Electronic Health Record Data Quality and Performance Assessments: A Scoping Review

Yordan P. Penev, Timothy R. Buchanan, Matthew M. Ruppert, Michelle Liu, Ramin Shekouhi, Ziyuan Guan, Jeremy Balch, Tezcan Ozrazgat-Baslanti, Benjamin Shickel, Tyler J. Loftus, Azra Bihorac

Submitted to: JMIR Medical Informatics
on: March 06, 2024

Table of Contents

Electronic Health Record Data Quality and Performance Assessments: A Scoping Review

Yordan P. Penev[1, 2] MTM; Timothy R. Buchanan[1, 2] BS; Matthew M. Ruppert[1, 2, 3] MS; Michelle Liu[1] BS; Ramin Shekouhi[1] MD; Ziyuan Guan[1, 2] MS; Jeremy Balch[4] MD; Tezcan Ozrazgat-Baslanti[1, 2] PhD; Benjamin Shickel[1, 2] PhD; Tyler J. Loftus[2, 4] MD; Azra Bihorac[5, 6] MD, MS

[1]Department of Medicine University of Florida Gainesville US
[2]Intelligent Clinical Care Center University of Florida Gainesville US
[3]College of Medicine University of Central Florida Orlando US
[4]Department of Surgery University of Florida Gainesville US
[5]Department of Medicine Division of Nephrology, Hypertension, and Renal Transplantation University of Florida Gainesville US

Corresponding Author:
Azra Bihorac MD, MS
Department of Medicine
Division of Nephrology, Hypertension, and Renal Transplantation
University of Florida
PO Box 100224
Gainesville
US

Abstract

Background: Electronic Health Records (EHRs) have an enormous potential to advance medical research and practice through easily accessible and interpretable EHR-derived databases. Attainability of this potential is limited by issues with data quality and performance assessment.

Objective: This review aims to streamline the current best practices on EHR Data Quality and Performance assessments as a replicable standard for researchers in the field.

Methods: PubMed was systematically searched for original research articles assessing EHR data quality and/or performance from inception until May 7, 2023.

Results: Our search yielded 26 original research articles. Most articles suffered from one or more significant limitations, including incomplete or inconsistent reporting (30%), poor replicability (25%), and lacking generalizability of results (25%). Completeness (81%), Conformance (69%), and Plausibility (62%) were the most cited indicators of Data Quality, while Correctness/Accuracy (54%) was most cited for Data Performance, with context-specific supplementation by Recency (27%), Fairness (23%), Stability (15%), and Shareability (8%) assessments. Artificial Intelligence (AI)-based techniques including natural language data extraction, data imputation, and fairness algorithms were demonstrated to play a rising role in improving both dataset quality and performance.

Conclusions: This review highlights the need for incentivizing data quality and performance assessments and their standardization. The results suggest utility of the adoption of AI-based techniques for enhancing data quality and performance to unlock the full potential of EHRs to improve medical research and practice.

(JMIR Preprints 06/03/2024:58130)
DOI: https://doi.org/10.2196/preprints.58130

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Electronic Health Record Data Quality and Performance Assessments: A Scoping Review

Yordan Penev, MTM[1,4]; Timothy R. Buchanan, BS[1,4]; Matthew M. Ruppert, MS[1,3,4]; Michelle Liu, BS[1]; Ramin Shekouhi, MD[1]; Ziyuan Guan, MS[1,4]; Jeremy Balch, MD[2]; Tezcan Ozrazgat-Baslanti, PhD[1,4]; Benjamin Shickel, PhD[1,4]; Tyler J. Loftus, MD[2,4] Azra Bihorac, MD MS[1,4]
[1]Department of Medicine, University of Florida, Gainesville, FL, USA
[2]Department of Surgery, University of Florida, Gainesville, FL, USA
[3]College of Medicine, University of Central Florida, Orlando, FL, USA
[4]Intelligent Clinical Care Center, University of Florida, Gainesville, FL, USA

Corresponding author: Azra Bihorac MD, MS, Department of Medicine, Division of Nephrology, Hypertension, and Renal Transplantation, PO Box 100224, Gainesville, FL 32610-0224. Telephone: (352) 294-8580; Fax: (352) 392-5465; Email: abihorac@ufl.edu

Word count: 4,705

Key words: Electronic Health Record, EHR, Data Quality, Data Performance, Clinical Informatics

Abstract

Background: Electronic Health Records (EHRs) have an enormous potential to advance medical research and practice through easily accessible and interpretable EHR-derived databases. Attainability of this potential is limited by issues with data quality and performance assessment.

Objective: This review aims to streamline the current best practices on EHR Data Quality and Performance assessments as a replicable standard for researchers in the field.

