including *Tools, Regulation, Guidance, Program Requirements, and Frameworks.* A "Rule" considered for inclusion is defined as any program requirement that explicitly defined the criteria for use of PROM data. Tools were classified under requirements as they produce data quality assessment information based on a standard set of end user requirements; therefore, tool requirements reflect the operational aspects of the existing standards. Rules for private programs that were accessible to the researcher as well as publicly available federal and state regulation were included in the database wherever possible. Some of the criteria was applied iteratively during the scan as each program identified offered additional criteria that was then applied back to the earlier findings when assessing their relevance.[1]

4. Charting the Data

A sortable database was created using MS Excel to organize the developing list of quality reporting program guidance into a cohesive structure such that each could be mapped to data quality framework definitions and cross referenced with the information gleaned from

informed stakeholders in AIM1. The initial list of concepts derived from AIM1 resulted in a list of ten common trust parameters and considerations for making decisions about PROM data quality. The excel file allowed for 'charting' of key concepts and terms used in quality reporting programs that can be derived from existing program guidance as well as any association of multiple relevant attributes to each under review.[15] Criteria for the program elements extracted for inclusion in the database were directly related to the use of PROM data for quality measurement or improvement, details of prescriptive processes required to validate PROM data, descriptive parameters for the secondary use of PROM data, and benchmarking information for PROM data's use in a program (i.e., sensitivity or specificity thresholds, operational testing parameters for PROM data composition, and named processes for validating PROM data). Several concept types were considered under this classification schema including element-level data, statistical parameters as well as narrative information on data quality assessment (Appendix A-2).

5. Collating, summarizing, and reporting the results

As the various program guidance elements were added to the excel database, they were assigned classifications by their concept 'Type' to organize them in such a manner to comprehend potential dependencies between the kinds of information when making fitness decisions. Concept type classification is important as many common data model (CDM) definitions are limited to data element-level concepts and do not encompass contextual or summary information in a standardized fashion. The list of concepts was then classified by the domains of the data quality categories outlined in the Kahn harmonized framework.[7] Many

concepts on the list applied to several DQA framework domains, thus all relevant domains were included adjacent to each concept to ensure all DQA relationships were represented. Clarification on specific program rules was sought whenever possible, findings were validated by consulting with sources directly involved with the programs under study to enhance our understanding of the extent to which the program guidelines were applicable.[16,17]

First, each annotated concept name was reviewed against the OMOP Common Data Model v5.4 to standardize the representation of each concept as a commonly accepted definition. The OMOP CDM was chosen as the primary source for concept definitions due to its applicability to a wide variety of data types and its extensive documentation on data quality that was assembled using the Kahn harmonized framework.[18]

Results

An initial scan of existing program requirements using a keyword search for PROMs yielded a lack of any directly relevant information in programs specific to PROMs forcing an expansion of the scoping definitions used to include patient generated health data (PGHD). The scoping review yielded 65 unique documents identified as potentially relevant to the topic of data quality. Of these, 26 were determined to be directly related to assessment of data quality for PROMs or PGHD. A total of 22 resources were identified from these documents and were categorized into five domains (Table 2.1).

Table 2.1: Programs Identified with Requirements Pertaining to the Use of Patient-Reported Outcome Data

Tools for assessing Data Quality:	Cypress[19], DQe-c[20], Achilles[21], Observational Medical Outcomes Partnership Data Quality Dashboard (OMOP DQD)[6]
Regulation Pertaining to Patient Reported Outcome Data:	Office of the National Coordinator for Health Information Technology Trusted Exchange Framework and Common Agreement (ONC TEFCA)[22], United States Core Data for Interoperability (USCDI)[4]
Guidance for Use of Patient Reported Outcome Data:	FHIR Resource Validation[23], Food & Drug Administration Guidance for Industry[3], Health Level Seven (HL7) QI Core Implementation Guide[24], Oregon Health Sciences University Patient Entered Data[25], ONC Patient Generated Health Data (PGHD)[26], ONC Patient Engagement Playbook[27], Defining and Measuring Completeness of Electronic Health Record Data[28], Data Quality Assessment Guidelines for EHR Data Reuse[29]
Program Requirements Related to Use Patient Reported Outcome Data:	NCQA Audit Process[30], HEDIS MY 2020 Volume 5[5]
Data Assessment Frameworks:	US Data Access Framework[31], International Monetary Fund Data Quality Assessment Framework[32], OMOP Common Data Model[18], The Entity-Attribute-Value Model[33], Data Element Function Model[34], Coleman Data Quality Assessment Framework[35]

Four resources stood out as being directly relevant in addressing the research question driving this scoping review: The US Food and Drug Administration (FDA) *Guidance for Industry*[3]; the Office of the National Coordinator (ONC) *United States Core Data for Interoperability (USCDI)*[4]; the National Committee for Quality Assurance (NCQA) *HEDIS VOLUME 5: HEDIS Compliance Audit™: Standards, Policies and Procedures (HEDIS Vol. 5)*[5]; and the Observational Medical Outcomes Partnership's *Data Quality Dashboard*[6] (OMOP DQD).

The FDA offers several resources providing guidance for the industry on patient reported outcome measures, however their focus is on their use for medical product development.[36,37] ONC publishes USCDI, a standard set of data classes and element definitions targeted for the exchange of health information.[4] USCDI does not currently define PROM data as part of its

standard set. USCDI v2 standard was out for public comment at the time of this search and the comments posted directly addressed a lack of any guidance on PROM data.

"*A gap in USCDI v2 is **patient reported information/outcomes data**, which represents a critical data source (the patient). Data reported by the patient provides unique and important context about the patient and their health status has been identified as critical information to consider during care. Patient reported information can be defined as structured data captured that comes directly from the patient, related to the status of a patient's health condition. This concept is represented in both USCDI version 2 submissions related to observation codes/values and questionnaires (captured by way of an observation or a questionnaire/questionnaire responses). Although these data may be less standardized than other data requested for USCDI consideration, this is an area where the requirements should push the digital capture and standardization of the data forward."*[4]

NCQA-certified HEDIS auditors adhere to a set of rules for validation and certification of data published by the National Committee for Quality Assurance (NCQA). *HEDIS Measurement Year 2020 VOLUME 5: HEDIS Compliance Audit™: Standards, Policies and Procedures* (HEDIS Vol. 5)[5] includes all the information HEDIS auditors require–including standards by which to evaluate information and the guidelines for the process of verifying it–to evaluate whether a payer's data is sufficient for HEDIS Health Plan reporting. HEDIS Vol. 5 was reviewed in detail for those standards and guidelines directly relevant to, or associated with, data that comes directly from a health plan member whether it came first via a provider or was sent to the payer directly by the member.

The OMOP Data Quality Dashboard[6] (OMOP DQD) is a parameterized data check process that is most useful as it provides a summary of the thousands of checks that are typically run and therefore is informative for data fitness evaluation. The OMOP DQD check-type concepts present information as "The number and percent of records with a value in *XX* field which is less than the lowest expected plausible value…" This provides the end user with a very specific profile of potential issues that creates relevant relationships between the various potential data errors which can be customized (within the DQA framework parameters) to each use case.

Discussion

NCQA's HEDIS program has sophisticated methods to ensure the accuracy and quality of data used through the *HEDIS Compliance Audit* which validates all forms of data for use in HEDIS.[5] In this specific context, NCQA certified auditors are responsible for assuring the integrity of both structured and unstructured clinical data included in quality measure reports through a rigorous process called Primary Source Verification (PSV). PSV evaluates both the chain of data stewardship as well as the accuracy of the information represented in the target dataset, allowing HEDIS Auditors to assure the accuracy of the information by understanding all aspects of data entry, editing and manipulation.[30] Individual organizations who might normally be unwilling to share specifics about their internal DQA processes are required to expose them to their HEDIS auditor for verification, otherwise they run the risk that their data will not be approved for inclusion.[38]

Even though the NCQA guidance represents the data quality assessment gold standard for quality measurement, there is minimal guidance in HEDIS for the use and validation of PROM data. The PSV process relies heavily on each auditor's experience and training to evaluate

copious volumes of health data, and this has led to some criticism of the approval of various datasets for HEDIS reporting. The complete guidance offered to HEDIS Auditors is detailed in Appendix B-2.

HEDIS currently has 5 PRO-PMs whose PRO submissions are annually assessed by HEDIS auditors. Despite this, auditors are making decisions on the data's fitness with extremely limited guidance from NCQA and must, therefore, rely heavily on their own judgements when validating its acceptability for HEDIS Health Plan Reporting. PROM data have clearly not yet achieved an equivalency status with administrative claims in terms of acceptability or usability, evident in the lack of effective guidance to those assessing the veracity of PRO datasets. While there is a long history of using administrative claims data for assessing healthcare quality at the national level,[39] the lack of sufficient detail in these data necessitates more comprehensive information be integrated into the program infrastructure.

Conclusion

There is a body of data quality research that could provide a foundation for a standardized PROM data quality assessment model, however few national quality programs incorporate much, if any, of this in their program guidelines. Clarifying data consumer expectations for PROM data consistency, completeness and plausibility through unambiguous requirements in quality program rules and regulations is needed to increase acceptance of PROM data's use for digital quality measurement.

References

1. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*. 2005;8(1):19-32. doi:10.1080/1364557032000119616

2. Tiase VL, Hull W, McFarland MM, et al. Patient-generated health data and electronic health record integration: protocol for a scoping review. *BMJ Open*. Dec 17 2019;9(12):e033073. doi:10.1136/bmjopen-2019-033073

3. FDA. *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. 2009. *FDA Guidance Documents*. Accessed September 20, 2019. https://www.fda.gov/media/77832/download

4. HealthIT.gov. United States Core Data for Interoperability (USCDI). The Office of the National Coordinator for Health Information Technology (ONC). Accessed April 2, 2021. https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi

5. NCQA. *HEDIS MEASUREMENT YEAR 2020 VOLUME 5: HEDIS COMPLIANCE AUDIT™: STANDARDS, POLICIES AND PROCEDURES*. National Committee for Quality Assurance; 2020. Accessed March 22, 2021. https://store.ncqa.org/hedis-my-2020-volume-5-epub.html

6. OHDSI. The Data Quality Dashboard Observational Health Data Sciences and Informatics. Accessed Oct. 20, 2020. https://data.ohdsi.org/DataQualityDashboard/

7. Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Washington, DC)*. 2016;4(1):1244. doi:10.13063/2327-9214.1244

8. Lee K, Weiskopf N, Pathak J. A Framework for Data Quality Assessment in Clinical Research Datasets. *AMIA Annual Symposium proceedings AMIA Symposium*. 2017;2017:1080-1089.

9. Feder SL. Data Quality in Electronic Health Records Research: Quality Domains and Assessment Methods. *Western journal of nursing research*. May 2018;40(5):753-766. doi:10.1177/0193945916689084

10. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. A Data Quality Ontology for the Secondary Use of EHR Data. *AMIA Annual Symposium proceedings AMIA Symposium*. 2015;2015:1937-46.

