eGEMs
Generating Evidence & Methods to improve patient outcomes

Reporting Data Quality Assessment Results: Identifying Individual and Organizational Barriers and Solutions

Tiffany Callahan, MPH; Juliana Barnard, MA; Laura Helmkamp, MS; Julie Maertens, PhD; Michael Kahn, MD, PhD[i]

ABSTRACT

Introduction: Electronic health record (EHR) data are known to have significant data quality issues, yet the practice and frequency of assessing EHR data is unknown. We sought to understand current practices and attitudes towards reporting data quality assessment (DQA) results by data professionals.

Methods: The project was conducted in four Phases: (1) examined current DQA practices among informatics/CER stakeholders via engagement meeting (07/2014); (2) characterized organizations conducting DQA by interviewing key personnel and data management professionals (07-08/2014); (3) developed and administered an anonymous survey to data professionals (03-06/2015); and (4) validated survey results during a follow-up informatics/CER stakeholder engagement meeting (06/2016).

Results: The first engagement meeting identified the theme of unintended consequences as a primary barrier to DQA. Interviewees were predominantly medical groups serving distributed networks with formalized DQAs. Consistent with the interviews, most survey (N=111) respondents utilized DQA processes/programs. A lack of resources and clear definitions of how to judge the quality of a dataset were the most commonly cited individual barriers. Vague quality action plans/expectations and data owners not trained in problem identification and problem-solving skills were the most commonly cited organizational barriers. Solutions included allocating resources for DQA, establishing standards and guidelines, and changing organizational culture.

Discussion: Several barriers affecting DQA and reporting were identified. Community alignment towards systematic DQA and reporting is needed to overcome these barriers.

Conclusion: Understanding barriers and solutions to DQA reporting is vital for establishing trust in the secondary use of EHR data for quality improvement and the pursuit of personalized medicine.

[i]University of Colorado Anschutz Medical Campus

Introduction

By 2020, the volume of health data generated worldwide is expected to reach 25,000 petabytes, a 50-fold increase from the amount of data generated in 2012.[1] The Health Information for Economic and Clinical Health (HITECH) Act,[2,3] the Patient Protection and Affordable Care Act,[4] and the American Recovery and Reinvestment Act[5] facilitated an unprecedented collection of electronic health data[6,7] in an effort to improve patient care. As a result of HITECH, the adoption and use of electronic health records (EHRs)[3] has drastically increased. At the same time, rapid advancements in genomic technology, and personal health data produced from initiatives like the Quantified Self movement (e.g., continual sharing of personal data via wearable devices and social networking) have made tremendous amounts of diverse health data available for secondary use by clinicians and clinical researchers.[8,9] Yet there remain barriers that prevent the meaningful use of these data.[7,10]

The progression of science is founded on the tenets of reproducible research, and data quality (DQ) issues pose a serious threat to both the validity and generalizability of research findings.[11] A recent survey investigating the factors associated with the reproducibility of research performed with EHR data found that an assessment of DQ dimensions is required in order to determine if the data support the author's conclusions.[12] Research findings and results lacking reproducibility have serious consequences including the retraction and/or amending of published manuscripts.[13–15] Clinical research based on EHR data of unknown quality can also have consequences for patient safety and care quality.[16–20]

The utility of EHR in data-driven decision making and comparative effectiveness research (CER) faces many challenges including: governance and stakeholder misalignment[21] and lack of mutually accepted guidelines for data sharing,[6,22] data integration and aggregation,[23] DQ,[18,24–27] data standards,[24] infrastructure and data management,[6,21] and record linkage.[18] Further, guidelines for security and privacy[6,10,28] and analytics and tool development are under-developed.[21,29,30] The most informative health data (e.g., data stored in EHRs, personal wellness devices, and medical research applications) are not 'off-the-shelf' ready for analysis;[31] these data are subject to different regulations and standards, and often violate fundamental assumptions of DQ.[26,32] Ensuring transparency regarding the fitness for use of these data for analytic studies is fundamental to the responsible utilization of EHR data.

To help address the issues around DQ transparency the Evidence, Data, and Methods (EDM) Forum supported the creation of a Data Quality Collaborative.[33] The primary objective of this collaborative was to create a data quality assessment (DQA) framework and guidelines specific to the CER community.[33] Uniting the many efforts dedicated to validating EHR data,[27,34–39] this collaborative has developed and published a harmonized DQ terminology[27] as well as standards for reporting DQA results,[11] *yet* standards for assessing and reporting DQ issues have yet to be thoroughly identified.

A number of potential barriers may impede institutional investments in DQA activities. For example, methods for assessing and reporting DQ lack standardization; organizational stakeholder requirements may mandate the utilization of very different tools, reporting methods, and assessment strategies than are used by professional researchers aiming to answer specific clinical questions. Individual and organizational priorities or constraints also may impact how DQ is assessed and how findings are reported. Attempts to understand barriers to performing DQA and reporting DQ findings remain largely unexplored; to the best of our knowledge, there is no existing literature examining

DQA and reporting practices employed in the field nor any work which investigates the barriers to performing DQA analytics and reporting DQA results. The primary goal of this work was to gain insight into the current state of DQA conducting and reporting practices in the field of biomedical informatics. To accomplish our goal, we utilized a multi-phase mixed-methods approach.

Methods

The project was conducted in four Phases (Figure 1):

- **Phase 1: First Stakeholder Engagement Meeting -** examined current DQA and reporting practices by analyzing discussions between informatics/CER stakeholders at an engagement meeting.
- **Phase 2: Interviews -** interviewed key personnel and data management professionals at organizations currently conducting DQA.
- **Phase 3: DQ Barriers and Solutions Survey -** developed and administered an anonymous web-based survey (DQ Barriers and Solutions Survey) to employees, data professionals, and

academic researchers currently using or interested in developing a DQA process.
- **Phase 4: Second Stakeholder Engagement Meeting -** validated survey results during a follow-up informatics/CER stakeholder engagement meeting.
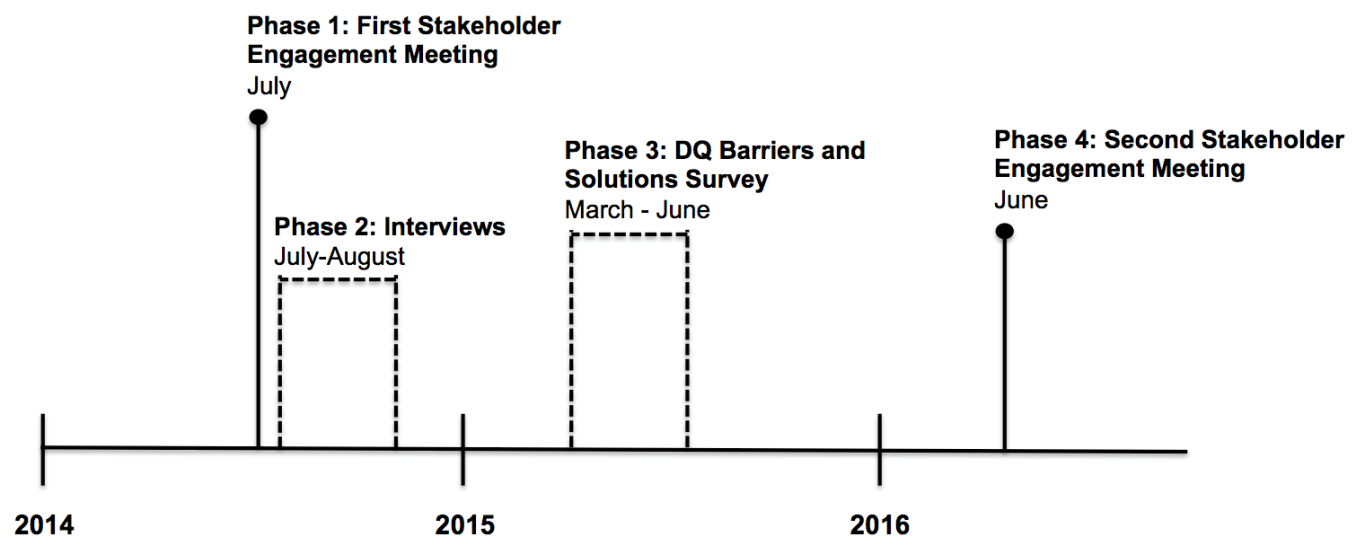
Phases 2 and 3 were developed using insight gained from the previous phases. The phases of this project were reviewed and approved by the Colorado Multiple Institutional Review Board (COMIRB #13-2917 and #14-1812).