Methods: PubMed was systematically searched for original research articles assessing EHR data quality and/or performance from inception until May 7, 2023.

Results: Our search yielded 26 original research articles. Most articles suffered from one or more significant limitations, including incomplete or inconsistent reporting (30%), poor replicability (25%), and lacking generalizability of results (25%). Completeness (81%), Conformance (69%), and Plausibility (62%) were the most cited indicators of Data Quality, while Correctness/Accuracy (54%) was most cited for Data Performance, with context-specific supplementation by Recency (27%), Fairness (23%), Stability (15%), and Shareability (8%) assessments. Artificial Intelligence (AI)-based techniques including natural language data extraction, data imputation, and fairness algorithms were demonstrated to play a rising role in improving both dataset quality and performance.

Conclusions: This review highlights the need for incentivizing data quality and performance assessments and their standardization. The results suggest utility of the adoption of AI-based techniques for enhancing data quality and performance to unlock the full potential of EHRs to improve medical research and practice.

Introduction

Adoption of electronic health records (EHR) optimistically promised easily searchable databases and accessible means for prospective and retrospective research applications.[1] The EHR has fallen short of these promises, often due to limited local data and poor data quality (DQ)[2,3] To overcome the first limitation, several institutions have harmonized databases and model ontologies, including PCORnet (The National Patient-Centered Clinical Research Network), All of Us, MIRACUM (Medical Informatics in Research and Care in University Medicine), and the EHDEN Project.[4–7] These programs strive to offer high-quality data for research purposes.[2] However, EHR data quality is often variable. For example, studies have shown completeness in EHR parameter values ranging from 60-100%.[8,9] Similar inconsistencies present a significant limitation to the generalizability and applicability of lessons learned across these datasets for broader research and medical use purposes.

Multiple initiatives have aimed to measure and improve EHR data quality.[10,11] Early efforts in data quality assessment (DQA) demonstrated inconsistent reporting and a need for universal terminology standards in DQA efforts.[11] In response, attempts at a standardized ontology for DQA have been developed, such as through the efforts of International Consortium for Health Outcomes Measurement, 3x3 DQA guidelines, and the terminologies proposed by Kahn et al. and Wang et al.[8,12–15] More recently, artificial intelligence (AI) and natural language processing (NLP) techniques have automated quality initiatives, including data assessment and augmentation.[16,17] Nonetheless, these techniques introduce their own set of quality requirements, including fairness metrics, handling non-tolerable or lost data, and mitigating data drift.[18] Finally, data performance assessment (DPA)–defined as standard metrics related to correctness–is crucial for quality improvement, especially in AI-based processes.[19]

In this review, we critically evaluate peer-reviewed literature on the intersection of DQA standardization, performance of DQA applications, as well as trends in automation techniques.[10–13,20–22] The purpose of this scoping review was to combine the three to formulate a more robust standard for DQA of EHR datasets for medical research and practice.

Methods

This scoping literature review was conducted according to the 2018 Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR), whose checklist is shown in (Supplementary Table S1).[23]

Literature search

A search was performed for all full-text research articles published in English in PubMed from inception to May 7, 2023. A list of the exact search terms is included in (Supplementary Table S2).

Article selection

Four investigators (JB, RS, TB, and YP) reviewed the selected studies during the title and

abstract screening. A further four investigators (ML, RS, TB, and YP) conducted the full-text review and final extraction of articles. Title/abstract screening, full-text review, and final extraction were based on the consensus opinion between two independent reviewers. Conflicts were resolved by a third reviewer. Article management and calculations of Inter-rater reliability and Cohen's kappa were performed using Covidence systematic review software (Covidence, Melbourne, AUS).

Inclusion criteria: Titles and abstracts were screened to include original research articles assessing the data quality and/or performance of all or part of a hospital's EHR system. We looked for studies reporting on aspects of Data Quality (the assessment of EHR data without consideration of follow-up actions) and Data Performance (the assessment of EHR data applications). We expanded on the definitions as defined and cited below (and Table 1):

Data Quality

1. Completeness (or conversely, missingness): the absence of requested data points, without reference to conformance or plausibility as defined below.[12]
2. Conformance: the compliance of data with expected formatting, relational, or absolute definitions[12]
3. Plausibility: the possibility that a value is true given the context of other variable(s) or temporal sequence(s) (i.e., patient date of birth must precede date of treatment or diagnosis)[12]
4. Uniqueness: the lack of duplicated records[8]

Data Performance

- Correctness/Accuracy: whether patient records are free from errors or inconsistencies when the information provided in them is true[10,13]
- Currency/Recency: whether data was entered into the EHR within a clinically relevant timeframe and/or is representative of the patient state at a given time of interest[10,13]
- Fairness (or conversely, bias): the degree to which data collection, augmentation, and application are free from unwarranted over- or underrepresentation of individual data elements or characteristics
- Stability (or conversely, temporal variability): whether temporally dependent variables change according to predefined expectations[10,12]
- Shareability: whether data can be shared directly, easily, and with no information loss[3]
- Robustness: the percent of patient records with tolerable (e.g., inaccurate, inconsistent, outdated information) vs. intolerable (e.g., missing required information) data quality problems.[24]

We additionally included studies reporting on data imputation methods, defined as techniques used to fill in missing values in an EHR, such as through statistical approximation and/or the application of AI.