11. Jain ST, B. Data Lineage: An Important First Step for Data Governance. *BeyeNetwork* blog. December 18, 2013. http://www.b-eye-network.com/view/17023

12. College C. Research Skills Tutorial: 1.2 Searching with Keywords. Columnbia College. 2021. https://columbiacollege-ca.libguides.com/research_skills/getting_started/keywords

13. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights*. 2013;6:61-8. doi:10.4137/HSI.S11093

14. Plastiras P, O'Sullivan D. Exchanging personal health data with electronic health records: A standardized information model for patient generated health data and observations of daily

living. *International journal of medical informatics*. Dec 2018;120:116-125. doi:10.1016/j.ijmedinf.2018.10.006

15. Ritchie JaS, L. Qualitative Data Analysis for Applied Policy Research. In: Bryman AaB, R.,, ed. *Analyzing Qualitative Data*. 1994:173-194.
16. Oliver S. Making research more useful: Integrating different perspectives and different methods. 2001:
17. Talbot W. Confirmation on the General Guidelines and HEDIS Audit Standards for Verification and Use of Person-Reported information in HEDIS Reporting. In: Hamlin B, editor. verification of the HEDIS General Guidelines and HEDIS Compliance Audit Standards that pertain to the use and validation of person-reported infomation in HEDIS ed2021.
18. OHDSI. OMOP Common Data Model. The Observational Health Data Sciences and Informatics (OHDSI). 2019. https://www.ohdsi.org/data-standardization/the-common-data-model/
19. CMS. cypress: Rigourous and repeatable testing of Electronic Health Records. Accessed September 12, 2019. https://www.healthit.gov/cypress/
20. Estiri H. DQe-c. GitHub. 2019. https://github.com/hestiri/DQe-c
21. OHDSI. ACHILLES for data characterization. Observational Health Data Sciences and Informatics (OHDSI). Accessed January 5, 2019. https://www.ohdsi.org/analytic-tools/achilles-for-data-characterization/
22. ONC. A User's Guide to Understanding Trusted Exchange Framework and Common Agreement (TEFCA) Draft 2. The Office of the National Cooridnator for Health Information Technology; 2019. Accessed June 2019. WWW.HEALTHIT.GOV/TEFCA
23. HL7. Validating Resources. HL7 International. Accessed July 2019, 2019. https://www.hl7.org/fhir/validation.html
24. HL7_International. QI-Core Implementation Guide: STU 4 (v4.0.0 for FHIR 4.0.1). 2020. http://hl7.org/fhir/us/qicore/index.html
25. OHSU. Patient entered data. American Medical Informatics Association. 2021. https://www.clinfowiki.org/wiki/index.php/Patient_entered_data
26. ONC. Patient-Generated Health Data. Healthit.gov. Accessed March 21, 2021. https://www.healthit.gov/topic/scientific-initiatives/patient-generated-health-data
27. ONC. Patient Engagement Playbook. Office of the National Coordinator for Health IT (ONC). Updated 2019. Accessed March 21, 2021. https://www.healthit.gov/playbook/pe/
28. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform*. Oct 2013;46(5):830-6. doi:10.1016/j.jbi.2013.06.010
29. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A Data Quality Assessment Guideline for Electronic Health Record Data Reuse. *EGEMS (Washington, DC)*. Sep 4 2017;5(1):14. doi:10.5334/egems.218

30. NCQA. Audit Process: Offsite Methods. *HEDIS 2019 Volume 5: HEDIS Compliance Audit™: Standards, Policies and Procedures*. 2019 ed. National Committee for Quality Assurance; 2019. *The Healthcare Effectiveness Data and Information Set*.
31. HL7.org. U.S. Data Access Framework (DAF) FHIR Implementation Guide (IG). HL7 International. Accessed October 16, 2018. http://hl7.org/fhir/2015May/daf.html
32. IMF. *IMF Data Quality Assessment Framework*. Department PbtSDatPDaR; 2001. *Fourth Reveiw of the Fund's Data Standards' Initiatives*. https://www.imf.org/external/np/sta/dsbb/2001/070201.pdf
33. Raszczynski R. The EAV Model. INVIQA; 2010. 2010. Accessed July 2019. https://inviqa.com/blog/eav-data-model
34. Rogers JR, Callahan TJ, Kang T, et al. A Data Element-Function Conceptual Model for Data Quality Checks. *EGEMS (Washington, DC)*. Apr 23 2019;7(1):17. doi:10.5334/egems.289
35. Sebastian-Coleman L. Data Quality Assessment Framework. presented at: The Fourth MIT Information Quality Industry Symposium; 2010; Accessed March 2019. http://mitiq.mit.edu/IQIS/Documents/CDOIQS_201077/Papers/03_08_4B-1.pdf
36. FDA. *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims* Administration FaD; 2009. *Guidance for Industry*. https://www.fda.gov/media/77832/download
37. Administration UFaD. Virtual Public Meeting - Patient-Generated Health Data Throughout the Total Product Life Cycle of Medical Devices. 2021:
38. Newsletter TDA. Metadata in Data Governance. Dec 18, 2012. http://tdan.com/metadata-in-data-governance/16101
39. Callahan TJ, Bauck AE, Bertoch D, et al. A Comparison of Data Quality Assessment Checks in Six Data Sharing Networks. *EGEMS (Washington, DC)*. Jun 12 2017;5(1):8. doi:10.5334/egems.223

Appendix A-2: Data Quality Assessment Concept Types and Definitions

Table A-1: Data Quality Assessment Concept Types and Definitions

Type	Definition
Data Element	Individual data points that must be present to effectively assess a dataset's fitness for use.
Test & Parameter	The preferred analytic processes and their respective pre-specified thresholds typically deployed to produce actionable knowledge allowing an end-user to make decisions about a dataset's fitness. *Tests* are frequently statistical analysis *Parameters* are those configurable thresholds for each test tailored to each end-user's preferences for a specific use case.
Summary Information	A summary result or descriptive report that helps inform an end-user's understanding of the data. These summary reports typically include data check results alongside other supplemental information on the dataset's assets. A report is considered a concept for this study as it informs the end-user's decision on the quality and utility of the dataset.

Table A-2: Concepts to be considered for PROM Trust Framework

Concept Name	Concept Description	Concept Type	Data Quality Domain
Database Format	Format of source dataset used to assess PGHD fitness (e.g., FHIR, JSON, SQL, CQL, etc.)		Conformance
Database Standard Format	A yes or no value indicating the dataset used to evaluate PGHD fitness is stored in a standard format that is useable, locatable, feasible to access and process	Test & Parameter	Conformance
Database Standards Use	Version of standards used to store information (e.g., FHIR 4.0.1, CQLv1.5.1, etc.)	Summary Information	Conformance
Database Format Interoperability	A yes or no value indicating the data format used meets ONC interoperability requirements for data exchange internally and externally (e.g., USCDI)	Test & Parameter	Conformance
Patient Demographics	The number and proportion of records with a value present in Patient Demographics fields (e.g., age, gender, ethnicity, primary language, culturally relevant information)	Data Element	Conformance Completeness Plausibility
Patient Directives	The number and proportion of records containing documentation of patient choices, (consent) permitting or denying recipient(s) or recipient role(s) to perform one or more actions within a given policy context, for specific purposes and periods of time.	Data Element	Conformance Completeness Plausibility

Concept Name	Concept Description	Concept Type	Data Quality Domain
Data Source Type	The number and proportion of records containing source of PGHD information (person, medical device, wearable device, other)	Data Element	Conformance Completeness Plausibility
Data Source Provenance	The number and proportion of records containing metadata on the activities taken upon the source information to transform to the target dataset. Provenance is prepared by the application that initiates the create/update of the PGHD	Data Element	Conformance Completeness Plausibility
Audit Log: Data Record Management	A yes or no value indicating the presence of the records management log assuring privacy and security safeguards are maintained and functioning.	Test & Parameter	Conformance
Audit Log: Data Quality Assessments	Number and type of measures deployed that are monitoring the quality of the target data at any point in time (ratio of data to errors, ETL transformation errors rate, data time to value, mandatory field enforcement, number of in-line corrections)	Summary Information	Conformance
Audit Log: DQA Report	Presentation of DQ measure results exposing discernable patterns of data inconsistencies and/or systematic errors	Summary Information	Plausibility
Audit Log: Provenance	A yes or no value indicating the presence of provenance information	Test & Parameter	Completeness
Audit Log: Persistence	A yes or no value indicating the persistence of provenance information that verifies the data is true to original form in the end user interface.	Test & Parameter	Conformance
Encounter Type	The number and proportion of records documenting the type of encounter during which patient reported outcome data was generated	Data Element	Completeness
Encounter Location	The number and proportion of records documenting the location where the patient reported outcome data collection takes place and/or the mode of data collection (face to face, asynchronous via online form, other)	Data Element	Completeness
Encounter class	The number and proportion of records documenting Period (Date/Time) information for Encounter Type	Data Element	Completeness
Encounter reference	The number and proportion of records relating the frequency of patient contributions to their health record for any patient reported outcome data element. (e.g., how many times PHQ-9 appears within a certain period for an individual.)	Data Element	Completeness Plausibility

Concept Name	Concept Description	Concept Type	Data Quality Domain
Encounter performer	The number and proportion of records containing the role\licensure of person documenting or collecting the PGHD in record (patient, Care Manager, MD, RN, etc.)	Data Element	Completeness
Questionnaire	The number and proportion of records documenting standard instrument type used to collect patient reported outcome data	Data Element	Completeness Conformance
Questionnaire Data Type	A yes or no value indicating if the Questionnaire is the expected data type based on the technical specification.[40]	Data Element	Conformance
Questionnaire Response Value	The number and proportion of records documenting presence of total score result indicating complete administration of questionnaire-asserting the structured group of questions that comprise the instrument have been successfully answered and a valid total score was calculated and present in the record.	Data Element	Completeness Plausibility
Questionnaire Response metadata	The number of records that contain supplemental information on acceptable use parameters for the instrument used (legal rights of use info, valid scoring range, positive thresholds, etc.).	Data Element	Completeness
Questionnaire Standard Terminology	The number and percent of records that have a standard, valid term present in all Questionnaire and Questionnaire Response fields as specified. (Use of preferred standard terminology to document instrument and result.)	Data Element	Conformance Plausibility
Questionnaire Response Null Value	The number and percent of records with a NULL value in the QuestionnaireReponse value field (considered not nullable)[40]	Data Element	Completeness Conformance Plausibility
Questionnaire Response Zero Value	The number and percent of records with a value of 0 in the QuestionnaireReponse value field[40]	Data Element	Completeness Conformance Plausibility
Questionnaire Response: Missing Value	The number and percent of records with missing QuestionnaireReponse values	Data Element	Completeness Plausibility
Questionnaire Response: Value Threshold	The prespecified threshold (expected) for the QuestionnaireReponse value field	Test & Parameter	Conformance
Questionnaire Response: Value Low	The number of records that have a value that is lower than a pre-specified threshold.[40]	Data Element	Conformance Plausibility
Questionnaire Response: Value High	The number of records that have a value that is higher than a pre-specified threshold.[40]	Data Element	Conformance Plausibility

Concept Name	Concept Description	Concept Type	Data Quality Domain
Data Quality: Thresholds	The reference to a technical specification or relational dataset that defines pre-specified thresholds asserting sufficient accuracy and precision for any specific use case	Test & Parameter	Conformance
Data Quality: Context Specificity	Use case pre-defined for correction and enhancement of datasets linked to its intended use.	Test & Parameter	Conformance
Data Quality: Adherence to Target	A yes or no value indicating target dataset complies with the pre-specified end-user expectations for intended use case	Summary Information	Conformance
Data Quality Test Type	Test type used to ascertain dataset adherence	Test & Parameter	Conformance
Dataset Governance Reference	Reference information used in the assessment and standardization of information within the dataset	Test & Parameter	Conformance
Data Quality: Dataset Governance Process	Documentation of the process that characterizes the necessity of data enhancement based on standard reference information.	Summary Information	Conformance
Data Quality: Measurement Report	A report that characterizes how well source data transformation to target dataset meets end user expectations.	Summary Information	Conformance Plausibility
Data Quality Criterion	What criterion/test is used to ascertain data dependencies?	Test & Parameter	Conformance Plausibility
Data Quality Dependency Report	Report of data dependencies results	Summary Information	Conformance Plausibility
Actionability	What do errors and inconsistencies signify		
Workflow	Workflow practices that must exist for end users to “trust” the information is accurate and can be used for clinical quality		

Appendix B-2: Guidelines for the Use of Member-reported[†] Data by Certified HEDIS Auditors

Table 2.4. Guidelines for the Use of Member-reported[‡] Data by Certified HEDIS Auditors

HEDIS General Guidelines for Data Collection and Reporting[41] (HEDIS Volume 2)

38. Member-Collected Samples

Test results from member-collected samples may be used for FOBT, urinalysis testing and blood spots for HbA1c, LDL-C, glucose and total cholesterol. Member-collected samples must be sent to the laboratory or provider's office for analysis.

39. Member-Reported Services and Biometric Values

Member-reported services and biometric values (height, weight, BMI percentile) are acceptable only if the information is collected by a primary care practitioner (refer to Appendix 3 for the definition of "PCP") or specialist, if the specialist is providing a primary care service related to the condition being assessed, while taking a patient's history. The information must be recorded, dated and maintained in the member's legal health record. Note: It is a "best practice" to collect data directly from members for the Language Diversity of Membership and Race/Ethnicity Diversity of Membership measures; therefore, a PCP or specialist is not required to collect this information as part of patient's history for these measures only

HEDIS Compliance Audit Standards: IS Standards[5] (HEDIS Volume 5)

IS 5.0 Supplemental Data—Capture, Transfer and Entry

IS 5.1 Nonstandard coding schemes are fully documented and mapped to industry standard codes.