Phase 1: First Stakeholder Engagement Meeting

Participants

The first stakeholder engagement meeting was held in Washington, D.C. in July of 2014. Meeting attendees included CER/informatics professionals from universities, research institutions, professional organizations, federal government agencies, an insurance company, and healthcare institutions.

Figure 1. Project Timeline

Phase 1: First Stakeholder Engagement Meeting
July

Phase 2: Interviews
July-August

Phase 3: DQ Barriers and Solutions Survey
March - June

Phase 4: Second Stakeholder Engagement Meeting
June

2014   2015   2016

Discussion

Qualitative methods were used to moderate the meeting. A researcher led the meeting using a discussion guide developed by the study team. The discussion guide included broad, open-ended questions and prompts to elicit detailed descriptions of the stakeholders' views and experiences. The meeting collected recommended additions and changes to a proposed DQA framework and guidelines specific to the CER community, and discussed limitations and implications of conducting DQA and reporting DQ results. Three established models of DQA systems (National Patient-Centered Clinical Research Network, the Observational Medical Outcomes Partnership's (OMOP) ACHILLES Heel, and the University of Washington's Find It) were showcased to elicit modifications.

Data Collection

Discussion amongst stakeholder meeting attendees was digitally recorded following consent from all meeting participants. The recordings were professionally transcribed verbatim and imported into ATLAS.ti (qualitative analytic software). Data were analyzed using qualitative content methods[40] and reflexive team analysis which emphasizes inclusion of emergent rather than *a priori* themes. This approach utilizes the broader study team to discuss emergent understandings of the data and check on analysts' preconceived assumptions and biases to identify themes and subthemes from the meeting discussions.[41]

Phase 2: Interviews

In July-August 2014, several sites engaged in DQA and reporting were contacted and asked if they would be willing to host a site visit and be interviewed regarding their current practices. Informed by the findings from Phase 1, the purpose of these interviews was to elicit information regarding each of the sites' current DQA and reporting efforts and investments.

Interview Questions

Using a previously established DQA framework,[42] a semi-structured interview instrument consisting of 30 questions was developed to elicit information regarding each of the sites' current DQA and reporting efforts and investments. The interview contained questions about current practices and stakeholder standards within the interviewee's organization, DQA requirements, DQA strategies, and remediation plans. Questions were pilot tested with biomedical professional researchers with past DQA experience and reviewed for applicable content, clarity, and completion time.

Data Collection

All interviews were conducted one-on-one and took approximately 30 minutes to complete. Detailed notes were taken during interviews and all participants were re-contacted after the initial interview and offered the opportunity to clarify their responses. To protect the privacy of the participating personnel and sites, names and locations were anonymized.

Data Analysis

Interview responses were aggregated within each site. Current DQA analytics and reporting practices at each of the four sites were evaluated descriptively. Iterative thematic content analysis was performed on the interview notes.

Phase 3: DQ Barriers and Solutions Survey

A survey was developed using the findings from Phases 1 and 2. Given the sensitivity around potential consequences of reporting negative DQA findings it was important to provide a mechanism that encouraged participants to answer questions

honestly. For these reasons, an anonymous survey was utilized. Prior to administration, questions were pilot tested with biomedical professional researchers with past DQA experience and reviewed for applicable content, clarity, and completion time. The final version of the anonymous web-based survey contained 44 self-report items organized within six separate subsections and was administered March-June, 2015 ("DQ Barriers and Solutions Survey" in Appendix B: Supplemental Materials).

Survey Questions

To ensure participants felt comfortable completing the survey, a response option of "I do not feel comfortable answering this question" was included for every item. In addition to questions about individual and organizational barriers and DQA conducting and reporting solutions described below, participants were asked four questions regarding demographics, five questions regarding current employment, and eight questions about current DQA practices.

Individual and Organization Barriers to Conducting and Reporting DQAs

Based on findings from Phases 1 and 2, questions about individual and organizational barriers to conducting and reporting DQAs as well as potential solutions to these barriers were developed. A five-point Likert scale ("Strongly Disagree" to "Strongly Agree") was used to examine agreement to 11 potential individual barriers. The organizational barriers questionnaire was created by modifying items from a questionnaire created to assess the barriers of implementing quality management in service organizations in Pakistan.[43] The same five-point Likert scale was used to examine agreement to nine potential organizational barriers. Higher scores indicated a greater perceived individual or organizational barrier.

Solutions to Conducting and Reporting DQAs

A four-point scale ("None" to "A Lot") was used to assess seven potential DQA analytics and reporting solutions. Higher scores indicated greater perceived organizational support for conducting and reporting DQAs. Additionally, participants were asked to provide any other solutions they felt would support conducting and reporting of DQAs.

Recruitment

Participants between the ages of 18-89 who currently work with data as a producer (i.e., someone who generates data) and/or consumer (i.e., someone who uses data generated by a data producer) were recruited. No identifying information (e.g., name, birthdate, social security number) was collected from participants and participation was completely voluntary. Following the Dillman method of survey research, participants received up to five reminder emails; participants were emailed once a week for up to five weeks starting from the date the initial email was sent.[44]

Data Collection

Study data was collected and managed using REDCap (Research Electronic Data Capture).[45] Interested participants were first presented with information about the survey and those who gave consent then proceeded to complete the rest of the survey. The database is hosted at the University of Colorado Denver Anschutz Medical Campus.

Statistical Analysis

Analysis was performed using SAS software (version 9.4). Univariate statistics were used to examine the frequencies of responses to survey questions. Chi-squared and Fisher exact tests were performed to investigate whether responses differed by job characteristics, including the numbers of hours per week spent working on DQ issues (0-9 hours/week

versus 10 or more hours/week), the type of position (data producer, data consumer, or both), and the number of data sites participating in the participant's network (small: 1-20 sites versus large: more than 20 sites).

For bivariate analyses, the responses were collapsed into two categories, "Strongly Agree/Agree" versus "Strongly Disagree/Disagree/Neutral" for all individual and organizational barriers items and "None/Some" versus "Mostly/A Lot" for all solutions items, to allow for statistical testing given the limited sample size.

Exploratory factor analysis, with principle components analysis as the method of factor extraction and varimax rotation, was used to explore the individual and organizational barriers scales for correlated variables that reflected an underlying factor structure in the data. The Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity were used to determine factorability of the two sets of items. Items that loaded strongly on a factor were averaged into a subscale for each participant; this method was selected over analyzing factor scores as it allowed the subscale values to be interpreted using the Likert scale of the original item responses. Analysis of Variance (ANOVA) was used to compare the overall individual and organizational barriers scales, as well as the subscales derived from factor analysis.

Phase 4: Second Stakeholder Engagement Meeting

Participants

The second stakeholder engagement meeting was held in Washington, DC in June of 2016. Attendees were similar to the first stakeholder engagement meeting and included: CER/informatics professionals from universities, research institutions, professional organizations, federal government agencies, an insurance company, and healthcare institutions.

Roughly half of the attendees had participated in Phase 1.

Discussion

Qualitative methods and group discussion procedures were similar to Phase 1. Attendees were presented with a summary of the DQ Barriers and Solutions Survey results. During the first part of the discussion, the research team asked stakeholders to report if the survey results were consistent or if they conflicted with their individual and organizational work as producers or consumers of data. The second part included a 'gallery walk', where attendees could circulate, view survey-identified barriers and solutions on a whiteboard, and validate, comment on, or add barriers and solutions based on their own work experiences.

Data Collection

Discussion amongst stakeholder meeting attendees was collected, recorded, and analyzed using the same procedures outlined for the first stakeholder engagement meeting (Phase 1).