Exclusion criteria: We excluded tangential analyses of data quality in articles focused primarily on clinical outcomes. As such, studies discussing data cleaning as part of quantifying clinical outcomes were excluded from our analysis. Proposals or study

protocols with no results were also excluded during the screening process.

Article quality assessment

Full text articles were additionally scored as having or missing the following criteria:

- Data integrity: comprehensiveness for each main outcome, including attrition and exclusions from the analysis and reasons for them
- Method clarity: clear description of DQA data sources, analysis steps, and criteria
- Outcome clarity: outcomes reporting in plain language, in their entirety, and without evidence for selective reporting
- Generalizability: applicability of DQ techniques described in the article to other clinical settings

Results

Article characteristics

Flow diagram for article selection is shown in (Figure 1). A total of 154 records were identified using the search terms defined in (Supplementary Table S2) using the PubMed library. After removal of 31 duplicates and the 72 articles identified as irrelevant, 51 studies proceeded to full-text review. Full-text review excluded a further 25 articles owing to reasons listed in (Figure 1), leaving a final total of 26 original research studies.[2–6,8,9,14,19,22,24–39] Cohen's kappa between the different pairs of reviewers ranged from 0.28 to 0.54 during the screening process and from 0.54 to 1.00 during the full-text review.

Study characteristics are shown in (Figure 2) and (Supplementary Table S3). Exactly half of the identified articles targeted general EHR data quality analysis[4–6,19,22,27–34], while the other half focused on a particular specialty or diagnosis (Figure 2a).[2,3,8,9,14,24–26,35–39] The latter included primary care (n=3, 12%)[37–39], cardiovascular disease (n=3, 12%)[8,35,36], anesthesia/ pain medicine (n=2, 8%)[14,26], intensive care units (n=2, 8%)[3,25], and pediatrics[24], oncology[2], and infectious disease (n=1 each, 4%).[9]

Article quality assessment conducted as part of our review process identified 14 (54%) of the articles[2–6,8,9,19,22,24–31,31–38] suffered from at least one common study design/reporting limitation, with 5 of the articles having more than one.[14,24,27,35,38] Among these, 6 (30% of all errors) articles did not clearly state their methods[3,28,29,31,35,38], 5 (25%) had incomplete data[24,27,30,35,38], 5 were not generalizable to other settings[4,24–26,35], and 4 did not clearly state their outcomes (Figure 2b).[27,31,33,36]

Commonly referenced Data Quality and Performance indicators are summarized in (Figure 3). Respective definitions, mitigation strategies, and references are listed in (Table 1) below.

Table 1. Data Quality and Performance indicator definitions, mitigation strategies, and references.

	Definition	**Mitigation Strategies**	**Relevant Studies**
Data Quality			
Completeness (or	The absence of data	Automated data	[2–6,8,9,24–26,28–30,32–39]

conversely, Missingness)	points, without reference to data type or plausibility[12]	extraction; Data Imputation	
Conformance	The compliance of data with expected formatting, relational, or absolute definitions[12]	Preemptively enforced data format standardization	2–6,8,14,24–28,30,32–35,38
Plausibility	The possibility that a value is true given the context of other variable(s) or temporal sequence(s) (i.e., patient date of birth must precede date of treatment or diagnosis)[12]	Periodic realignment with logic rule sets or objective truth standards; Thresholding	4–6,8,14,25,27–29,31–35,37,39
Uniqueness	The lack of duplicate data among other patient records[8]	Two-level encounter/ visit data structure	8
Data Performance			
Correctness/ Accuracy	Whether patient records are free from errors or inconsistencies when the information provided in them is true[10,13]	Periodic validation against internal and/or external gold standards	2,7–9,14,23,24,29,33–3
Currency/Recency	Whether data was entered into the EHR within a clinically relevant timeframe and/ or is representative of the patient state at a given time of interest[10,13]	Enforcing predetermined hard and soft rule sets for timeline of data entry	2,4,9,28,34,36,38
Fairness (or conversely, Bias)	The degree to which data collection, augmentation, and application are free from unwarranted over- or underrepresentation	Periodic review against a predetermined internal gold standard or bias criterion	3,19,22,24,28,37

	of individual data elements or characteristics		
Stability (or conversely, Temporal variability)	Whether temporally dependent variables change according to predefined expectations[10,12]	Periodic measurement of data drift against a baseline standard of data distribution	[4,8,19,33]
Shareability	Whether data can be shared directly, easily, and with no information loss[3]	Preemptively enforced data standardization	[2,3]
Robustness	The percent of patient records with tolerable (e.g., inaccurate, inconsistent, outdated information) vs. intolerable (e.g., missing required information) Data Quality problems.[24]	Timely identification of critical data quality issues	[24]