- Mapping documents show that all nonstandard codes and code systems are identified and mapped according to the requirements in the specifications.
- Program code ensures mapping documents are executed accurately.

IS 5.2 The organization has effective procedures for submitting measure-relevant information for data entry. Electronic transmissions of data have checking procedures to ensure accuracy.

- Policies, procedures and log forms for data submission ensure accuracy and completeness and verify that the organization has mechanisms for transferring information to the appropriate location within the organization.
- Forms—including samples of completed forms, policies, procedures and instructions for completing the forms—ensure all fields relevant to measure reporting.

[†] HEDIS specifications and general guidelines use "member-reported" to describe any patient reported or person generate information. The General Guidelines do not include any specific designation distignuishing Patient/person/clinician/etc.

[‡] HEDIS specifications and general guidelines use "member-reported" to describe any patient reported or person generate information. The General Guidelines do not include any specific designation distignuishing Patient/person/clinician/etc.

- Electronic file formats and protocols ensure capture of all data fields listed in the Roadmap. Policies and procedures for collecting supplemental data specify:
  - Exclusions are not collected for previous reporting years for members with clinical conditions that can change.
  - Information obtained by the provider's office or clinician directly from the member was entered in the medical record by the deadline established for the measure.
  - Information obtained by the provider's office or clinician directly from the member is verified when taking a patient history of a disease management system.
  - Information obtained from a simple provider attestation is not used.
  - Information obtained from member surveys is not used.
- Policies and procedures for submission and transmission of electronic information:
- The organization effectively monitors the quality and accuracy of its electronic submissions.
- Transmissions are properly controlled by logs, record count verification, redundancy checking receipts, retransmissions and sign-offs.

IS 5.4 The organization continually assesses data completeness and takes steps to improve performance.

- Policies, procedures and performance standards require:
  - Complete submission and entry of data.
  - Proper control of transmissions by logs, record count verification, redundancy checking receipts, retransmissions and sign-offs to ensure that all data are received.
- Contracts with vendors require data for measure reporting and provide inspection and auditing of data, correction and resubmission of data and backlog control standards and procedures.
- Policies, procedures and performance standards require reconciliation of data between the originating system and the repository.

Chapter 3: A Consensus Data Quality Assessment Model Defining Trustworthiness of Patient-Reported Outcome Measures (PROMs) for Performance Measurement.

Abstract

Introduction

Fitness assessment of datasets intended for patient-reported outcome performance measurement (PRO-PM) requires clear guidance on what decision makers are seeking when determining the useability of PROM data. The guidance must explicitly define minimum characteristics a dataset must exhibit to be considered fit for a specific purpose. Each purpose has differing objectives for defining trustworthiness which presents some difficulty when developing uniform guidance for assessing the credibility of data. This study set out to answer several essential questions: First, to identify what specific information must be present for evaluators to effectively assess PROM data quality; and second, to understand what external references are relied upon when making judgements about the usability of PROM data for the purposes of performance measurement.

Methods

A modified Delphi process to obtain structured feedback from a group of diverse stakeholders was used to assess the relevance and importance of specific data quality concepts when making decisions about data fitness. The process sought to narrow down the list of concepts to a manageable size while still adequately representing the varied perspectives of decision makers on a multidimensional issue. A purposive sample of informed stakeholders in the field of data quality, PRO-PMs and clinical decision support were invited to participate in the development of a trust model for PROM data. A three-step process was used to identify and prioritize the

attributes of PROM data and concurrently assess their importance in the context of data fitness decisions.

Results

A total of 40 individual data quality concepts were initially reviewed. Of these 40 concepts presented for evaluation, 19 received a not relevant (N/A) designation, 10 of those 19 receiving greater than two N/A votes from individual respondents. Respondents strongly agreed on the importance of 22 of the initial 40 concepts, and only moderately agreed on the importance of the remaining 18 concepts. None of the original 40 concepts were calculated as having poor agreement amongst respondents. Using this information, a list of 25 concepts presented to a panel of subject matter experts was further refined to a list of 18 core concepts organized into 8 key domains.

Conclusion

It is rare for decision-makers to have guidance available to them when assessing the fitness of PROM data. They frequently rely upon a variety of statistical and operational test results to make these decisions and are often presented with seemingly inconsistent information which requires substantial expertise to interpret. A uniform, well-defined list of PROM data quality concepts provides much needed information that would enhance the consistency and reliability of data fitness decisions. Despite varying objectives, decision makers can agree on a model providing uniform parameters for evaluating data fitness that are central to their decisions about the useability of PROM data.

Background

A standardized data quality assessment model that assures acceptable consistency and completeness of patient reported outcome measure (PROM) data found in electronic health records is indispensable for accurately evaluating healthcare quality. Standardized clinical data are those data which have undergone a process by which it is transformed into a format that can be processed or shared electronically between systems of care.[1,2] Standard clinical data is that which is stored using common interoperability definitions using preferred terminologies (codes) that represent the stored knowledge in a manner that can be used for clinical decision making.[2] A patient reported outcome (PRO) is a report of a person's health that is documented directly as it was reported by them without any interpretation of the person's response by a member of the care team.[3] Patient reported outcome measures (PROMs) are the standardized instruments used to collect PROs generally through the administration of self-report questionnaires.[3] If outcomes such as symptoms, functional status, or well-being are documented based on the report of a caregiver or health professional, they are not considered PROs, but observer-reported outcomes or clinician-reported outcomes, respectively.[4] There are many modalities for administering PROMs including paper questionnaires, web-based applications (web apps) or by the patient directly entering responses into their electronic medical record form. Variation in PROM administration practices lead to the data being documented in the patient's record in multiple formats and locations. The typically non-standard nature of PROM data in electronic health records requiring multiple transformations presents a considerable barrier to broad implementation of PROM-based performance measures.

These data are also contextually dependent, meaning the fitness of a PROM dataset may be deemed sufficient for research while being inadequate for quality measurement.[5-9] Kahn et al (2015) proposed that data quality assessment (DQA) terms be organized into three major categories: 1) *Conformance*, representing the compliance of the data representation to formatting and relational definitions; 2) *Completeness*, describing the frequencies of attributes for data elements without contemplating their actual values; and, 3) *Plausibility*, which determines the level of trust in the element values' faithfulness in representing the intended construct.[5] Decisions regarding data fitness require explicit information about context of use matched with those intrinsic values considered by evaluators to be necessary to achieve a level of trustworthiness acceptable for any specific purpose. Establishing clear expectations for data quality provides effective guidance on how to preserve the necessary information as data is transformed into a format useable by quality metrics. This guidance also serves as the basis for other verification procedures, ensuring complete transparency of evaluators' expectations for both the quality and usability aspects of the information within a dataset.

A major focus of existing DQA models resides at the structural, or data element level. These models primarily address the process of defining data quality through a comprehensive process of flagging inconsistencies in the structure at each data point then dichotomously categorizing as meeting/not meeting expected parameters. Evaluators are frequently left to both interpret the pre-specified parameters of the structural analysis, and to evaluate whether the flagged issues are consistent with their own expectations- a process reliant on them having substantial prior experience in the field with a specific datatype. The necessity to draw upon personal experiences is particularly challenging when assessing PROM data because it frequently exists

in many different non-standard formats. While individual consumers may have varying objectives for understanding data fitness there is a single commonality–to determine whether a dataset is suitable to sufficiently answer the questions about care quality.

Lincoln and Guba describe ideals for attaining trustworthiness: "In order to demonstrate 'Truth Value,' the naturalist must show that he or she has represented those multiple constructions adequately…that the reconstructions that have been arrived at via the inquiry are credible to the constructors of the original multiple realities".[10] To improve the quality and useability of PROM data, a list of high-value DQA concepts was sought in the context of its use for performance measurement; addressing the question of what information decision-makers require when asked to determine the fitness of these datasets. The identification and harmonization of these values will provide a pathway guiding future enhancement of PROM data, thereby increasing its useability for clinical quality assessment.

Methods

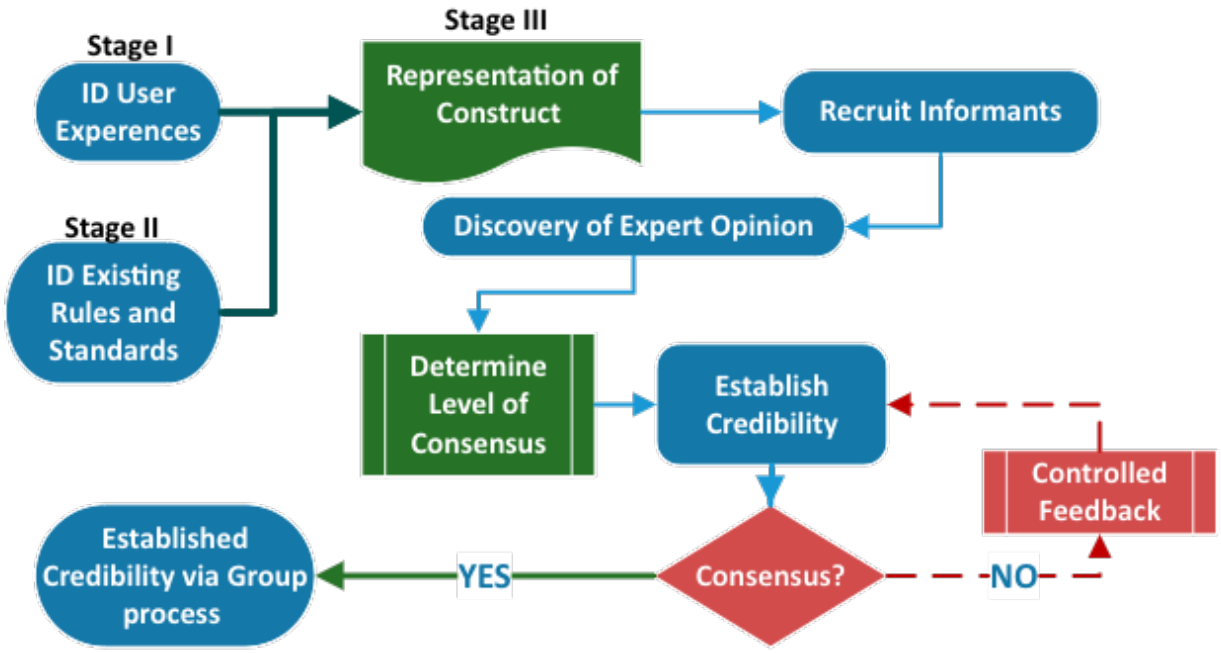
This research included three individual phases, each building upon the previous component to identify and prioritize what evaluators need when making decisions about the quality of PROM data (Figure 3.1). The first stage of this project sought to identify and describe data consumers' experiences with assessing sufficiency of electronic PROM data, specifically on elements they required to make decisions confirming that a dataset was satisfactory for use in performance measurement. The project utilized a series of semi-structured key informant interviews with individuals who regularly use information from DQA for policy decisions, payment decisions, or to inform quality improvement activity.

Stage II sought to identify the existing rules and standards for PROM data and cross reference these with the Phase I findings that demarcated evaluators' expectations for PROM-based DQA. A scoping study was performed to identify and document existing requirements for the use of PROM data in quality measurement. These existing rules reflect the current state of expectations and experiences with these data, the results of which informed the development of a draft trust model for PROM data.

Stages I and II yielded a list of 42 concepts to be considered for a consensus data quality assessment model specific to PROM data for performance measurement (Appendix A). Stage III research further refined the draft list of PROM-DQA concepts using a modified Delphi process to obtain structured feedback from a group of diverse stakeholders, prioritizing guidance on data quality that acknowledges the varied perspectives on a complex issue. [11] [12] Delphi fosters consensus building through the sharing of collective wisdom even in cases where the problem requires prioritization of multidimensional and multidisciplinary issues.[13]

Linstone and Turoff defined the Delphi technique for composing the structure of a new model: "Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem."[11] In this case, the model is a structured set of DQA concepts representing a collective set of informed opinions reflecting substantial experiences of data evaluators assessing PROM data.