Results

Phase 1: First Stakeholder Engagement Meeting

While the first stakeholder engagement meeting (n=23) elicited stakeholder's opinions on current DQA practices and the critique of a new DQA framework, the discussion inductively elicited identification of professional barriers and solutions to conducting and reporting DQAs. The meeting attendees identified unintended consequences as a primary barrier to conducting and reporting DQAs. Attendees were concerned that poor DQA results could spur questions regarding the organization's caliber of care, could result in the potential loss of future funding, collaboration opportunities, could raise questions about publications used for professional promotion, or could result in damage

to the organization's reputation. When discussing solutions that would address DQA barriers, participants recommended enacting guidelines to protect against negative repercussions when reporting DQA findings, developing remediation plans to deal with DQA-related issues, requiring peer-reviewed journals to mandate the inclusion of DQA results with submissions, and identification of DQA resources by institutions and funding agencies. Importantly, attendees believed that systematically conducting and reporting DQA would require a culture change in expectations regarding how DQAs are conducted, reported, and the interpretation of findings.

Contributions to the Development of Phase 2

The findings from this phase provided valuable insight into stakeholder identified DQA barriers. Attendee's also provided a range of solutions that they felt would help to alleviate some of the barriers that they identified. Interpretation of these findings was limited by a lack of information about the organizations (e.g., internal vs. external stakeholder management, existing/desired DQA processes, and tools/software used). Phase 2 was designed to elicit this type of information from well-established private sector and academic organizations currently engaged in this type of work.

Phase 2: Interviews

A total of 19 interviews were completed in four separate site visits. Participating sites included one academic site (Site 1: N=3), two medical groups (Sites 2: N=1 and 3: N=7), and one commercial research site (Site 4: N=8). An overview of site characteristics can be found in Table 1. Participants were asked to discuss key DQ dimensions without prompting from a list of DQ elements. This was done intentionally to reduce confirmation bias during interviews. The number of DQA employees varied across sites (n=1-8 employees). All four sites utilized one or more common data model (CDM) (OMOP, Sentinel, and the Health Care Systems Research Network (HCSRN) Virtual Data Warehouse (VDW)). Sites 2-4 had a standardized remediation plan in place to resolve DQ anomalies.

Sites were compared according to whether they were managed by internal stakeholders (Site 4), external stakeholders (Site 2), or internal and external stakeholders (Sites 1 and 3). Internally/externally managed sites tended to have fewer DQA employees, did not use custom tools, and valued completeness, consistency, and timeliness when evaluating DQ. Site 2 was externally managed and valued data validity more than any other DQ dimensions. This site used a custom CDM and reporting tools as well as employed an entire team dedicated to conducting DQA.

Site 4 was internally managed and was the only site that conducted commercial research within a large collaborative of independent researchers. Like Site 2 (managed by an external stakeholder), this site had developed its own CDM and DQA conducting and reporting tools. While this site employed a team to conduct DQA, its outside collaborators conducted their own independent DQA work.

Contributions to the Development of Phase 3

Results from the interviews (Phase 2) facilitated a general characterization of sites currently conducting DQAs. A general theme that arose across all sites was the notion that DQA barriers and solutions could exist at both the level of the individual as well as the organization. Phase 3 was designed to elicit more specific information regarding DQA barriers and solutions at both the individual and organizational levels. This assessment also included questions about demographics, employment characteristics, and current DQA practices.

Table 1. Interview Participants and Site Characteristics

CHARACTERISTIC	SITE 1	SITE 2	SITE 3	SITE 4
Site type	Academic	Medical group	Medical group	Commercial
# DQ employees	1	8	2	8
Stakeholders[a]	Internal/External	External	Internal/External	Internal
Common data model[b]	OHDSI, i2b2	Sentinel	HCSRN VDW	OHDSI
Network type	Site-based	Distributed	Distributed	Site-based
Formal DQ plan	No	Yes	Yes	Yes
DQA tools	R, Python, SAS, MySQL, DB2, OHDSI tools[c]	Sentinel tools[d], Excel	SAS	OHDSI tools[c], R
Visualization tools	R, Google Graphics, Spotfire, Cytoscape, OHDSI tools[c]	COMPARE[d]	SAS, Excel	OHDSI tools[c], R, Spotfire
Key DQ dimensions	Completeness, Consistency, Correctness, Accuracy, Timeliness	Validity, Completeness, ETL processing integrity	Completeness, Timeliness, Consistency, ETL processing integrity	Accuracy, Completeness, Consistency
DQ remediation plan	No	Yes	Yes	Yes

[a]An internal stakeholder functions within an organization (e.g., employees, researchers, and/or project managers). In contrast, an external stakeholder (e.g., regulatory organizations, community members, and government agencies) has functionality that is external to the organization.
[b]Current CDM(s) utilized as of summer 2014. Observational Health Data Sciences and Informatics (OHDSI); Informatics for Integrating Biology and the Bedside (i2b2); Sentinel; The Health Systems Research Network (HSCRN) Virtual Data Warehouse (VDW).
[c]OHDSI tool information can be on the OHDSI home page: http://www.ohdsi.org/.
[d]Sentinel Data QA SAS tools and information can be found here: http://www.mini-sentinel.org/.

Phase 3: DQ Barriers and Solutions Survey

Of the 141 participants who provided consent, 30 did not complete the survey and were thus excluded from analysis. The remaining 111 participants completed the entire survey (with < six missing responses). Demographic characteristics are reported in Table 2. Most the sample was white or Asian, middle-aged (31-50 years old), and highly educated (i.e., had a PhD or MD).

A majority of the participants reported currently implementing DQA processes and most planned to share their findings outside of their organization (Table 3), only about half used a CDM, and less than half were required to assess DQ by funders or stakeholders. Most important and most commonly evaluated aspects of DQ were consistency and completeness of the data.

DQA and Reporting Barriers

Responses to individual barriers are presented in Appendix A (Figure A1). Over three-fourths of the participants strongly agreed or agreed that an important individual barrier was a lack of resources (i.e., funding or time to complete DQAs). A fear that DQ would invalidate prior research or fear of colleagues leaving collaborations was considered less frequently. Responses to the DQA organizational barriers are presented in Appendix A (Figure A2). Nearly two-thirds of the participants (64 percent) strongly agreed or agreed that the quality action plans/expectations are vague and a lack of data owner training in problem identification and problem solving skills were noted as barriers to expanding or improving DQA. Excess layers of interfering management, frequent data owner turn over, and cost of implementing DQA outweighing the benefits, were cited less frequently.

Most of the responses to the barrier questions did not significantly differ by participant job characteristics (i.e., number of hours/week spent working on DQ issues, position type, or the number of data sites participating in the participant's network). Those who participated in small networks (88 percent; 1-20 sites) were significantly more likely to agree or strongly agree that a "culture change in how data are currently managed, assessed, and reported" was needed ($p<0.01$) than those in large networks (50 percent; more than 20 sites).

Responses to the survey items regarding solutions to encourage the conducting and reporting of DQAs are shown in Appendix A (Figure A3). The majority of the participants (>85%) felt that having available resources and having DQA be part of the standard process of the network would encourage them mostly or a lot. The least encouraging solution was professional and financial protections.

Factor Analysis

Individual and organizational barrier item factorability was confirmed by a large number of correlations among the items (>0.3), a KMO measure of sampling adequacy of 0.80 for individual barriers and 0.81 for organizational barriers, and a significant Bartlett's test of sphericity $p<0.0001$ for both scales. Three factors were selected for the individual barriers and two factors for the organizational barriers using the Kaiser's criterion (eigenvalue >1) and an assessment of the scree plot; these factors accounted for 61 percent and 54 percent of the total variance in the individual and organizational barriers scales, respectively (Table A1 in Appendix A).

For individual barriers (Figure 2), three factors were obtained. The first factor (*Personal Consequences*) comprised six items covering serious career consequences as a result of DQA, including invalidating prior work or colleagues leaving collaborations. The second factor (*Process Issues*) comprised three items and concerned the implementation of DQA in the analysis pipeline.