Data Quality Assessment

Completeness

Completeness was the most cited element of data quality analysis, with references in 21 (81%) of all articles.[2–6,8,9,24–26,28–30,32–39] Importantly, 19 (73%) studies integrated data from multiple clinical sites[2,4–6,9,19,22,24,26,27,31–39], which was associated with issues in data collection and missingness "across organizational structure, regulation, and data sourcing."[33] Clinical domains reported to be prone to low data completeness included patient demographics, with Estiri et al.[30] highlighting the issue for records of patient ethnicity and Thuraisingam et al.[37] for mortality records (e.g., missing year of death), and medication management, with Thuraisingam et al. highlighting the issue for dosage/strength/frequency of prescriptions and Kiogou et al. for missing dates/reasons for discontinuation of medications.[36]

To combat data missingness, Lee et al.[22] used NLP algorithms to automatically extract data from patient records, while a further 5 studies made use of data imputation techniques. Among the latter, two articles generated synthetic data, while another three supplemented datasets through information from external datasets. Fu et al.[3] generated synthetic data by modeling providers' assessments of EHR data based on different information sources according to their individual characteristics (e.g., tendency to ascertain delirium status based on Confusion Assessment Method (CAM) vs. prior ICD coding or nursing flowsheet documentation), while Zhang et al.[19] used a generative adversarial network (GAN) trained on real longitudinal EHR data to create single synthetic EHR episodes (e.g., outpatient or

inpatient visit). Meanwhile, Lee et al.[35] supplemented existing EHR records on heart failure by aggregating data from open-source datasets of heart failure biomarkers (including Database of Genotypes and Phenotypes and the Biologic Specimen and Data Repository Information Coordinating Center) and using literature guidelines to create a standard set of cardiovascular outcome measures, while Curtis et al.[2] supplemented missing EHR mortality records with data from US Social Security Death Index and the National Death Index, and Mang et al.[32] used a manually-generated standalone synthetic dataset to test the development of a new software tool for DQ assessment.

Conformance

Conformance was the second most cited element of DQA with references in 18 (69%) articles.[2–6,8,14,24–28,30,32–35,38] Similarly to completeness, data quality checks on conformance were performed automatically across most studies. Mitigation strategies included enforcing strict formatting rules at the time of data entry; e.g., by using International Statistical Classification of Diseases (ICD) codes to define cause of death or a diagnosis of delirium.[2,3]

Plausibility

Plausibility was the third most cited element of DQA with references in 16 (62%) articles.[4–6,8,14,25,27–29,31–35,37,39] Clinical domains prone to issues with plausibility included patient baseline physical characteristics, medication and laboratory records. Estiri et al.[30] and Wang et al.[31] reported significant rates of plausibility issues for baseline physical characteristics, with higher error rates for records of patient height as compared to weight, likely due to the multiple flowsheet fields for height, including “estimated”, “reported”, and “measured” which are generally averaged or selectively dropped. Pharmacologic data were prone to issues with plausibility due to timeliness (e.g., ART was dispensed before or >30 days after visit date[9]) or discrepancies between diagnosis and drug (e.g., NSAID prescription on date of gastroduodenal ulcer diagnosis[6]). Finally, lab results were also prone to issues with plausibility due to value ranges, units, timing (e.g., lab time was at an invalid time of day or in the future), discrepancies between diagnosis and lab records (e.g., drug was documented as present but there was no lab record) or drug prescriptions and lab records (e.g., metformin was prescribed prior to a documented HbA1c lab or warfarin was prescribed without follow up INR lab).[6] Notably, this may reflect poorly integrated healthcare systems where labs are being drawn at disparate institutions.

A total of 18 (69%) studies utilized logic statements to assess plausibility[2,4–6,8,9,14,24,27–29,33–39], including rules to determine temporal plausibility (e.g., labs drawn at an invalid time of day (e.g. 10:65 AM)[6], extubation occurring prior to intubation[14], or death date occurring before birth date[34]), diagnostic/procedural plausibility (e.g., a procedure marked as outpatient when it is only performed on an inpatient basis[27] or an obstetric diagnosis given for a biologically male patient[6,9,27]), alignment with external standards or expectations (e.g., laboratory result absent for diagnosis/drug[6] or demographic alignment of medication name and dose with expected value ranges[36]) and others. Eleven studies (42%) utilized thresholding to identify data of low or questionable quality,[4,6,8,9,14,19,29,31,34,37,39] including clinical and physiological value ranges (e.g., BMI between 12 and 90 kg/ m^2[37] or FiO2 between 10 and 100%[14]) and logical thresholds (e.g., recorded date of arrival prior to date

of data collection initiation[8] or difference of >730d when comparing age in years and date of birth fields[9]).