*Figure 3.1: Framework of a Modified Delphi Process**



*(*modified from Habibi et al 2014: Delphi Theoretical Framework in Qualitative Research[13])*

The modified Delphi method is particularly effective when managing a heterogenous group with varying perspectives, allowing for conclusions to not be dominated by a particular personality or subset of the group.[11] Modifications of the Delphi technique has also been successful in maintaining sufficient reliability and validity when seeking clinical consensus.[14-23] In this case, the specific modification of the Delphi technique chosen to obtain consensus on a draft DQA model was the "reactive Delphi" where respondents are asked to provide their informed opinions on a list of prepared concepts.[24] This modification of the traditional Delphi technique allowed for the suitability of the structured information derived from the first two phases of this project to be judged by respondents while still allowing for the inclusion of respondent suggested additions to the concepts being considered. The two major criteria judged by respondents were relevance and importance. *Relevance* being judged as a dichotomous

variable and *Importance* on a scale, both based upon the respondents' experiences with each specific datatype.

The goal of the modified Delphi method is a sufficient level of agreement on a draft list of concepts that are important to a broad range of end users of PROM data quality information. Narrowing the DQA concept list to a manageable size while maintaining its relevance to multiple quality activities, requires consensus from a diverse group of end users. The PROM-based DQA model must also be flexible to allow for future refinement as information about its utility is collected throughout deployment. The design principles for the methods chosen to construct the consensus PROM-based DQA model were intentionally calibrated to support a quality use case. The primary goal in this instance was to achieve a moderate to strong level of consensus as a credibility marker across as many related disciplines as possible.

Participants:

A purposive sample of informed stakeholders in the field of data quality, PROMs and clinical decision models were invited to participate for the development of a harmonized PROM-based DQA trust framework.[25] Criteria used to choose participants included: 1) representatives of a related profession (HEDIS-certified auditor, healthcare professional, data scientist, clinical researcher, QI lead/manager, and public health expert); 2) affiliation with a particular setting or work field; 3) representative from an affiliated standards organization (AMIA[iv], HL7[v], NCQA[vi],

[iv] https://amia.org/
[v] https://www.hl7.org/
[vi] https://www.ncqa.org/

OMG[vii]); 4) recognized subject matter expert or authority and 5) level of experience in respective field.[26]

Selected participants represented an array of decision-makers who are the target audience for DQA information and are typically responsible for reviewing presentations of key evidence pertaining to the acceptability or fitness of a dataset and making final decisions about its appropriateness. This role may include decisions that either establish or pre-specify the parameters that are needed to make their final decision. Alternatively, they may rely upon pre-defined and evidence-based parameters published as the practice standard and review them against their own expectations.[27,28]

A three-step process was used to prioritize the attributes of PROM data and concurrently assess the importance of these attributes to the proposed trust framework for data quality.

Step 1: Recruitment of Informed Stakeholders

An initial questionnaire (Appendix C-3) was developed in SurveyMonkey® to identify a set of end-users with the appropriate expertise and experience with making data fitness decisions. The brief survey included information to orient potential participants to the goals of the study, terms and definitions used for a data fitness framework and asked participants to provide information about themselves and their current profession to ascertain their potential to contribute to the study aims.

The questionnaire specifically ascertained stakeholders':

a. number of years' experience with PROM data
b. how frequently they encounter these data in their professional capacity and in what context

[vii] https://www.omg.org/

c. what their primary role is with regard to use of these data– patient facing, patient secondary use (point of care QI or measurement), payer secondary use (HEDIS, VBC, other),
d. history of participation in research or other contributions to body of knowledge, request list of relevant references or other citations of work
e. descriptive information and demographics of the information sets they work in

Equal representation from related disciplines and end-user affiliation was sought, although significant overlap in multiple areas occurred due to the specific expertise sought on the topic. A purposive sample of informed stakeholders was sought using an email sent to those contributors from the first phase of the study who provided substantive input for the initial list of PROM-DQA concepts. The email contained specific information on what information was being sought, the time commitment required and how the feedback they provide would be used.[29] Additional email requests for participation were forwarded to a select group of registrants to the 2021 Digital Quality Summit whose professional affiliation (drawn from registration form field) indicated they might possess relevant knowledge and/or relevant experience in reviewing person-reported outcome data. Finally, stakeholders with specific experience in evaluating PROMs for HEDIS dQMs were recruited through the HEDIS Certified Auditor group via a request published in their monthly newsletter.

Step 2: Discovery of Key Informant Opinion

Respondents to the initial questionnaire were asked to complete a follow up questionnaire (Appendix C-3) focused on ascertaining their collective views on the list of elements and their relevance and importance to making decisions regarding the fitness of patient reported outcome data.[29] The introduction page of the questionnaire included a dictionary of specific terms and acronyms, key references and a brief explanation of the goals for this research.

Participants were asked to rank each concept's relevance and importance based upon their own determinations of its priority, credibility and trustworthiness in the context of data fitness. Each respondent was asked to rate Importance using a seven-point Likert scale which included an "N/A" option to allow respondents to indicate if they felt the concept was not relevant when considering the specific context of data fitness (Figure 3.2). The 7-point scoring range was set up with *Strongly Disagree* equaling one and *Strongly Agree* equal to 7. "N/A" designations were given a score of 0. A seven point Likert was chosen as it provides the best accuracy and ease of use.[30] Respondents were specifically asked to rank each item relative to their own personal experiences and expertise. Open text fields were provided to allow respondents to include their own thoughts about each concept's importance and to elicit specific observations based on their own personal experiences and expertise (Appendix B-3). Several open-ended questions at the end of each page provided respondents an opportunity to also introduce concepts that were inadvertently left out of the original list under review.

Figure 3.2: Sample Question for Stakeholder Feedback

Patient Information

The next set of questions deals with specifics about patients. Please review the description of each concept and provide your perspective on its relevance and importance to data quality decision-making. *If a concept is not relevant to your decision about a dataset's quality, please select* ***N/A***.

Concept Name	Concept Description	Concept Type	Data Quality Domain
Patient Demographics	The number and proportion of records with a value present in Patient Demographics fields (e.g., age, gender, ethnicity, primary language, culturally relevant information)	Data Element	Conformance Completeness Plausibility
Patient Directives	The number and proportion of records containing documentation of patient choices, (consent) permitting or denying recipient(s) or recipient role(s) to perform one or more actions within a given policy context, for specific purposes and periods of time.	Data Element	Conformance Completeness Plausibility

9. Concept Importance: Patient Demographics

The number and proportion of records with a value present in Patient Demographics fields is important when assessing the quality of PGHD. 0

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neither Disagree or Agree	Somewhat Agree	Mostly Agree	Strongly Agree	N/A
○	○	○	○	○	○	○	○

Prior to distributing the questionnaire, the study's data collection instrument, the study definitions, and the supplemental guidance was reviewed by a four-person advisory group comprised of research scientists specializing in the field of quality measurement. The advisory group provided feedback on the survey's presentation of the concepts to be ranked, the level of comprehension required to provide meaningful feedback and the methods used to evaluate each concept's relevance to the topic.

Step 3: Determining Level of Consensus (establishing credibility)

The third and final round for the process convened the group of informed stakeholders in a private virtual Zoom meeting room for discussion and voting on the concepts that are important when making decisions about the fitness of PROM data. Respondents to the data collection survey received a personalized email invitation to participate in the expert panel

discussion on a consensus definition for a harmonized patient-reported outcome data quality framework. A total of 22 invitations were sent with a final group of 10 individuals participating.[13] Reasons respondents provided for their not being able to participate included availability during the scheduled time, perception they did not possess the specific expertise required, or they lacked experience with data quality assessment.

Panelists were convened for a 90-minute session during which they were presented with the option to discuss then vote on those concepts which resulted in moderate to low agreement on importance in the online survey. Concepts were determined for panelist review when their importance score (*weighted* $mean$) was less than or equal to 5.5. Each respondents' vote on a data quality element was equally weighted, except for the "N/As" which was the proxy used for "not relevant" which were assigned a zero weight to eliminate them from the overall *Importance* score. Additionally, any concept receiving at least two "N/A" votes in the survey was automatically included for further discussion by the expert panel, regardless of its actual importance score. Each concept was presented individually to the panel with a graphical representation of the distribution of the results as well as basic descriptive statistics (min, max, median, mean and weighted mean). Panelists were asked if they had any questions about the concept, the data presented or whether they wished to discuss the concept prior to embarking on voting on its importance–conducted using the poll feature in Zoom with panelists making their selection directly from their home screens. The on-screen scoring process used a seven-point Likert scale and panelists were given up to 30 seconds to provide their input. Graphical results of each vote were then immediately presented back to all panelists for reaction and further discussion.

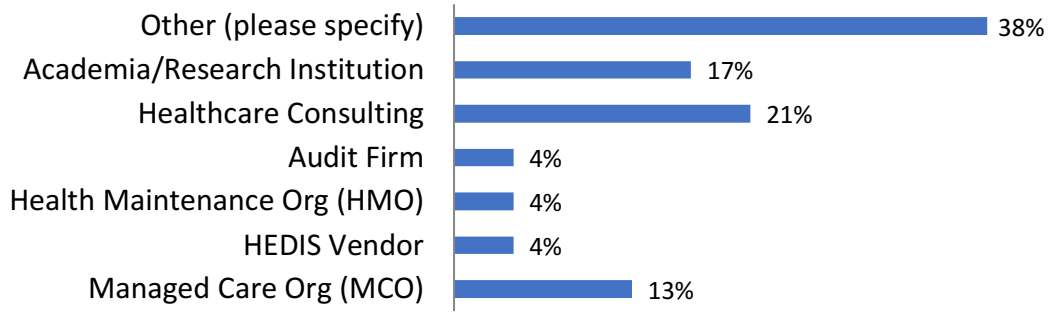
After voting was completed for each concept, panelists were asked if they wished to comment on the results or if any questions or another round of discussion and voting was required to allow panelists to adjust, enhance or modify their prior response. In those cases where panelist responses were distributed between agreement and disagreement, a second vote was performed after discussion. If the results of the second vote were identical to the first, the voting was considered complete, and the next concept was considered until all concepts in the proposed matrix were evaluated.

The session was recorded, and the recording was transcribed in *AWS Transcribe* to allow for documentation and analysis of the key points of the panelist discussions. The transcriptions were reviewed against the recording to ensure the accuracy and clarity of the session notes. These were then used to provide greater detail to the panelist comments on their considerations of the concept's importance.

Results

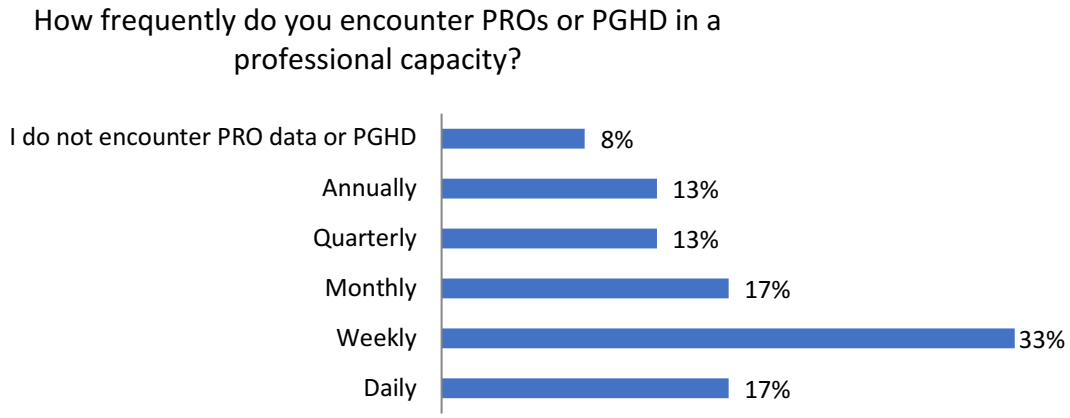
Step 1: The initial survey of informed stakeholders resulted in 24 respondents and a 100% completion rate. Average time to complete the survey was six (6) minutes. Respondents represented several different end-user types (Figure 3.3) and a range of experience with patient-reported outcome data (Figure 3.4).