Table 2. Description of Participant Demographic and Employment Characteristics

VARIABLE	N	%
WHAT IS YOUR GENDER?		
Female	56	50.5
Male	53	47.7
Missing/ unknown	2	1.8
HOW OLD ARE YOU?		
30 or younger	14	12.6
31-50	62	55.9
51 or older	33	29.7
Missing/ unknown	2	1.8
WHAT IS YOUR RACE?		
White	89	80.2
Asian/Pacific Islander	9	8.1
Hispanic/Latino	2	1.8
Multiracial	5	4.5
Missing/ unknown	6	5.4
WHAT IS THE HIGHEST LEVEL OF EDUCATION YOU HAVE COMPLETED?		
College graduate	13	11.7
Master's Degree	31	27.9
PhD/MD	67	60.4
HOW MANY HOURS A WEEK DO YOU WORK ON DATA QUALITY-RELATED ISSUES?		
0-9 hours/week	51	45.9
10-19 hours/week	34	30.6
20-39 hours/week	16	14.4
40 or more hours/week	10	9.0

Table 2. Description of Participant Demographic and Employment Characteristics (Cont'd)

VARIABLE	N	%
HOW LONG HAVE YOU BEEN WITH YOUR CURRENT EMPLOYER?		
6 months - 1 year	11	9.9
1 year - 2 years	10	9.0
2 years - 4 years	31	27.9
5 years or more	58	52.3
Missing/ unknown	1	0.9
WHAT TYPE OF POSITION TO DO YOU CURRENTLY HOLD?		
Data Producer i.e., creates/populates databases for a consumer/analyst	14	12.6
Data Consumer i.e., uses data that is provided by a data producer	38	34.2
Both a Data producer and Consumer	54	48.6
Missing/ unknown	5	4.5
HOW MANY CUSTOMERS USE YOUR DATA PER MONTH?		
0-25 customers/month	38	34.2
26-50 customers/month	4	3.6
51-100 customers/month	5	4.5
Greater than 100 customers/month	10	9.0
I am not sure how many consumers use the data that I produce each month	9	8.1
Missing/ unknown	45	40.5
HOW MANY DATA PARTNERS OR SITES PARTICIPATE IN YOUR NETWORK?		
1-4 sites	32	28.8
5-10 sites	21	18.9
11-14 sites	9	8.1
15-20 sites	10	9.0
More than 20 sites	21	18.9
Does not apply	16	14.4
Missing/ unknown	2	1.8

Table 3. Description of Current Data Quality Assessment Practices

VARIABLE	N	%
WHICH COMMON DATA MODEL DO YOU CURRENTLY UTILIZE I.E., OMOP, HMORN HCSRN VDW, MINI-SENTINEL, I2B2, PCORNET, OPEN EHR, OTHER IN YOUR ORGANIZATION?		
OMOP	16	14.4
HCSMORN VDW	7	6.3
i2b2	8	7.2
PCORnet	6	5.4
OpenEHR	2	1.8
Other	15	13.5
I do not utilize a common data model in my organization	53	47.7
Missing/ unknown	4	3.6
DO YOU CURRENTLY IMPLEMENT ANY PROCESSES TO VERIFY OR ASSESS THE QUALITY OF YOUR DATA EITHER PRODUCED OR CONSUMED/USED BY YOUR ORGANIZATION?		
No	14	12.6
Yes	94	84.7
Missing/ unknown	3	2.7
DO YOU CURRENTLY HAVE ANY DATA QUALITY REQUIREMENTS DICTATED BY YOUR PRIMARY FUNDERS OR STAKEHOLDERS IN YOUR ORGANIZATION?		
No	57	51.4
Yes	52	46.8
Missing/ unknown	2	1.8
THINKING ABOUT THE FIVE MOST RECENT ANALYSIS PROJECTS YOU COMPLETED IN YOUR ORGANIZATION, ON AVERAGE, WHAT PORTION OF TIME DURING AN ANALYSIS DID YOU SPEND DOING DATA QUALITY ASSESSMENT?		
< 10%	22	19.8
10-40%	53	47.7
41-70%	17	15.3
71-100%	9	8.1
Does not apply	7	6.3
Missing/ unknown	3	2.7

Table 3. Description of Current Data Quality Assessment Practices (Cont'd)

VARIABLE	N	%
HOW LIKELY ARE YOU TO ASSESS THE QUALITY OF THE DATA YOU WORK WITH IN THE NEXT 12 MONTHS?		
Probably won't happen	4	3.6
Probably will happen	19	17.1
Definitely will happen	86	77.5
Missing/ unknown	2	1.8
IF YOU ARE ALREADY ASSESSING THE QUALITY OF THE DATA YOU WORK WITH, HOW LIKELY ARE YOU TO SHARE A DATA QUALITY ASSESSMENT REPORT WITH INDIVIDUALS WHO ARE NOT A PART OF YOUR ORGANIZATION IN THE NEXT 12 MONTHS?		
Will not happen	9	8.1
Probably won't happen	28	25.2
Probably will happen	45	40.5
Definitely will happen	20	18.0
I am not currently assessing the quality of the data I work with	4	3.6
WHEN DETERMINING THE VALUE OF A DATA QUALITY REPORT, WHICH DATA QUALITY CATEGORIES ARE MOST IMPORTANT TO YOUR ORGANIZATION?		
Consistency	80	72.1
Completeness	83	74.8
Non-redundancy	40	36.0
Processing Integrity	64	57.7
Quality of Documentation and Metadata	51	45.9
PRIOR TO CONDUCTING AN ANALYSIS, WHICH DATA QUALITY CATEGORIES DO YOU MOST OFTEN EVALUATE WHEN DETERMINING THE APPROPRIATENESS OF A DATASET IN YOUR ORGANIZATION?		
Consistency	71	64.0
Completeness	84	75.7
Non-redundancy	35	31.5
Processing Integrity	50	45.0
Quality of Documentation and Metadata	31	27.9

The third factor (*Lack of Resources*) contained just two items, dealing with a lack of funding, time, or knowledge in conducting DQAs.

For organizational barriers (Figure 3), two factors were obtained. The first factor (*Environment/Support*), contained four items describing limits on DQA imposed by other parties, such as data owners, investigators, management, and funding agencies. The second (*Practices*) contained three items and covered issues such as a lack of "best practices" for DQ and vague DQ requirements. Two organizational barriers items were excluded from the factors due to strong cross-loadings.

Figure A4 (Appendix A) shows the agreement level with the overall individual barriers scale and with each of the three factor subscales. Mean values

differed significantly (*p*<0.0001), with participants tending to agree with the *Lack of Resources* items (factor 3) and disagree with the *Personal Consequences* items (factor 1). The organizational barriers showed values in the neutral-to-agree range for the overall scale and for both factor subscales.