Uniqueness

Finally, one study (4%) reported on data Uniqueness. Aerts et al.[8] measured the frequency of patient record duplications (i.e., when patient records were erroneously copied during data merging or reprocessing). To reduce the rate of record duplications, the researchers in the study suggest a 2-level data structure, with more general patient data being recorded at the encounter level (which can include multiple visits during a single clinical episode) and diagnosis/ procedure-specific data at the level of the particular visit.

Data Performance Assessment

Correctness/Accuracy

Correctness/Accuracy was the most cited element in data performance analysis, with references in 14 (54%) of all articles[2,8,9,14,19,25,26,31,34–39] The metric was evaluated via manual review in 8 (57%) out of the 14 articles which reported the measure.[2,8,14,25,26,31,36,38] Five (36%) articles evaluated it in comparison to an external standard, including national registries[2,37], EHR case definitions based on billing codes[38], literature guidelines with high research utility[35], or, in the case of a newly proposed AI technique for synthetic data augmentation, comparison to a previously published GAN model performance.[19] A further 3 (21%) assessed Correctness/Accuracy against an internal standard by calculating the proportion of records satisfying internally predetermined rule sets.[9,34,39] Of note, Curtis et al.[2] and Terry et al.[38] used both manual review and comparison to an external gold standard for validation.

Currency/Recency

Recency was the second most cited data performance element, with references in 7 (27%) articles. [2,4,9,28,34,36,38] Among these, 5 (71%) studies evaluated the metric according to internally predetermined hard rule sets (e.g., whether an obese patient had a weight recording within one year of the previous data point or whether data was entered into the EHR within 3 days of the clinical encounter[9,34,38]) or soft rule sets (e.g., whether the data was entered into the EHR within a subjectively determined clinically actionable time limit[4,36]), while 2 (29%) used external standards including national registries and guidelines.[2,28]

Fairness/Bias

The third most cited data performance element was Fairness or Bias, with references in 6 (23%) articles. [3,19,22,24,28,37] Among these, Lee et al.[22], Thuraisingam et al.[37], Tian et al.[28], and García-de-León-Chocano et al.[24] assessed fairness by manual review, while Fu et al.[3] and Zhang et al.[19] did so through automated review against a predetermined internal gold standard (i.e., distribution of data characteristics within a real EHR dataset) or data bias

criterion (i.e., critic model measuring Jensen–Shannon divergence between real and synthetic data over time), respectively.

Stability

Data stability was the fourth most cited performance element, referenced in 4 (15%) articles. [4,8,19,33] All four articles which measured data stability did so via temporal statistical analyses of data drift according to a predetermined internal baseline standard of data distribution.[8,9,34,39]

Shareability

Shareability was referenced in 2 (8%) articles from our analysis.[2,3] Both studies measured the performance metric by way of manual review in a pre- and post-test analysis of data standardization.[2,3]

Robustness

Finally, García-de-León-Chocano et al.[24] reported on information robustness by way of statistical estimation of critical (e.g., missing or null required values) vs. noncritical (all other) data quality issues which may obstruct subsequent data applications & performance measures.

Interventions for Improving Data Quality and Performance

Three articles included in our analysis reported effective interventions to improve data quality and performance.[4,9,39] In terms of Data Quality, Walker et al.[39] reported an increase in compliance with 155 completeness and plausibility data checks from 53% to 100% across six clinical sites after 3 rounds of DQA. In terms of Data Quality and Performance, Puttkamer et al.[9] reported both higher data completeness and recency following a continuous data reporting & feedback system implementation. Finally, Engel et al.[4] reported increased shareability (concept success rate i.e., whether data partners converted information from their individual EHRs to the shared database) increase from 90% to 98.5% and percentage of sites with over three data quality errors reduction from 67% to 35% across 50+ clinical sites over two years.

Discussion

Principle Contributions and Comparison with Prior Work

This scoping review provides an overview of the most common and successful means of EHR data quality and performance analysis. The review adds to a growing body of literature on the subject, most recently supplemented by a systematic review by Lewis et al.[40] To our knowledge, ours is the first review of specialty-specific applications of data quality alongside performance assessments. We identified and analyzed a total of 26 original research articles recently published on the topic. The results serve to characterize the most

common medical fields making use of such assessments, the methodologies they use for conducting them, and areas for specialty-specific as well as generalizable future improvement. Finally, the discussion proposes a set of six unique and practical recommendations for minimizing modifiable data quality and performance issues arising during data extraction and mapping.