Figure 3.3: Respondents Professional Affiliation



38% of respondents self-identified as a data scientist or informaticist, 21% as a quality measure developer, 17% identified as a quality manager and 7% as a healthcare consultant. The remaining respondents were evenly split between HEDIS auditor, primary care provider, healthcare researcher, and "other".

Figure 3.4: Respondents self-reported exposure to PROM data



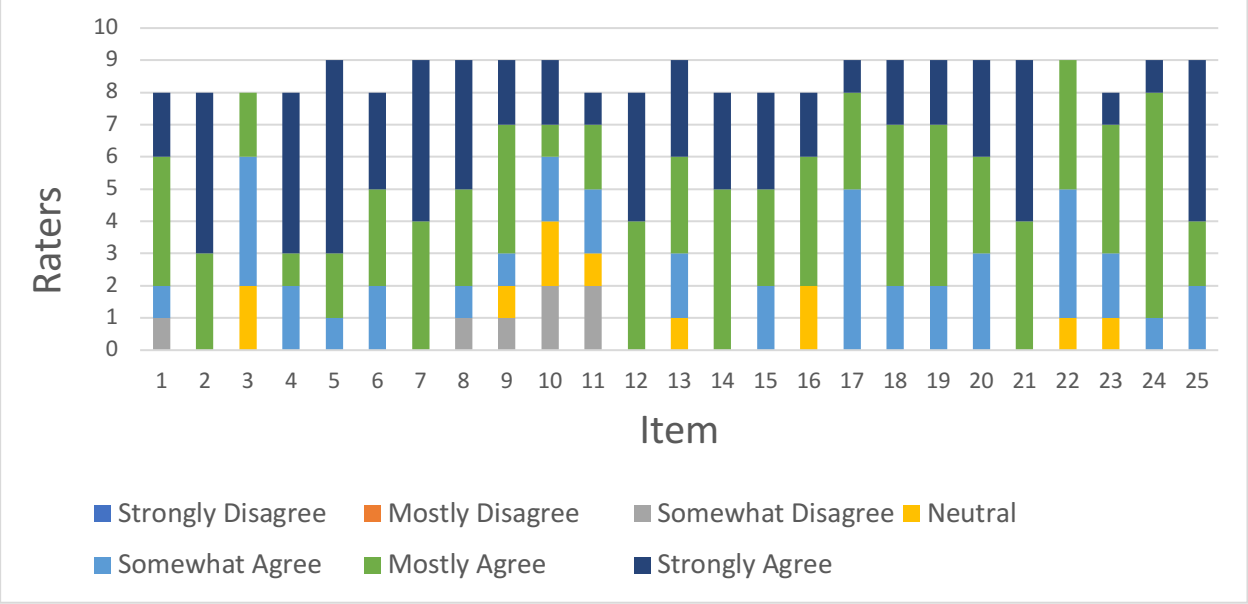
Respondents' level of experience with patient-reported outcome data varied, with 48% having greater than 10 years' experience. 46% reported between 4- and 10-years' experience and only 13% reported having 3 years or less. Eight of the respondents listed their relevant contributions to the general body of knowledge around person reported information and data quality in the form of research published in peer reviewed journals, operational reports for government or other grant supported quality projects, participation in the design and execution of PROM-

based PM development projects, development of data quality tools, and development of quality program requirements that include PROM information.

Step 2: The list of concepts reviewed during the online review and voting of the PROM-DQA concepts consisted of 62 individual questions reviewing a total of 40 individual concepts (Appendix A-3). A total of 23 stakeholders completed all the required questions in the online survey, 74% of whom also provided additional information for optional questions. The optional questions were open ended and primarily involved the respondent providing free text to further elaborate on a concept or any additional details. All respondent data was downloaded from SurveyMonkey® into MS Excel for analysis.

Step 3: Expert Panel review of the draft concept list. A total 25 concepts were posted to the screen with a brief introduction by the session moderator, a brief discussion of the results for the concept's evaluation in Step 2, followed by the participants using the Zoom poll feature to vote on importance (Figure 3.5).

Figure 3.5: Expert Panel Item Agreement



Representation of the groups' expertise was intentionally broad, and the discussions provided some insight into the context for the varying level of agreement on importance. From a certified HEDIS auditor perspective, the responses tended towards a priority on the analytics of the datasets as a whole:

> *"I tend to apply a deductive approach that if there's good data quality analytics that are there showing low number of anomalies, referential integrity looks quite good."*

> *"We do a lot of data characterization and if you have a large number of data points, particularly for exactly the same information over time, you can definitely make some assessments with regard to the consistency of that data."*

Whereas the practicing clinicians were more interested in the individual scenario in which the data elements were being collected:

> *"If I'm doing it through a web app in the privacy of my own space versus directly to my physician, I might answer differently. I feel more compelled to give the truth when I was speaking with my clinician because I know that there's an expectation of whatever I give him is important parcel to my therapy and treatment."*

Analysis

Data returned from respondents in Step 2 was analyzed for completeness utilizing the completion statistics report in SurveyMonkey and agreement using a computed weighted mean ($\bar{x}$) score (Figure 3.6). Weighted mean instead of arithmetic mean was chosen to account for the frequency of respondent choice contributing to the overall importance rank.[31,32] The weighted mean calculation does not include any of the "N/A" designations. Basic descriptive statistics were also computed for each element (minimum, maximum, median, mean, SD), all of which were presented to the panel during Step 3 of the consensus process.[13,29,33]

Figure 3.6: Weighted mean ($\bar{x}$)

$$weighted\ mean = \frac{x_1 w_1 + x_2 w_2 + x_3 w_3 \ldots x_n w_n}{Total}$$

w = weight of answer choice (1=strongly disagree, 7=strongly agree)
x = response count for the question

The benchmark for high agreement amongst respondents was selected as a weighted $\bar{x} \geq 5.5$, moderate agreement when $\bar{x}$ < 5.5 and >3, and low agreement when $\bar{x} \leq 2$. The benchmark was selected as the score represented approximately 80% of the points on the scale.[34,35] Any element that received at least two N/A votes was included for further discussion by the panel. Of the 40 concepts presented for evaluation, 19 received a not relevant (N/A) designation from any respondent, 10 of those 19 receiving at least two N/A votes from respondents. Only 1 concept (Data Quality Test Type: N/A = 3) received two or more N/A votes from respondents. Of those ten concepts receiving two or more N/A votes, seven had a $(\bar{x})$ > 5.5 which pushed them into a final round of review by an expert panel of key informants.

Respondents strongly agreed on the importance of 22 of the 40 concepts ($\bar{x}$>5.5), and only moderately agreed on the importance of the remaining 18 concepts (3<$\bar{x}$<5.5). None of the original 40 concepts were calculated as having poor agreement ($\bar{x} \leq 2$) amongst respondents. 20 concepts received no "Disagree" votes on importance, however in 15 of these 20 cases, at least one respondent submitted a neutral vote (Neither Agree nor Disagree). The five concepts that received no negative or neutral votes included:

1. *Database Format Interoperability*: The indication that the data format used meets ONC interoperability requirements for data exchange
2. *Data Quality Assessment*: A record of the number and type of measures deployed that are monitoring the quality of the target data at any point in time
3. *Questionnaire Response Value Low*: the number of records that have a value that is lower than a pre-specified threshold
4. *Questionnaire Response Value High*: The number of records that have a value that is higher than a pre-specified threshold

5. *Data Quality Thresholds*: The reference to a technical specification or other relational dataset that defines the use-case threshold for accuracy and precision

Discussion

The development of a consensus-based definition for what constitutes data quality, and the integrated hierarchy of that information relies upon agreement by the end-user that the critical and foundational aspects of a dataset needed to determine its fitness are present and within expected margins. This study set out to answer several essential questions: First, to identify what specific information must be present and made clear to evaluators assessing PROM data quality; and second, to understand what external references (published evidence, accepted norms, repositories, etc.) are relied upon when making judgements about the usability of PROM data for the purposes of performance assessment. Discovery of informed stakeholder opinion and expectations was the method chosen for addressing these questions. The main purpose of this study was nicely summed up by one of the experts during the final round of review.

> *"…it's rare to have the guidance. I mean, that's basically what were employed to do…to apply our own expertise and understanding because it's rare that we have clearly defined instruction and what exactly the data quality should look like under these scenarios."*

Establishing transparency around a core set of expectations for PROM-DQA creates an opportunity for both the recorders of the information and the vendors of the systems within which it is recorded to enhance the documentation of relevant outcomes. Creating uniform set of data expectations that can be continuously assessed against a standard benchmark improves the credibility, reliability, and utility of the information. Ensuring the representations of data quality achieve a sufficient level of face validity for all decision-makers, regardless of their

fluency in data science methods, will allow broader acceptance of PROM information in the assessment of clinical quality and will substantially inform quality improvement efforts. As there is a broad range of end user needs, even in the narrow context of quality improvement, it can be challenging to create a completely uniform set of requirements for PROM data quality. The data quality concepts reviewed in Steps 2 and 3 were grouped into related categories using the information provided by informed stakeholders and the original classifications or associations from the FHIR Resource definitions or OMOP DQD reference tables. Related concepts that were originally split for their independent evaluation of relevance and importance to the DQA model were recombined based on the informed stakeholder information. For example: the concepts related to questionnaire information were assessed by stakeholders as 11 individual concepts (Questionnaire, Questionnaire Data Type, Questionnaire Response Value, Questionnaire Standard Terminology, etc.) however for inclusion in the framework, those elements that were determined to be both relevant and important to data fitness decisions were aggregated into the Questionnaire category comprised of five components (Table 3.1). The consolidation of individual concepts into domains was guided by the stakeholder discussions during the expert review.

Weiskopf's *Desiderata* defined the critical parameters for an ideal DQA model. The model developed in this study was collated with broad stakeholder input from across those disciplines that are directly involved with data quality decision making as well as those involved in using the results to inform policy. This model provides guidance about best practices for improving the consistency of decision making within a specific context, however its adherence to open standards allows for continued iteration enabling the parameters to be adjusted as users gain

more experience with it. The model was built in harmony with well-established and open data quality and digital measurement standards–stewarded by international Standards Development Organizations (SDOs)–and are well supported by both the SDOs and the communities that use them. Finally, the model does provide detail about minimum expectations but does not reference any gold standard or other benchmarks.

Table 3.1: Concepts Important to Patient-Reported Outcome Data Fitness Decision-making

Database Format	Use of standard interoperability format for data structure (e.g., FHIR, OMOP) Format used meets ONC interoperability requirements for healthcare data exchange (e.g., USCDI)
Data Governance	Presence of reference information used for the assessment and transformation of data Presence of process documentation characterizing the necessity of data enhancement during transformation process
Data Source	Records must contain the primary source of patient-reported outcome data The presence of provenance information regarding all transformations made to the dataset prior to the quality assessment Presence of information verifying the end-user interface represents the data in a manner that is true to its original form (original veracity of data persists in target dataset)
Context of Use	Intent (use-case) for the target dataset must be well defined
Patient Encounter	Documentation of actual encounter datetime information must be present
Questionnaire	Documentation for instrument type, or method used to collect patient-reported outcome data must be present Presence of a total score (result) asserting completion of a standard PROM instrument Prevalence report detailing any missing values in the score or result field of the dataset Documentation of, and reference to, the specification for any pre-defined threshold for indicated use-case (e.g., quality measure defined threshold indicating positive observation) The number of records for whom the score or result field is higher or lower than the pre-specified thresholds for indicated use-case. This also includes the number of results in any record that are found to be out of range (not outliers) for the PROM used.
Workflow	The presence of information detailing those workflow practices that must exist for end users to “trust” the information is accurate and can be used for clinical quality.
Data Quality Assessment	A record of the number and type of data quality assessment (DQA) measures used to ascertain its fitness for an intended use-case Presence of a DQA measure report exposing inconsistencies and critical errors within the dataset that would affect fitness determinations for an indicated use-case Presence of report detailing results of tests used to determine data dependences

Study Limitations and Strengths

This study sought to define a consensus-based model for assessing patient-reported outcome data for use in quality measurement programs; however, the lack of any current DQA model for PROM information and a global lack of clarity on the terms used in the field of patient reported outcome performance measurement (PRO-PM) presented several challenges during this study. A complete lack of guidance in terms of how to validate PRO or PROM data for the purposes of assessing the quality of care made it very difficult to recruit key stakeholders. Due to the extensive, and often inconsistent, use of the many different terms to describe patient reported outcome data–such as patient-generated health data–we included a variety of terms in the literature searches, surveys, and interviews. This is a limitation of the study as these terms are not interchangeable with the specific concept of "patient-reported outcome data" and may have been interpreted differently by different stakeholders. The need for the study to obtain consensus from a group of stakeholders with a broad set of perspectives on the use of patient-reported information (a category including patient reported outcome data) required the expansion of the definition to match the stakeholders understanding. While the intent for use of this model to be focused on PROMs, the inconsistency in the use context for the various terms presents some challenges for the final model. This is a limitation of the interpretation of the results for patient-reported outcome data specifically and could limit its applicability in the framework. Future research and more precise application in digital quality measurement should utilize these terms correctly and accurately which would improve the quality of this framework. Qualitative studies using a similar method suggest a median consensus threshold of greater than 88% however, for this particular study, a moderate mean threshold of 75% was chosen for

this study to ensure the final framework represented the widest array of perspectives possible.[12,36-38] The main goal of this study was the formulation of a core list of concepts, rather than final agreement on the components of the list, the lower threshold was necessary to ensure inclusion of the widest array of viewpoints possible without creating a series of "sub-lists" for any one stakeholder group. While it can be difficult to address reliability of the conclusions made by these qualitative findings, it would be possible to independently verify the credibility of informed stakeholders' expertise. The small size of the respondent sample and the lack of repeated rounds of questioning and voting on every single concept might lessen the repeatability however the close alignment with international standards (e.g., OMOP, FHIR) of the concepts under review lessens the chance of variation in respondents' interpretation and improves the probability that the importance rating reflects a common understanding.