Phase 4: Second Stakeholder Engagement Meeting

To confirm the survey results, responses were queried with stakeholders attending a second stakeholder engagement meeting. Individuals from universities (n=3), federal government agencies (n=3), healthcare research institutions (n=3), and professional healthcare organizations (n=3) were in attendance. From an individual barriers perspective, attendees agreed with survey participants. They experienced a lack of guidelines and resources for
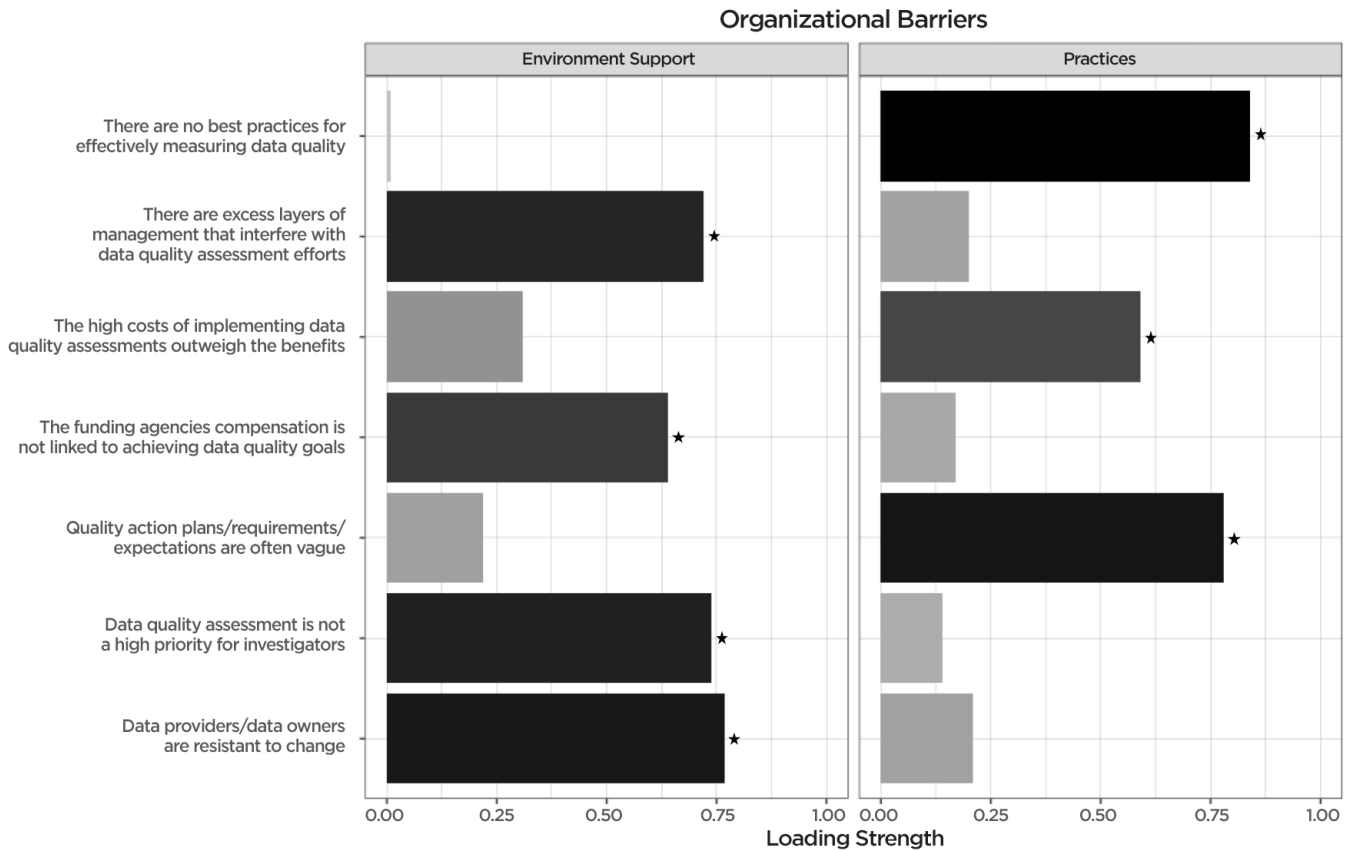
Figure 2. Factors Obtained for the Individual Barriers Questions
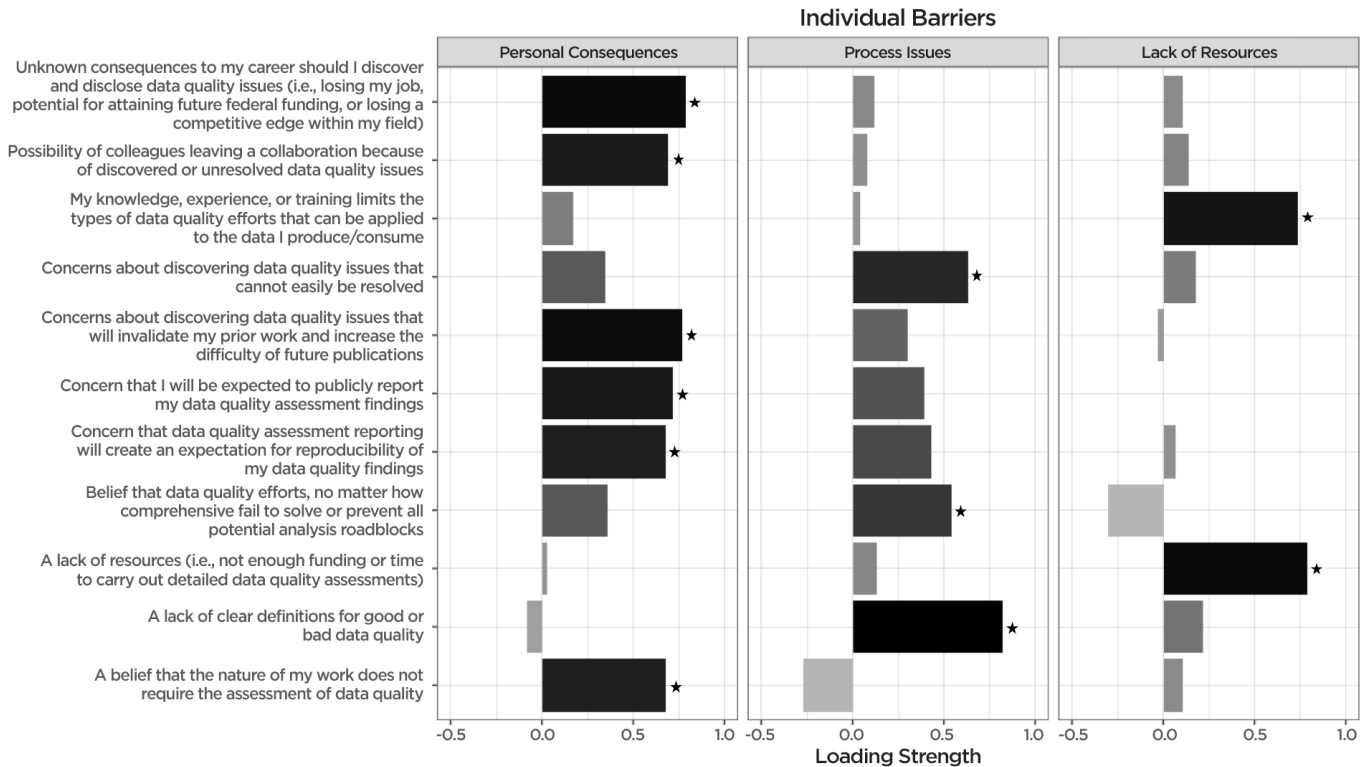
Note: Darker colored bars represent a stronger loading between a question and a factor. The questions that represent each factor are marked with a star (i.e., the Personal Consequences factor represents six items; the Process Issues factor represents three items; and the Lack of Resources factor represents two items).

Figure 3. Factors Obtained for the Organizational Barriers Questions

Organizational Barriers

| | Environment Support | Practices |

Note: Darker colored bars represent a stronger loading between a question and a factor. The questions that represent each factor are marked with a star (i.e., The Environment Support factor represents four items and the Practice factor represents three items).

conducting and reporting DQA and added that there was a general feeling of powerlessness to impact the quality of data sets received. Meeting attendees suggested possible solutions to DQA barriers, including the ability to access case studies using data findings that would demonstrate the importance and applicability of DQA to different user types. From an organizational standpoint, attendees agreed that there was an overall lack of organizational understanding about DQ issues, few resources to support DQA, and a lack of guidelines or 'gold standards' against which to compare/validate DQA results. They also added data source

issues as a barrier, where there are often difficulties determining the source, or controlling the quality, of data received.

Discussion

Applications for secondary use of EHR data are immense and include everything from investigation of rare and chronic diseases and quality improvement to repurposing of medications and hospital accreditation.[46] Assessing and reporting the quality of these data will play an important role in determining their utility. Understanding the barriers to conducting and reporting DQA

experienced by data professionals may focus efforts to alleviate current barriers and ultimately could increase trust in secondary use of EHR data, the sharing of DQA results within organizations, and the pursuit of personalized medicine. To the best of our knowledge, this exploration is the first formalized attempt to understand barriers to performing DQA and reporting DQ results. Themes of consequences of reporting DQA findings and support for DQA and reporting were consistent across all participants and phases. The findings from each phase are discussed in reference to each of these themes below.

Consequences of Reporting DQA Findings

The idea of unintended consequences resulting from reporting negative DQA findings was first discussed during the initial stakeholder engagement meeting conducted during Phase 1. Meeting attendees were concerned about personal as well as organizational consequences that could result from reporting poor DQA. For example, specific concerns expressed included damage to the organization's reputation, the potential loss of funding, and the loss of promotions. Conversely, survey participants (Phase 3) were much less concerned about potential negative personal, professional or institutional reputation-related consequences of reporting poor DQA results. The mixed findings related to consequences of reporting poor DQA between the different phases is interesting and warrants future investigation. These results imply that the consequences of reporting negative DQA findings may be more complex than initially hypothesized, operating on a level other than the individual or organization. This specific idea was not thoroughly examined in the current work.