Article characteristics

Our review noted a paucity of data quality assessments within clinical specialties, where expert domain knowledge plays a key role in identifying logic inconsistencies. Half of all identified articles concerned general EHR data assessments, while the other half focused on medical fields such as Primary care, Cardiovascular diseases, or ICU/ Anesthesia, with notable absence of Psychiatry, Emergency Medicine, and any of the surgical specialties. This points to a lack of peer-reviewed research and underutilization of data quality and performance strategies across a wide spectrum of the medical field. There is a wide knowledge gap between how data is entered and acted upon clinically and how it appears in silico. Therefore, more efforts need to be directed towards supporting EHR data assessment initiatives in these specialties, with close collaboration between clinical users and data scientists.

More than half of the articles included in this scoping review suffered from common limitations, including using or reporting incomplete data, methods, and/or outcomes. Among the articles scoring high for incomplete data, the chief issues include data attrition during extraction[24,30] and unclear or missing reporting[27,35,38], pointing to a need for higher information interoperability and reporting standards, such as those put forth by Kahn etal.[12] These standards recommend using a harmonized and inclusive framework for the reporting of DQ assessments, including standardized definitions for Completeness, Conformance, Plausibility and other measures as discussed previously.

Similar issues were observed with methods reporting, with several articles under-reporting steps in their data extraction or analysis, thereby limiting the replicability and generalizability of their findings.[3,28,29,35] Unclear reporting or underreporting was a substantial issue for outcomes as well, with low scoring articles reporting only partial or too high-level results suggesting selective reporting bias.[14,27,33,36] To align with the standards set forth by articles scoring high in reporting quality, we recommend stating all data sourcing, methods, and results according to predetermined definitions of Data Quality or Performance (see above) in enough detail such that they would be easily replicated by researchers at an unrelated institution.

A final article quality pitfall concerned articles which were too specific to a particular health system or clinical context. The chief issues among original research articles which scored “low” in our generalizability assessment concerned their overreliance on internal data quality checks or measures that could only be implemented within their specific institutional EHR.[4,24–26,35] To increase generalizability, we recommend relying on external data quality standards such as societal guidelines, previously published measures, or open-source databases, to the extent possible before resorting to the development of new inhouse tools which impose limitations to generalizability outside the local clinical context.[8,12–15]

Data Quality Assessment

The marked drop between the use of Completeness, Conformance, and Plausibility vs. other indicators (Figure 3a) demonstrates that the field has settled on these measures as the main components of EHR data quality analysis. Taking this into consideration, we recommend measuring all three for a general assessment of clinical data quality. Of note, there is a significant drop-off between 81% of studies reporting on Completeness vs. 69% on Conformance and 62% on Plausibility, which indicates an opportunity for limited but quick data quality "checks" utilizing completeness measures only. More specialized analyses may require further reporting, including Uniqueness in the event of data merger with the possibility of duplicate results. These may be particularly important in the case of EHR data quality assessments following information reconciliation from the merger of multiple data sources including patient demographics or baseline physical characteristics, laboratory or pharmacological data which were shown to be particularly prone to errors in data quality.

Our review additionally demonstrates that issues with data completeness, conformance, and plausibility may be at least partially addressed with data imputation methods. While previously these methods were either too limited in scope (completeness only), crude (e.g., augmenting missing data with the mean of the entire dataset or a value's K-nearest neighbors) or computationally expensive (e.g., individual values calculated via regression models based on predetermined sets of correlated features), our review suggests that these tasks are being increasingly automated. Specifically, data attrition contributing to missingness and/ or conformity at the extraction stage may be minimized with AI data extractor algorithms, such as the one described by Lee et al.[22] In cases where further extraction is no longer feasible, the dataset may be augmented by 1) using Large Language Models (LLMs) for extracting structured data available in other formats (e.g., lab values recorded in the text of media files from outside patient records), 2) incorporating or cross-referencing data from well-established outside data repositories (e.g., US Social Security Death Index for mortality records[2] or Database of Genotypes and Phenotypes and the Biologic Specimen for biomarkers of heart failure and other conditions[35]), or 3) generating synthetic data, for example, by modeling providers' behaviors with respect to different information types or sources[3] and/or by using generative adversarial networks to create synthetic care episodes based on longitudinal EHR observations.[19]