It is possible when using a reactive Delphi process to impose the researcher's preconceptions when presenting an already-specified list of concepts for review without allowing for respondent perspective related to the problem.[11] In this case, this was managed by using open-ended questions during the Stage I survey of user experiences and thoughts on patient-reported outcome data quality topics and by providing multiple free text fields within every category being assessed during Stage 3 to allow for additions and alternative options to be suggested by participants.

This study confirms the rarity of decision-makers access to guidance when assessing fitness of PROM data. They must rely upon a variety of statistical and operational test results to make these decisions and are often presented with seemingly inconsistent information which requires substantial expertise to interpret. The model's focus is specific to patient-reported

outcome measures (PROMs) and while this is a very narrow context, the general parameters may apply to other domains of healthcare where patient reported information is self-documented in their medical records. In the context of social determinants of health (SDoH), where a person's own assignment of their race and ethnicity might be very different than what a clinician or other proxy might observe and document in the medical record, the same PROM principles for data quality assessment may apply. The PROM-based DQA domains such as *Source* and *Workflow* are potentially very valuable to those studying social determinants, especially their intersection with patient reported outcomes.

Conclusion

A uniform, well-defined list of data quality concepts applicable to using PROM data for the purpose of quality measurement would provide much needed guidance and enhance the consistency and reliability of data fitness decision-making. This study confirmed the desire for more specific guidance to support evaluators' decision making about PROM data fitness. Despite varying objectives, decision makers can agree on a model providing uniform parameters for evaluating data fitness that are central to their decisions about the useability of PROM data. Clarifying expectations for data consistency, completeness, and plausibility through deployment of unambiguous requirements defined in quality program rules and regulation is needed to increase acceptance of PROM data in the quality measurement ecosystem.

This additional detail still needs to be defined by those programs for which PROM data is being validated and these benchmarks should be established through systematic analysis of real-

world PROM data submitted for quality reporting informed by the experiences of evaluators in the field making data quality decisions for these programs.

References

1. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med*. Dec 1 2009;151(11):793-803. doi:10.7326/0003-4819-151-11-200912010-00007

2. Williams SB, O'Connor EA, Eder M, Whitlock EP. Screening for child and adolescent depression in primary care settings: a systematic evidence review for the US Preventive Services Task Force. *Pediatrics*. Apr 2009;123(4):e716-35. doi:10.1542/peds.2008-2415

3. Lee K, Weiskopf N, Pathak J. A Framework for Data Quality Assessment in Clinical Research Datasets. *AMIA Annual Symposium proceedings AMIA Symposium*. 2017;2017:1080-1089.

4. Feder SL. Data Quality in Electronic Health Records Research: Quality Domains and Assessment Methods. *Western journal of nursing research*. May 2018;40(5):753-766. doi:10.1177/0193945916689084

5. (US) IoM. 4, Health Care Data Standards. In: Aspden P CJ, Wolcott J, et al.,, ed. *Patient Safety: Achieving a New Standard for Care*. National Academies Press; 2004.

6. FDA. *Patient-Focused Drug Development: Collecting Comprehensive and Representative Input Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders*. Services USDoHaH; 2018. https://www.fda.gov/media/139088/download

7. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A Data Quality Assessment Guideline for Electronic Health Record Data Reuse. *EGEMS (Washington, DC)*. Sep 4 2017;5(1):14. doi:10.5334/egems.218

8. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*. 2005;8(1):19-32. doi:10.1080/1364557032000119616

9. FDA. *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. 2009. *FDA Guidance Documents*. Accessed September 20, 2019. https://www.fda.gov/media/77832/download

10. HealthIT.gov. United States Core Data for Interoperability (USCDI). The Office of the National Coordinator for Health Information Technology (ONC). Accessed April 2, 2021. https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi

11. NCQA. *HEDIS MEASUREMENT YEAR 2020 VOLUME 5: HEDIS COMPLIANCE AUDIT™: STANDARDS, POLICIES AND PROCEDURES*. National Committee for Quality Assurance; 2020. Accessed March 22, 2021. https://store.ncqa.org/hedis-my-2020-volume-5-epub.html

12. OHDSI. The Data Quality Dashboard Observational Health Data Sciences and Informatics. Accessed Oct. 20, 2020. https://data.ohdsi.org/DataQualityDashboard/

13. Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Washington, DC)*. 2016;4(1):1244. doi:10.13063/2327-9214.1244

14. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. A Data Quality Ontology for the Secondary Use of EHR Data. *AMIA Annual Symposium proceedings AMIA Symposium*. 2015;2015:1937-46.

15. Jain ST, B. Data Lineage: An Important First Step for Data Governance. *BeyeNetwork* blog. December 18, 2013. http://www.b-eye-network.com/view/17023

16. Tiase VL, Hull W, McFarland MM, et al. Patient-generated health data and electronic health record integration: protocol for a scoping review. *BMJ Open*. Dec 17 2019;9(12):e033073. doi:10.1136/bmjopen-2019-033073

17. College C. Research Skills Tutorial: 1.2 Searching with Keywords. Columnbia College. 2021. https://columbiacollege-ca.libguides.com/research_skills/getting_started/keywords

18. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights*. 2013;6:61-8. doi:10.4137/HSI.S11093

19. Plastiras P, O'Sullivan D. Exchanging personal health data with electronic health records: A standardized information model for patient generated health data and observations of daily living. *International journal of medical informatics*. Dec 2018;120:116-125. doi:10.1016/j.ijmedinf.2018.10.006

20. Ritchie JaS, L. Qualitative Data Analysis for Applied Policy Research. In: Bryman AaB, R.,, ed. *Analyzing Qualitative Data*. 1994:173-194.

21. Oliver S. Making research more useful: Integrating different perspectives and different methods. 2001:

22. Talbot W. Confirmation on the General Guidelines and HEDIS Audit Standards for Verification and Use of Person-Reported information in HEDIS Reporting. In: Hamlin B, editor. verification of the HEDIS General Guidelines and HEDIS Compliance Audit Standards that pertain to the use and validation of person-reported infomation in HEDIS ed2021.

23. OHDSI. OMOP Common Data Model. The Observational Health Data Sciences and Informatics (OHDSI). 2019. https://www.ohdsi.org/data-standardization/the-common-data-model/

24. CMS. cypress: Rigourous and repeatable testing of Electronic Health Records. Accessed September 12, 2019. https://www.healthit.gov/cypress/

25. Estiri H. DQe-c. GitHub. 2019. https://github.com/hestiri/DQe-c

26. OHDSI. ACHILLES for data characterization. Observational Health Data Sciences and Informatics (OHDSI). Accessed January 5, 2019. https://www.ohdsi.org/analytic-tools/achilles-for-data-characterization/

27. ONC. A User's Guide to Understanding Trusted Exchange Framework and Common Agreement (TEFCA) Draft 2. The Office of the National Cooridnator for Health Information Technology; 2019. Accessed June 2019. WWW.HEALTHIT.GOV/TEFCA

28. HL7. Validating Resources. HL7 International. Accessed July 2019, 2019. https://www.hl7.org/fhir/validation.html

29. HL7_International. QI-Core Implementation Guide: STU 4 (v4.0.0 for FHIR 4.0.1). 2020. http://hl7.org/fhir/us/qicore/index.html

30. OHSU. Patient entered data. American Medical Informatics Association. 2021. https://www.clinfowiki.org/wiki/index.php/Patient_entered_data

31. ONC. Patient-Generated Health Data. Healthit.gov. Accessed March 21, 2021. https://www.healthit.gov/topic/scientific-initiatives/patient-generated-health-data

32. ONC. Patient Engagement Playbook. Office of the National Coordinator for Health IT (ONC). Updated 2019. Accessed March 21, 2021. https://www.healthit.gov/playbook/pe/

33. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform*. Oct 2013;46(5):830-6. doi:10.1016/j.jbi.2013.06.010

34. NCQA. Audit Process: Offsite Methods. *HEDIS 2019 Volume 5: HEDIS Compliance Audit™: Standards, Policies and Procedures*. 2019 ed. National Committee for Quality Assurance; 2019. *The Healthcare Effectiveness Data and Information Set*.

35. HL7.org. U.S. Data Access Framework (DAF) FHIR Implementation Guide (IG). HL7 International. Accessed October 16, 2018. http://hl7.org/fhir/2015May/daf.html

36. IMF. *IMF Data Quality Assessment Framework*. Department PbtSDatPDaR; 2001. *Fourth Reveiw of the Fund's Data Standards' Initiatives*. https://www.imf.org/external/np/sta/dsbb/2001/070201.pdf

37. Raszczynski R. The EAV Model. INVIQA; 2010. 2010. Accessed July 2019. https://inviqa.com/blog/eav-data-model

38. Rogers JR, Callahan TJ, Kang T, et al. A Data Element-Function Conceptual Model for Data Quality Checks. *EGEMS (Washington, DC)*. Apr 23 2019;7(1):17. doi:10.5334/egems.289

Appendix A-3: Concepts Considered for PROM Trust Framework

Table A3-1: Concepts Considered for PROM Trust Framework

Concept Name	Concept Description	Concept Type	Data Quality Domain
Database Format	Format of source dataset used to assess PGHD fitness (e.g., FHIR, JSON, SQL, CQL, etc.)		Conformance
Database Standard Format	A yes or no value indicating the dataset used to evaluate PGHD fitness is stored in a standard format that is useable, locatable, feasible to access and process	Test & Parameter	Conformance
Database Standards Use	Version of standards used to store information (e.g., FHIR 4.0.1, CQLv1.5.1, etc.)	Summary Information	Conformance
Database Format Interoperability	A yes or no value indicating the data format used meets ONC interoperability requirements for data exchange internally and externally (e.g., USCDI)	Test & Parameter	Conformance
Patient Demographics	The number and proportion of records with a value present in Patient Demographics fields (e.g., age, gender, ethnicity, primary language, culturally relevant information)	Data Element	Conformance Completeness Plausibility
Patient Directives	The number and proportion of records containing documentation of patient choices, (consent) permitting or denying recipient(s) or recipient role(s) to perform one or more actions within a given policy context, for specific purposes and periods of time.	Data Element	Conformance Completeness Plausibility
Data Source Type	The number and proportion of records containing source of PGHD information (person, medical device, wearable device, other)	Data Element	Conformance Completeness Plausibility
Data Source Provenance	The number and proportion of records containing metadata on the activities taken upon the source information to transform to the target dataset. Provenance is prepared by the application that initiates the create/update of the PGHD	Data Element	Conformance Completeness Plausibility
Audit Log: Data Record Management	A yes or no value indicating the presence of the records management log assuring privacy and security safeguards are maintained and functioning.	Test & Parameter	Conformance
Audit Log: Data Quality Assessments	Number and type of measures deployed that are monitoring the quality of the target data at any point in time (ratio of data to errors, ETL transformation errors rate, data time to value, mandatory field enforcement, number of in-line corrections)	Summary Information	Conformance
Audit Log: DQA Report	Presentation of DQ measure results exposing discernable patterns of data inconsistencies and/or systematic errors	Summary Information	Plausibility
Audit Log: Provenance	A yes or no value indicating the presence of provenance information	Test & Parameter	Completeness
Audit Log: Persistence	A yes or no value indicating the persistence of provenance information that verifies the data is true to original form in the end user interface.	Test & Parameter	Conformance