While this work is the first to provide empirical evidence suggesting that reputation-related punitive consequences are a barrier to DQA within the field of biomedical informatics, the fear of reporting negative findings is not a new phenomenon within the medical field. A study examining adverse event reporting in hospitals found that people were unlikely to report an error due to reluctance to accuse oneself and the fear of malpractice suits.[47] While these studies investigated different populations than the current study, the reporting of the fear of punitive consequences was present across all phases of the current project. Investigating effective solutions from these different fields may be beneficial for reducing these types of barriers within the domain of clinical research informatics. For example, a study investigating the frequency of medical error reporting by pharmacists working in an inpatient setting, found that pharmacists who felt they could openly communicate were 40 percent more likely to have reported a medical error within the past year.[48]

Supporting DQA and Reporting

Similar to attendees of the first stakeholder meeting (Phase 1), survey participants (Phase 3) agreed that a strong potential barrier to DQA and reporting was a lack of adequate funding and time to perform DQAs. They also agreed that a lack of guidelines around desirable versus undesirable DQA results was a likely barrier, both to defining DQ issues and to designing appropriate DQ action plans. Although limited to a few settings, there is some initial evidence to support the effect of resources on DQA practices found from the key personnel interviews (Phase 2). Specifically, the sites utilizing customized programs and tools for conducting and reporting DQAs (Sites 2 and 4) were those who had multiple employees with dedicated time for DQA.

Potential solutions to DQA barriers identified by the survey group mirrored those identified by attendees at both the first and second stakeholder engagement meetings (Phases 1 and 4). Solutions included the need for organizational resources

and support as well as established standards and processes for conducting DQAs to help data handlers determine whether DQA and reporting guidelines have been met. Additional solutions mentioned by attendees at the second stakeholder engagement meeting (Phase 4), included the development of a set of 'gold standards' against which to validate DQA results. While attendees and participants from all phases of the project mentioned wanting better guidelines for conducting DQA, a stronger theme heard from participants was the idea of needing a significant culture change within their data community or their organization. Examples included: the development of guidelines to protect those who report DQA findings, whether they are positive or negative, the development of remediation plans for identified DQA issues, and a mandate from peer reviewed journals to require the reporting of DQAs with submitted manuscripts. This finding suggests that the DQA community should develop and adopt an infrastructure that standardizes and facilitates the conducting and reporting of DQAs. Several efforts are currently making progress in this area.

Recent work by Kahn et al. resulted in the development of a set of recommendations for reporting DQA results when conducted on observational healthcare data.[11] The framework provides a set of guidelines for characterizing the DQ of a data source to determine its fitness for use. The framework includes documentation and reporting recommendations that span data capture, processing/provenance, data element characterization, and analysis-specific DQ documentation. The same authors recently proposed a harmonized DQA terminology in an effort to encourage the standardization of different DQ characteristics.[49] Integrating the harmonized DQA terminology with the reporting framework proposed by Kahn et al. could provide a foundation for establishing a DQA and reporting architecture for secondary health data and data professionals. Finally, Callahan et al. recently proposed a DQ Maturity model as a mechanism for helping organizations and individuals establish new or improve existing DQA processes.[50] This work provides one mechanism to foster community alignment towards systematically conducting DQA work by encouraging collaboration between organizations and individuals, regardless of how mature their current DQA processes are.

Limitations

The current study has important limitations. First, we were unable to perform extensive pilot testing of the interview and survey questions using a similar population of participants. This type of testing is important as it can be used to identify potential issues in the development of the survey (e.g., confusion regarding question wording, inclusion of all important questions). Second, scale items for the survey were developed through literature reviews, expert discussions, and the modification of other measures. While these are reliable sources, the survey has yet to receive formal validation or testing. Finally, the individual and organizational barriers and solutions were based on hypothetical scenarios; the degree to which they reflect actual DQA and reporting practices is uncertain.

Future Work

Using feedback from the expert discussions, the survey items should be modified and validated. Results from the current study can be used to explore ways to incorporate needs assessment into a pragmatic DQA and reporting plan. Also, a set of common practices for individuals and organizations not currently implementing DQ checks, but who want to implement DQA and reporting practices, should be drafted. An initial set of recommendations has been published.[11] Finally, since DQ issues can

arise at different stages of data use (e.g., at time of initial extract, transform, and load, at distribution to a study investigator or analyst, or at dynamically via database query) and potentially by different individuals who are responsible for performing DQA at these different stages of use, further investigation into the needs by data use stage and by varying professional roles is warranted.

Conclusion

This study is the first of its kind, facilitating an in-depth examination of DQA practices with specific focus on individual and organizational barriers to conducting and reporting DQAs. With this survey, data from over 100 data consumers and/or producers was collected. The results from this survey facilitated the identification of several individual and organizational barriers as well as helped to identify solutions. The results of this work can be used to inform the development of DQA and reporting standards as well as provide recommendations for clinicians, clinical researchers, and organizations intending to leverage health data sources in need of DQ evaluation.

Acknowledgements

References

1. Piai S, Claps M. Bigger Data for Better Healthcare. IDC Health Insights [Internet]. 2013; Available from: http://www.intelcore.cz/content/dam/www/public/us/en/documents/white-papers/bigger-data-better-healthcare-idc-insights-white-paper.pdf

2. Office for Civil Rights. HITECH Act Enforcement Interim Final Rule [Internet]. HHS.gov. 2009 [cited 2016 Jul 12]. Available from: http://www.hhs.gov/hipaa/for-professionals/special-topics/HITECH-act-enforcement-interim-final-rule/index.html

3. Blumenthal D. Launching HITECH. N Engl J Med. 2010 Feb 4;362(5):382–5.

4. Assistant Secretary for Public Affairs, U.S. Department of Health and Human Services. The Affordable Care Act [Internet]. HHS.gov. 2015 [cited 2016 Jul 12]. Available from: http://www.hhs.gov/healthcare/about-the-law/read-the-law/index.html

5. National Institutes of Health, U.S. Department of Health and Human Services. Recovery Act Grant Information - Supported by the American Recovery & Reinvestment Act of 2009 (ARRA) [Internet]. [cited 2016 Jul 12]. Available from: http://grants.nih.gov/recovery/

6. Holve E, Segal C. Infrastructure to support learning health systems: are we there yet? Innovative solutions and lessons learned from American Recovery and Reinvestment Act CER investments. J Comp Eff Res. 2014 Nov;3(6):635–45.

7. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA. 2013 Apr 3;309(13):1351–2.

8. Bietz MJ, Bloss CS, Calvert S, Godino JG, Gregory J, Claffey MP, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc JAMIA. 2016 Apr;23(e1):e42-48.

9. Panahiazar M, Taslimitehrani V, Jadhav A, Pathak J. Empowering Personalized Medicine with Big Data and Semantic Web Technology: Promises, Challenges, and Use Cases. Proc IEEE Int Conf Big Data IEEE Int Conf Big Data. 2014 Oct;2014:790–5.

10. Gray EA, Thorpe JH. Comparative effectiveness research and big data: balancing potential with legal and ethical considerations. J Comp Eff Res. 2015 Jan;4(1):61–74.

11. Kahn M, Brown J, Chun A, Davidson B, Meeker D, Ryan P, et al. Transparent Reporting of Data Quality in Distributed Data Networks. EGEMs Gener Evid Methods Improve Patient Outcomes [Internet]. 2015 Mar 23;3(1). Available from: http://repository.edm-forum.org/egems/vol3/iss1/7

12. Richesson RL, Horvath MM, Rusincovitch SA. Clinical Research Informatics and Electronic Health Record Data. Yearb Med Inform. 2014 Aug 15;9(1):215–23.

13. Furlanello C. Emerging data waves in biomedicine: the challenge of reproducibility. Keynote Lecture presented at: Intelligent Data Analysis in Medicine and Pharmacology; 2012 Nov 22; Pavia, Italy.