Data Performance Assessment

Correctness/Accuracy was by far the most reported measure among the data performance indicators examined in our review. While certainly integral to assessing a dataset's usability and potential for downstream clinical or research impact, Correctness alone is insufficient to guarantee the success of said applications. A technically "correct" dataset may still be practically limited if it is outdated, biased, inconsistent, or entirely idiosyncratic. We therefore recommend that future data assessments consider including additional measures of Recency, Fairness, Stability, and Shareability, respectively, among their core set of performance indicators as they each contribute a unique measure of a dataset's applicability. The predominance of internal standards comparisons for measuring Recency and Stability in our review demonstrates that these indicators may be essential for

individualized EHR data performance assessments and should therefore be considered on a case-by-case basis (e.g., in epidemiology where the timing and consistency of reporting can be of essential importance, or quality improvement initiatives where a researcher might want to compare pre- vs. post- intervention results). Likewise, Shareability ought to be considered in the case of assessing dataset performance for interoperability purposes (e.g., with data integrations, sharing and reporting).

As discussed previously, Data Fairness assessments can and should be considered for monitoring overall EHR bias as well as the bias inherent to any data imputation methods as discussed above. Our review points to the fact that this is a rapidly developing field, with fairness assessments to date mostly requiring manual review against national guidelines or disease registries, or, in the case of synthetic data, real EHR datasets.[41–43] Nonetheless, such gold standards are not always readily available (e.g., what is the standard distribution of age/ race in the real world?) so tech-savvy researchers have more recently resorted to detecting fairness during the validation of ML models or algorithms instead of the data itself.[41–43] Several research articles from our analysis proposed ways of automating the process. Fu et al.[3] presents a straightforward way of measuring agreement of AI-generated synthetic data against a gold standard dataset. Zhang et al.[19] suggests that while such straightforward analysis may be valuable, it is insufficient to measure true Fairness, and goes on to propose a method of measuring bias via Jansen-Shannon divergence which can be calculated for comparisons of real-world and synthetic data. The latter article also suggests a way of preventing synthetic data drift through condition regularization (i.e., minimizing contrastive loss by regularizing the synthetic dataset against a real dataset distribution) and fuzzying (i.e., adding controlled noise to broaden the dataset distribution before the AI training phase). To our knowledge, this is the most recently proposed technique for Fairness assessment in the field. More research is needed to validate and/or augment the technique. Whether through Jansen-Shannon divergence or alternative methods, we recommend that all future data assessment projects measure and report model performance and fairness for sensitive groups.

Lastly, Garcia-a-de-Leon-Chocano et al.[24] proposes a way of calculating Data Robustness. The calculation draws on comparing tolerable vs. non-tolerable issues with data quality, which may be particularly important prior to utilizing the information. We highly suggest that data quality assessments conduct a Robustness calculation immediately before calculating data performance measures for downstream applications, which will allow for timely intervention in the case of significant issues with data completeness, conformity, or plausibility that merit additional data collection, review, or imputation steps as discussed above.

The above findings and recommendations are summarized in (Table 2) below.

Table 2. Recommendations for future EHR Data Quality and Performance Assessments

Issue	**Recommendation**
Article Characteristics	
Paucity of specialty-focused EHR data	Incentivize more EHR data assessments, particularly in Psychiatry, Emergency Medicine, and Surgical specialties

assessments	
Incomplete reporting	Use standardized frameworks for measuring and reporting data quality and performance assessments
Poor replicability	Describe DQA methods in enough details such that they could be replicated by a research team at a different institution
Limited generalizability	Utilize already available data quality tools and standards before developing proprietary methodologies
Data Quality Assessment	
Inconsistent methodologies	Analyze Completeness, Conformance, and Plausibility at every DQA (Completeness only may be applicable for quick data quality checks)
Data missingness and non-conformity	Utilize available AI-based data extraction algorithms and augment data using external and synthetic datasets
Data Performance Assessment	
Inconsistent methodologies	Augment Correctness/Accuracy measurement with Recency, Fairness, Stability, Shareability performance metrics
EHR data bias	Automate data fairness assessments by measuring agreement against an external gold standard dataset and/or preventing drift via condition fuzzying and regularization
Timeliness of analysis	Calculate dataset Robustness prior to detailed data quality and performance analysis

Further recommendations

Based on the above review and our team's experience with data quality improvement initiatives, we recommend that administrators minimize modifiable data quality and performance issues arising during extraction by:

1. Using devices that directly upload measurements/settings to the EHR (instead of requiring manual data entry)
2. Anchoring the EHR's interface to a predefined data workflow and ontological structure, e.g., encounters start at time of patient check in instead of when a physician first sees the patient and all encounter times are recorded in one location using standard units
3. Periodically validating the plausibility of automatically entered data such that corrections can be made when necessary, e.g., if an electrocardiography lead falls off a patient's chest and needs to be replaced to record accurate measurements. Wherever possible, provide a reference data format for the validation.