Concept Name	Concept Description	Concept Type	Data Quality Domain
Encounter Type	The number and proportion of records documenting the type of encounter during which health data was generated	Data Element	Completeness
Encounter Location	The number and proportion of records documenting the location where the PROM collection takes place and/or the mode of data collection (face to face, asynchronous via online form, other)	Data Element	Completeness
Encounter class	The number and proportion of records documenting Period (Date/Time) information for Encounter Type	Data Element	Completeness
Encounter reference	The number and proportion of records relating the frequency of patient contributions to their health record for any PROM element. (e.g., how many times PHQ-9 appears within a certain period for an individual.)	Data Element	Completeness Plausibility
Encounter performer	The number and proportion of records containing the role\licensure of person documenting or collecting the PROM in record (patient, Care Manager, MD, RN, etc.)	Data Element	Completeness
Questionnaire	The number and proportion of records documenting standard instrument type used to collect PROs	Data Element	Completeness Conformance
Questionnaire Data Type	A yes or no value indicating if the Questionnaire is the expected data type based on the technical specification.[39]	Data Element	Conformance
Questionnaire Response Value	The number and proportion of records documenting presence of total score result indicating complete administration of questionnaire-asserting the structured group of questions that comprise the instrument have been successfully answered and a valid total score was calculated and present in the record.	Data Element	Completeness Plausibility
Questionnaire Response metadata	The number of records that contain supplemental information on acceptable use parameters for the instrument used (legal rights of use info, valid scoring range, positive thresholds, etc.).	Data Element	Completeness
Questionnaire Standard Terminology	The number and percent of records that have a standard, valid term present in all Questionnaire and Questionnaire Response fields as specified. (Use of preferred standard terminology to document instrument and result.)	Data Element	Conformance Plausibility
Questionnaire Response Null Value	The number and percent of records with a NULL value in the QuestionnaireReponse value field (considered not nullable)	Data Element	Completeness Conformance Plausibility
Questionnaire Response Zero Value	The number and percent of records with a value of 0 in the QuestionnaireReponse value field	Data Element	Completeness Conformance Plausibility
Questionnaire Response: Missing Value	The number and percent of records with missing QuestionnaireReponse values	Data Element	Completeness Plausibility

Concept Name	Concept Description	Concept Type	Data Quality Domain
Questionnaire Response: Value Threshold	The prespecified threshold (expected) for the QuestionnaireReponse value field	Test & Parameter	Conformance
Questionnaire Response: Value Low	The number of records that have a value that is lower than a pre-specified threshold.[39]	Data Element	Conformance Plausibility
Questionnaire Response: Value High	The number of records that have a value that is higher than a pre-specified threshold.[39]	Data Element	Conformance Plausibility
Data Quality: Thresholds	The reference to a technical specification or relational dataset that defines pre-specified thresholds asserting sufficient accuracy and precision for any specific use case	Test & Parameter	Conformance
Data Quality: Context Specificity	Use case pre-defined for correction and enhancement of datasets linked to its intended use.	Test & Parameter	Conformance
Data Quality: Adherence to Target	A yes or no value indicating target dataset complies with the pre-specified end-user expectations for intended use case	Summary Information	Conformance
Data Quality Test Type	Test type used to ascertain dataset adherence	Test & Parameter	Conformance
Dataset Governance Reference	Reference information used in the assessment and standardization of information within the dataset	Test & Parameter	Conformance
Data Quality: Dataset Governance Process	Documentation of the process that characterizes the necessity of data enhancement based on standard reference information.	Summary Information	Conformance
Data Quality: Measurement Report	A report that characterizes how well source data transformation to target dataset meets end user expectations.	Summary Information	Conformance Plausibility
Data Quality Criterion	What criterion/test is used to ascertain data dependencies?	Test & Parameter	Conformance Plausibility
Data Quality Dependency Report	Report of data dependencies results	Summary Information	Conformance Plausibility
Actionability	what do errors and inconsistencies signify		
Workflow	Workflow practices that must exist for end users to “trust” the information is accurate and can be used for clinical quality		

Appendix B-3: Stakeholder References and Additional Feedback

Table B3-1: External References & Processes Used to Test Patient-Reported Outcome Data Quality

Database Format	• *Pre-built test cases in Information Analyzer tool[1] with thresholds for acceptance* • *conduct review of primary source documents to applicable database fields* • *conformance to standard is first test, then checks are made for the code validity* • *Test that keys work to link related files. Sometimes do univariable tests (ranges, etc.) or intervariability consistency checks (skip patterns, etc.)* • *Conformance test results, if available. Otherwise, import/export validation logs.*
Data Source/Patient Information	• *Descriptive-based statistics and logical relationship review based on established thresholds* • *missingness and bias in missingness; contradicting information; evaluate correlations among volume of patient records with volume of patient demographics/directives* • *Primary Source Verification[2]* • *association rule confirmation, trending, monitored form completion (time responses), ancillary data reconciliation (confirmation against existing records/redundant collection)* • *commonly [used] test is null value counts. Required fields should have 0, optional fields can have not 0. If reference tables are not used for fields (e.g., gender), then value frequencies should be generated to identify anomalies* • *demographics: [for example] if the person's visits all occur after their birth year and ensuring that they have a valid sex* • *typically examine a set of various fields for the determination of data source in 3 dimensions: The source organization, the source person(s) and the source technology. No thresholds are used but outreach is often made to sending systems when this information is not available.* • *Depending on volume and impact of data, verifying sample of records from target Db to source Db. Threshold is dependent on error type, but less than 1% is standard* • *Reviewing ETL source code and logs from adjacent Db to target Db is also useful* • *Provenance data is excellent for error resolutions, but usually code review and review of reference tables is needed for systemic data issues*

[1] https://www.ibm.com/products/infosphere-information-analyzer

[2] https://www.ncqa.org/programs/data-and-information-technology/hit-and-data-certification/hedis-compliance-audit-certification/about-the-hedis-compliance-audit/

Patient Encounter	• *We look for standardized codes for encounters to determine the type of encounter, such as SNOMED, CPT and HL7 encounter codes. We do not reject data lacking this information, but it may make other processing and use of the data less meaningful or informed* • *Encounter type, location, and date/time are necessary to assess the reasonableness of data. Date/time is very important, especially in determination of the timeliness metric, expectations about relevance and completeness of data can be set based on type and location.*
Instrument and Instrument Score	• *Provenance: Not Null, and valid value. Persistence is secondary to expected data management adjuncts.* • *Look for missing data. Look for outliers. Identify who reported the data, why and at what point. Look for patterns within the responses*
Data Quality Assessment	• *Data quality reports can be misleading and tricky. If thresholds for data acceptance are too high, excessive data could be dropped in ETL resulting in a biased data set and unreliable research. If no errors are reported, parameters may be overly permissive.* • *Tailoring of validations and reporting of validation outcomes should be a very contextual exercise.* • *Audit logs are commonly used in QM[1] audits unless anomalies arise in data interrogation and root cause analysis is needed*

Table B3-2: Additional Considerations for Patient-Reported Outcome Data Quality Assessment

Database Format	• *A standard data format does not necessarily mean the source of truth data is accurately reflected in the data source* • *The ONC USCDI standards[2] are pretty strong for EHR data but have not been as well established for the many types of patient-reported outcome data* • *Integrity of the structured data is only as valuable as the integrity of the data to the producer's intent [which] can be diminished by excessive efforts toward standard formatting.* • *Reviewing data model, including LDM[3], Referential Integrity[4], and field constraints[5]*
Data Source/Patient Information	• *Identifying why a value is missing, and to what extent the missingness is systematic.* • *I don't differentiate between patient reported and provider documented data, as long as the data is recorded in the legal health record*

[1] quality measure (QM)
[2] https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi
[3] Local data manager (LDM) system
[4] https://database.guide/what-is-referential-integrity/
[5] https://www.tutorialspoint.com/sql/sql-constraints.htm#:~:text=Constraints%20are%20the%20rules%20enforced,level%20or%20a%20table%20level.

	• *Patient directives are often important in the context of data access but are often not present in the data delivery* • *Additional study is needed to assess best practices for overcoming patient hesitancy to share personal information* • *check that any sex-related records (like prostate cancer) are properly attributed to the correct sex* • *The technology (whether EHR or other IT system) is important in the provenance of information (both patient-reported outcome data and otherwise). Significant and meaningful data quality issue may be present based upon technology utilized.* • *Intrinsic expression vs. extrinsic can also be important. Was the data collected actively by intent or passively by trigger or random collection? Was the subject aware of the data collection?*
Patient Encounter	• *Location of the encounter (PCP, urgent care, express care) is important.* • *Dates and updates are the most important thing here, [observations may] change over time (previously a smoker, no longer a smoker)* • *The location when the information was biologically relevant is rarely available. Most systems do not rigorously record and transmit such information, even though it may be valuable for context.* • *Recorder information can often be misleading. The recorder may be a scribe during an encounter with a provider. It must be considered carefully and should be collected with consideration to the intended use.*
Workflow	• *Optimal clinical workflow and quality PROM measures will also likely require contributions from AI/ML to structure unstructured data that is the preferred forms of both patient and provider expression*
Data Quality Assessment	• *data quality checks are not shareable in a fully computable context. This makes it difficult to accept the audits/checks performed by others when dealing with healthcare data* • *Very few healthcare organizations have mature data management programs in which data quality metrics are used.* • *prefer to have some metadata about the full data source to get a better understanding of what to expect from a database. This is usually the name, type of data (EHR, etc.) date of the source data version and OMOP vocabulary version*

Appendix C-3: Patient-Reported Outcome Data Quality Surveys

Round 1: Evaluating the quality of patient reported outcome (PRO) data - Recruitment of Informed Stakeholders

Section 1: Email outreach to recruit informed stakeholders

[subject] Requesting your assistance with data quality research
[body]

[name],
I am conducting research into perceptions of patient reported data quality. Specifically, I am seeking to identify a set of intrinsic values considered by decision-makers to be the "gold standard" for high quality datasets. I am hopeful that the identification, harmonization and standardization of expectations would provide a roadmap for the enhancement of patient reported data, thereby increasing its value to quality programs.

I am seeking input from experts across several disciplines, specifically those with substantial experience evaluating data quality and/or auditing healthcare data management practices. If you are interested in supporting this research project, please complete the brief screening questionnaire HERE. Following the initial questionnaire, a second link to the actual data collection survey will be provided. The initial survey is intended solely for the purposes of helping categorize your responses and is not intended to evaluate your expertise in any way.

If you are not able to participate or feel that I should also seek feedback from someone else in your organization, please feel free to forward this email.

Thank you in advance for your time and willingness to support my research.

Sincerely,
Ben

Section 2: Online data collection using SurveyMonkey

Thank you for being willing to contribute your own personal expertise to this project. This study seeks to identify and characterize data consumer perspectives regarding the quality of person-generated health outcome data. The goal of this research is to improve the value of this important outcome data and promote its effective use in quality improvement.

The information obtained from this research will assist in developing a consensus-based definition for data quality, harmonizing different end-user perspectives around a common data quality assessment framework (DQAF).

Participants will be asked to evaluate a list of quality concepts and provide their own perspective of priority, credibility and trustworthiness in the context of data fitness. The results will be used to ascertain initial agreement for a of a structured consensus definition for PRO data quality.

Thank you.