14. Bellazzi R. Big Data and Biomedical Informatics: A Challenging Opportunity. Yearb Med Inform. 2014 May 22;9(1):8–13.

15. JAMA authors retract (and replace) paper about moves and kids' mental health [Internet]. Retraction Watch. 2016 [cited 2016 Jul 13]. Available from: http://retractionwatch.com/2016/06/17/jama-authors-retract-and-replace-paper-about-moves-and-kids-mental-health/

16. Kahn MG, Ranade D. The impact of electronic medical records data sources on an adverse drug event quality measure. J Am Med Inform Assoc JAMIA. 2010 Apr;17(2):185–91.

17. Sittig DF, Singh H. Electronic Health Records and National Patient-Safety Goals. N Engl J Med. 2012 Nov 8;367(19):1854–60.

18. McCoy AB, Wright A, Kahn MG, Shapiro JS, Bernstam EV, Sittig DF. Matching identifiers in electronic health records: implications for duplicate records and patient safety. BMJ Qual Saf. 2013 Mar;22(3):219–24.

19. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc. 2013 Jun 1;20(e1):e2–8.

20. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PRO, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care. 2013 Aug;51(8 Suppl 3):S30-37.

21. Randhawa GS. Building electronic data infrastructure for comparative effectiveness research: accomplishments, lessons learned and future steps. J Comp Eff Res. 2014 Nov;3(6):567–72.

22. Siu LL, Lawler M, Haussler D, Knoppers BM, Lewin J, Vis DJ, et al. Facilitating a culture of responsible and effective sharing of cancer genome data. Nat Med. 2016 May;22(5):464–71.

23. Kho AN, Rasmussen LV, Connolly JJ, Peissig PL, Starren J, Hakonarson H, et al. Practical challenges in integrating genomic data into the electronic health record. Genet Med. 2013 Oct;15(10):772–8.

24. Friedman DJ, Parrish RG, Ross DA. Electronic Health Records and US Public Health: Current Realities and Future Promise. Am J Public Health. 2013 Jul 18;103(9):1560–7.

25. 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. 2012 Jul;50 Suppl:S21-29.

26. Brown JS, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. Med Care. 2013 Aug;51(8 Suppl 3):S22-29.

27. Kahn MG, Callahan TJ, Barnard JJ, Bauck A, Brown JS, Davidson B N, et al. A harmonized data quality assessment terminology for the secondary use of electronic health record data. eGEMs.

28. Sittig DF, Singh H. Legal, Ethical, and Financial Dilemmas in Electronic Health Record Adoption and Use. Pediatrics. 2011 Apr;127(4):e1042–7.

29. Berger ML, Doban V. Big data, advanced analytics and the future of comparative effectiveness research. J Comp Eff Res. 2014 Mar;3(2):167–76.

30. Harper E. Can big data transform electronic health records into learning health systems? Stud Health Technol Inform. 2014;201:470–5.

31. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;2:3.

32. Bae C, Griffith S, Fan Y, Dunphy C, Thompson N, Parchman A, et al. The Challenges of Data Quality Evaluation in an EHR-based Data Registry. EDM Forum [Internet]. 2011; Available from: http://repository.academyhealth.org/symposia/june2014/panels/15

33. EDM Forum. Collaborative Projects - EDM Forum Home [Internet]. [cited 2016 Jul 13]. Available from: http://www.edm-forum.org/edmhome/collaborate/collaborativeprojects

34. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev MCRR. 2010 Oct;67(5):503–27.

35. 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. 2012 Jul;50 Suppl:S21-29.

36. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc JAMIA. 2013 Jan 1;20(1):144–51.

37. Weiskopf NG, Hripcsak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. J Biomed Inform. 2013 Oct;46(5):830–6.

38. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. A Data Quality Ontology for the Secondary Use of EHR Data. AMIA Annu Symp Proc AMIA Symp AMIA Symp. 2015;2015:1937–46.

39. Zozus M, Hammond W, Green B, Kahn MG, Richesson R, Rusincovitch S. Assessing Data Quality for Healthcare Systems Data Used in Clinical Research (Version 1.0) [Internet]. Available from: http://sites.duke.edu/rethinkingclinicaltrials/assessing-data-quality/

40. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today. 2004 Feb;24(2):105–12.

41. Mauthner NS, Doucet A. Reflexive Accounts and Accounts of Reflexivity in Qualitative Data Analysis. Sociology. 2003 Aug 1;37(3):413–31.

42. Sebastian-Coleman L. Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework. 1 edition. Waltham, MA: Morgan Kaufmann; 2013. 376 p.

43. Khan M. An empirical study of barriers in implementing total quality management in service organizations in Pakistan. Asian J Bus Manag Stud. 2011;2(4):155–61.

44. Dillman DA. Mail and Internet Surveys: The Tailored Design Method -- 2007 Update with New Internet, Visual, and Mixed-Mode Guide. John Wiley & Sons; 2011. 572 p.

45. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009 Apr;42(2):377–81.

46. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007;14(1):1–9.

47. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf. 2013 Oct 1;22(Suppl 2):ii21-ii27.

48. Associations Between Communication Climate and the Frequency... : Journal of Patient Safety [Internet]. LWW. [cited 2016 Dec 13]. Available from: http://journals.lww.com/journalpatientsafety/Fulltext/2013/09000/Associations_Between_Communication_Climate_and_the.3.aspx

49. Kahn M, Callahan T, Barnard J, Bauck A, Brown J, Davidson B, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMs Gener Evid Methods Improve Patient Outcomes [Internet]. 2016 Sep 11;4(1). Available from: http://repository.edm-forum.org/egems/vol4/iss1/18

50. Callahan TJ, Bauck A, Bertoch D, Brown JS, Khare R, Ryan PB, et al. A Comparison Of Data Quality Assessment Checks In Six Data Sharing Networks. eGEMS. 2017;5(1):8. DOI: https://doi.org/10.13063/2327-9214.1287.
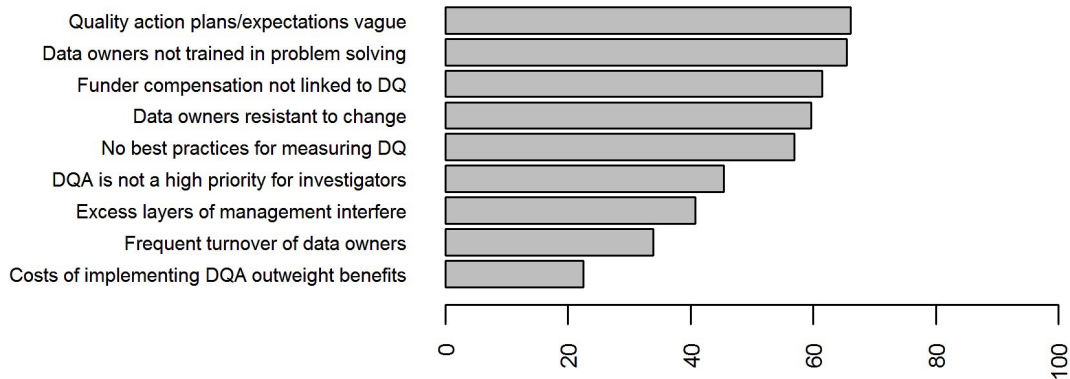
Appendix A. Supplementary Materials