Section 3: Survey to collect Informed Stakeholder Information

1) Name [First, Last]
2) Name of firm [organization with which you are most recently affiliated]
    1. If not currently affiliated w/org utilizing or evaluating PRO or PGHD, please ID the org for which you were affiliated at the time

3) Org Type
    1. Managed Care organization
    2. HEDIS vendor
    3. Health Maintenance organization
    4. Audit Firm
    5. Primary Care Practice (incl PCMH)
    6. ACO
    7. Healthcare Consulting
    8. Hospital
    9. Academia

4) How frequently do you encounter PROs or PGHD in a professional capacity?
    1. Daily
    2. Weekly
    3. Monthly
    4. Quarterly
    5. Annual Review
    6. I do not encounter PRO or PGHD

5) What is your primary role in terms of reviewing and/or evaluating PROs or PGHD?
    1. Primary care provider
    2. Quality Manager
    3. Auditor
    4. Data Scientist/Informaticist
    5. Healthcare Researcher
    6. Health Policy decision maker
    7. Quality measure developer
    8. Consultant
    9. Other (please explain)

6) Number of years' experience in working with PRO data and/or PGHD
    1. 1-3,
    2. 4-6,
    3. 7-10,
    4. 10+

7) Is the PRO data or PGHD you encounter limited to a particular demographic or geographic boundary? If yes, please describe briefly:

8) Please provide a list of your relevant research or other contributions to the body of knowledge pertaining to data quality assessment of PRO data or PGHD.

Round 2: Review and voting on Data Quality Concepts by Informed Stakeholders

Section 1: Email outreach providing informed stakeholders access to data collection instrument

[subject] Research into person-reported data quality
[body]
[name],
Thank you for agreeing to contribute to my research on person-reported data quality. Please provide all your responses using the online survey (link to survey or use QR code below). The survey should not take more than 45 minutes to complete.

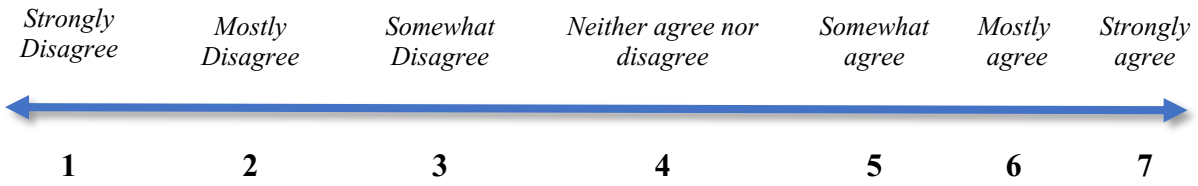
If you have any difficulties accessing the data collection survey or find an issue with one of the questions, please email me ASAP at hamlin@ncqa.org
If you indicated in the original invite that you would be willing to be part of the expert panel review, please input your availability for a 90-minute Zoom panel meeting using this link.
Thanks again for supporting this important research into perceptions of patient generated data.

Sincerely,
Ben

Section 2: Online data collection using SurveyMonkey

Thank you for agreeing to participate in this study characterizing data consumer perspectives regarding the quality of person-generated health outcome data. The information obtained from this research will assist in developing a consensus-based definition for data quality, harmonizing different end-user perspectives around a common data quality assessment framework (DQAF). Please provide your assessment of relevance and importance to every concept provided. Rank each concept's importance based upon your own views of its priority, credibility and value in the context of evaluating data fitness. Please review each concept definition carefully in the context of whether it is relevant and important to the end-user's ability to make informed decisions about a dataset's fitness for use in quality. For those concepts that you deem not relevant to a dataset's fitness, please use N/A and provide any additional details on the free-text field.

Figure C3-1: Seven-point Likert scale: Importance



Finally, please use the free text field to provide any additional contextual information on each concept's rank or other comments you wish to provide that will assist in the development of the final model. Free text fields are optional, but all information that you provide is greatly appreciated and will be evaluated in the next round. The end of the survey contains several open fields for respondents to suggest additional concepts that are not included in the original list but are thought to be important for evaluating data fitness.

Definitions

Before completing the survey, please familiarize yourself with this set of definitions:

Patient generated health data (PGHD) - are health-related data created, recorded, or gathered by or from patients (or patient caregiver). PGHD include, but are not limited to, health history, treatment history, symptoms, biometric data and patient-reported outcomes or outcome measures. PGHD are distinct because patients generally decide how to share these data in the context of their health concerns.[1]

Patient reported outcome (PRO) – A reporting of personal health status coming directly from an individual without manipulation or interpretation of the information by another.[2]

Patient reported outcome measure (PROM) – Instruments used to collect PRO information. PROMs are standardized questionnaires completed by patients which are used to inform a care team of the patient's perceptions about their health.[3]

End-user/data consumer – The target audience for data quality assessment information.

FHIR- Fast Healthcare Interoperability Resources (FHIR) are an HL7 International standard describing data formats and elements as "Resources" which enable the exchange of electronic healthcare data across institutions.

Questionnaire/QuestionnaireResponse Resources- A *Questionnaire* is an organized collection of questions intended to solicit information from patients, providers or other individuals involved in the healthcare domain. A Questionnaire can guide a user through a data collection process that ensures appropriate information is collected based on answers to questions asked, how they are ordered and grouped, any intervening instructional text and what the constraints are on the allowed answers. The results of a *Questionnaire* can be communicated using the *QuestionnaireResponse* Resource.[4]

OMOP- The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) standardizes the format and content of observational data, allowing capture and use of healthcare information in a consistent and reliable manner for research.

Harmonized Data Quality Assessment (DQA) Terms: [5]

Conformance **–** Do data values adhere to specified standards and formats?
Completeness **–** Are data values present?
Plausibility **–** Are data values believable?

Data Quality Concepts for Review and User Feedback

Please review each concept definition presented carefully in the context of whether it is relevant and important to the end-user using the information to make informed decisions. Several concept types are to be considered including element-level data, statistical parameters as well as summary display information. For test concepts where you normally rely on a pre-specified parameter to make a judgement, please include in the text box both the name, the pre-specified threshold typically applied in your setting for this test and the reference for the selected parameter.

Concept Name	Concept Description	Concept Type	Data Quality Domain
Database Format	Format of source dataset used to assess PGHD fitness (e.g., FHIR, JSON, SQL, CQL, etc.)		Conformance
Database Standard Format	A yes or no value indicating the dataset used to evaluate PGHD fitness is stored in a standard format that is useable, locatable, feasible to access and process	Test & Parameter	Conformance
Database Standards Use	Version of standards used to store information (e.g., FHIR 4.0.1, CQLv1.5.1, etc.)	Summary Information	Conformance
Database Format Interoperability	A yes or no value indicating the data format used meets ONC interoperability requirements for data exchange internally and externally (e.g., USCDI)	Test & Parameter	Conformance
Patient Demographics	The number and proportion of records with a value present in Patient Demographics fields (e.g., age, gender, ethnicity, primary language, culturally relevant information)	Data Element	Conformance Completeness Plausibility

Concept Name	Concept Description	Concept Type	Data Quality Domain
Patient Directives	The number and proportion of records containing documentation of patient choices, (consent) permitting or denying recipient(s) or recipient role(s) to perform one or more actions within a given policy context, for specific purposes and periods of time.	Data Element	Conformance Completeness Plausibility
Data Source Type	The number and proportion of records containing source of PGHD information (person, medical device, wearable device, other)	Data Element	Conformance Completeness Plausibility
Data Source Provenance	The number and proportion of records containing metadata on the activities taken upon the source information to transform to the target dataset. Provenance is prepared by the application that initiates the create/update of the PGHD	Data Element	Conformance Completeness Plausibility
Audit Log: Data Record Management	A yes or no value indicating the presence of the records management log assuring privacy and security safeguards are maintained and functioning.	Test & Parameter	Conformance
Audit Log: Data Quality Assessments	Number and type of measures deployed that are monitoring the quality of the target data at any point in time (ratio of data to errors, ETL transformation errors rate, data time to value, mandatory field enforcement, number of in-line corrections)	Summary Information	Conformance
Audit Log: DQA Report	Presentation of DQ measure results exposing discernable patterns of data inconsistencies and/or systematic errors	Summary Information	Plausibility
Audit Log: Provenance	A yes or no value indicating the presence of provenance information	Test & Parameter	Completeness
Audit Log: Persistence	A yes or no value indicating the persistence of provenance information that verifies the data is true to original form in the end user interface.	Test & Parameter	Conformance
Encounter Type	The number and proportion of records documenting the type of encounter during which health data was generated	Data Element	Completeness
Encounter Location	The number and proportion of records documenting the location where the PGHD collection takes place and/or the mode of data collection (face to face, asynchronous via online form, other)	Data Element	Completeness

Concept Name	Concept Description	Concept Type	Data Quality Domain
Encounter class	The number and proportion of records documenting Period (Date/Time) information for Encounter Type	Data Element	Completeness
Encounter reference	The number and proportion of records relating the frequency of patient contributions to their health record for any PGHD element. (e.g., how many times PHQ-9 appears within a certain period for an individual.)	Data Element	Completeness Plausibility
Encounter performer	The number and proportion of records containing the role\licensure of person documenting or collecting the PGHD in record (patient, Care Manager, MD, RN, etc.)	Data Element	Completeness
Questionnaire	The number and proportion of records documenting standard instrument type used to collect PGHD	Data Element	Completeness Conformance
Questionnaire Data Type	A yes or no value indicating if the Questionnaire is the expected data type based on the technical specification.[6]	Data Element	Conformance
Questionnaire Response Value	The number and proportion of records documenting presence of total score result indicating complete administration of questionnaire-asserting the structured group of questions that comprise the instrument have been successfully answered and a valid total score was calculated and present in the record.	Data Element	Completeness Plausibility
Questionnaire Response metadata	The number of records that contain supplemental information on acceptable use parameters for the instrument used (legal rights of use info, valid scoring range, positive thresholds, etc.).	Data Element	Completeness
Questionnaire Response Null Value	The number and percent of records with a NULL value in the QuestionnaireReponse value field (considered not nullable)[6]	Data Element	Completeness Conformance Plausibility
Questionnaire Response Zero Value	The number and percent of records with a value of 0 in the QuestionnaireReponse value field[6]	Data Element	Completeness Conformance Plausibility
Questionnaire Response: Missing Value	The number and percent of records with missing QuestionnaireReponse values	Data Element	Completeness Plausibility
Questionnaire Response: Value Threshold	The prespecified threshold (expected) for the QuestionnaireReponse value field	Test & Parameter	Conformance

Concept Name	Concept Description	Concept Type	Data Quality Domain
Data Quality: Thresholds	The reference to a technical specification or relational dataset that defines pre-specified thresholds asserting sufficient accuracy and precision for any specific use case	Test & Parameter	Conformance
Questionnaire Response: Value Low	The number of records that have a value that is lower than a pre-specified threshold.[6]	Data Element	Conformance Plausibility
Questionnaire Response: Value High	The number of records that have a value that is higher than a pre-specified threshold.[6]	Data Element	Conformance Plausibility
Questionnaire Standard Terminology	The number and percent of records that have a standard, valid term present in all Questionnaire and Questionnaire Response fields as specified. (Use of preferred standard terminology to document instrument and result.)	Data Element	Conformance Plausibility
Data Quality: Context Specificity	Use case pre-defined for correction and enhancement of datasets linked to its intended use.	Test & Parameter	Conformance
Data Quality: Adherence to Target	A yes or no value indicating target dataset complies with the pre-specified end-user expectations for intended use case	Summary Information	Conformance
Data Quality Test Type	Test type used to ascertain dataset adherence	Test & Parameter	Conformance
Dataset Governance Reference	Reference information used in the assessment and standardization of information within the dataset	Test & Parameter	Conformance
Data Quality: Dataset Governance Process	Documentation of the process that characterizes the necessity of data enhancement based on standard reference information.	Summary Information	Conformance
Data Quality: Measurement Report	A report that characterizes how well source data transformation to target dataset meets end user expectations.	Summary Information	Conformance Plausibility
Data Quality Criterion	What criterion/test is used to ascertain data dependencies?	Test & Parameter	Conformance Plausibility
Data Quality Dependency Report	Report of data dependencies results	Summary Information	Conformance Plausibility
Actionability	what do errors and inconsistencies signify		
Workflow	Workflow practices that must exist for end users to “trust” the information is accurate and can be used for clinical quality		