Figure A1. A Current Barrier to Implementing Data Quality Processes is
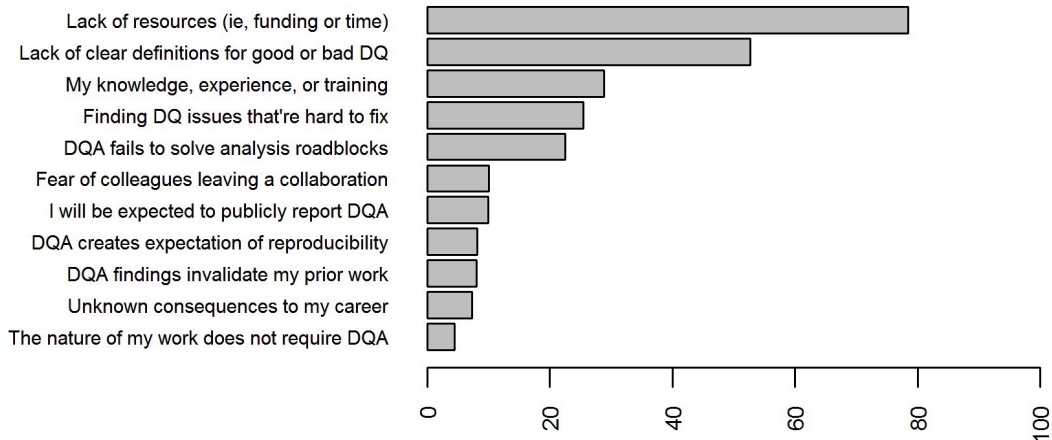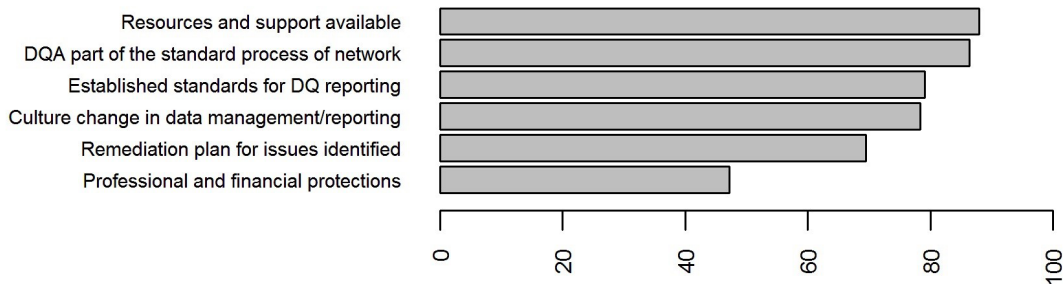

Note: Bars represent the percent of individuals who responded, "agree" or "strongly agree".

Figure A2. A Current Barrier to Implementing Data Quality Processes within my Organization is

Note: Bars represent the percent of individuals who responded, "agree" or "strongly agree".

Figure A3. How Much Would Each of the Following Encourage you and your Organization to conduct and report data quality assessments (DQAs)?

Note: Bars represent the percent of individuals who responded, "mostly" or "a lot".

Table A1. Factor Structure Matrix for Individual and Organizational Barriers Items

ITEM	FACTOR LOADINGS		
	F1	F2	F3
INDIVIDUAL BARRIERS			
Factor 1: Personal Consequences			
Unknown consequences to my career should I discover and disclose data quality issues (i.e., losing my job, potential for attaining future federal funding, or losing a competitive edge within my field)	0.79	0.12	0.11
Concerns about discovering data quality issues that will invalidate my prior work and increase the difficulty of future publications	0.77	0.30	-0.03
Concern that I will be expected to publicly report my data quality assessment findings	0.72	0.39	0.00
Possibility of colleagues leaving a collaboration because of discovered or unresolved data quality issues	0.69	0.08	0.14
A belief that the nature of my work does not require the assessment of data quality	0.68	-0.27	0.11
Concern that data quality assessment reporting will create an expectation for reproducibility of my data quality findings	0.68	0.43	0.07
Factor 2: Process Issues			
A lack of clear definitions for good or bad data quality	-0.08	0.82	0.22
Concerns about discovering data quality issues that cannot easily be resolved	0.35	0.63	0.18
Belief that data quality efforts, no matter how comprehensive, fail to solve or prevent all potential analysis roadblocks	0.36	0.54	-0.30
Factor 3: Lack of Resources			
A lack of resources (i.e., not enough funding or time to carry out detailed data quality assessments)	0.03	0.13	0.79
My knowledge, experience, or training limits the types of data quality efforts that can be applied to the data I produce/consume	0.17	0.04	0.74

The three columns of numbers under the Factor Loading heading represent the loading values for each question onto each of the identified factors shown above.

Table A1. Factor Structure Matrix for Individual and Organizational Barriers Items (Cont'd)

ITEM	FACTOR LOADINGS		
	F1	F2	F3
ORGANIZATIONAL BARRIERS			
Factor 1: Environment/Support			
Data providers/data owners are resistant to change	0.77	0.21	—
Data quality assessment is not a high priority for investigators	0.74	0.14	—
There are excess layers of management that interfere with data quality assessment efforts	0.72	0.20	—
The funding agencies compensation is not linked to achieving data quality goals	0.64	0.17	—
Factor 2: Practices			
There are no best practices for effectively measuring data quality	0.01	0.84	—
Quality action plans/requirements/expectations are often vague	0.22	0.78	—
The high costs of implementing data quality assessments outweigh the benefits	0.31	0.59	—
Excluded items			
Data providers/data owners are not trained in problem identification and problem solving skills	0.56	0.47	—
There is frequent turnover of data providers/data owners	0.31	0.46	—

The three columns of numbers under the Factor Loading heading represent the loading values for each question onto each of the identified factors shown above.

Figure A4. The Agreement Between the Individual Barriers Scale and Each of the Individual Factor Subscales

Box Plots - Individual Barriers mean and factor means
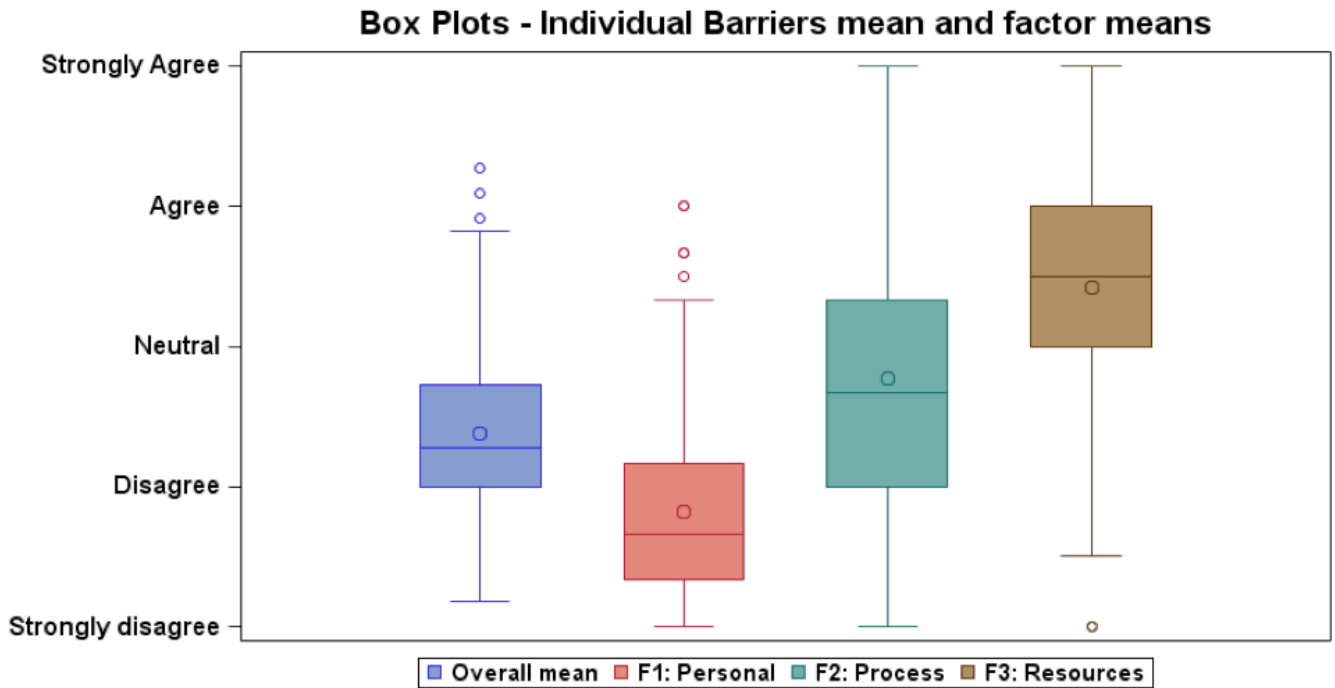
Appendix B. Supplementary Materials

DQ Barriers and Solutions Survey

Please use the following link to download a PDF of the Survey:
doi.org/10.13063/egems.1297.s1