To minimize modifiable issues arising during data mapping, we furthermore recommend:

4. Establishing rules for how to treat a) "missing", b) "modified", or c) "overlapping"

data, e.g., whether a) fields with no value should be regarded as data points or artefact, b) data points which have been subsequently modified should be updated or retained, and c) one data source should take precedence over another in case of duplicate records

5. Instituting standards for parent-child encounters, e.g., if a post-operative outpatient clinic visit should be assigned as unique or a child encounter to the parent surgery visit
6. Maintain provenance of outside facility records which can be used to identify potential issues with externally collected data, e.g., when an outside lab measures a patient's test result using a more or less accurate laboratory technique

Limitations

While this scoping review provides valuable insight into the existing literature on EHR Data Quality Analytics, it has several limitations. Foremost, it is important to acknowledge the limited sample size of 154 articles using our original search criteria, and consequently also the limited number of 26 original research articles which were included in our final analysis after full-text review. Among these articles, there was significant heterogeneity in settings and outcomes of interest, which may limit the validity of direct comparisons between the studies as well as the generalizability of our findings. The review was furthermore restricted to articles available in the PubMed library, which may introduce a potential publication bias, as well as to articles available only in English, which may introduce a language bias to our study selection and subsequent analysis. Finally, while the review focused on EHR data quality and performance assessments, it did not include adjacent areas which may have a pronounced impact on clinical data recording and/ or use such as EHR implementation or utilization. Future research should consider broader inclusion criteria and explore additional dimensions of EHR data quality to provide a more comprehensive understanding of this important topic.

Conclusions

The findings of this scoping review highlight the importance of EHR data quality analysis in ensuring the accuracy and reliability of clinical data. Our review identified a need for specialty-specific data assessment initiatives, particularly in the fields of Psychiatry, Emergency Medicine, and Surgery. We additionally identified a need for standardizing data quality reporting to enhance the replicability and generalizability of outcomes in the field. Based on our review of the existing literature, we recommend analyzing Data Quality in terms of completeness, conformance, and plausibility, and Data Performance in terms of correctness as well as use-case specific metrics such as recency, fairness, stability, and shareability. Notably, our review demonstrated several examples of Data Quality improvement with the use of AI-enhanced data extraction and supplementation techniques. Future efforts in augmenting data quality through AI should make use of data fairness assessments to prevent the introduction of synthetic data bias.

Acknowledgements

Yordan Penev: Investigation, Data curation, Writing – original draft, Writing – review and editing. Timothy R. Buchanan: Investigation, Data curation, Writing – original draft. Matthew M. Ruppert: Data curation, Investigation, Writing – review and editing. Michelle Liu: Investigation. Ramin Shekouhi: Investigation. Ziyuan Guan: Investigation, Methodology, Writing – review and editing. Jeremy Balch: Data curation, Methodology, Writing – review and editing, Supervision. Tezcan Ozrazgat-Baslanti: Data curation, Methodology, Supervision. Benjamin Shickel: Data curation, Methodology, Supervision. Tyler J. Loftus: Data curation, Methodology, Supervision. Azra Bihorac: Data curation, Methodology, Supervision.

T.O.B. was supported by OT2 OD032701 by National Institute of Health (NIH), K01 DK120784 and R01 DK121730 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK), R01 GM110240 and R01 GM149657 from the National Institute of General Medical Sciences (NIH/NIGMS), R01 EB029699 from the National Institute of Biomedical Imaging and Bioengineering (NIH/NIBIB), and R01 NS120924 from the National Institute of Neurological Disorders and Stroke (NIH/NINDS), by DRPD-ROSF2023 (00132783) from University of Florida (UF) Research, and by AWD10247 from the University of Florida Clinical and Translational Science Institute, which was supported in part by the NIH National Center for Advancing Translational Sciences under award number UL1TR001427. A.B. was supported by OT2 OD032701 by National Institute of Health (NIH), by R01 GM110240 and R01 GM149657 from the National Institute of General Medical Sciences (NIH/NIGMS), R01 EB029699 from the National Institute of Biomedical Imaging and Bioengineering (NIH/NIBIB), R01 NS120924 from the National Institute of Neurological Disorders and Stroke (NIH/NINDS), and by R01 DK121730 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK). T.J.L. was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01 GM149657. B.S. was supported by OT2 OD032701 by National Institute of Health (NIH), by R01 DK121730 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK) and by R01 GM110240 and R01 GM149657 from the National Institute of General Medical Sciences (NIH/NIGMS). J.B. was supported by T32 GM008721 by National Institute of Health (NIH). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health and other funding sources.

Conflicts of Interest

The authors declare that they have no competing interests.

Abbreviations

AI: Artificial Intelligence
ART: Anti-Retroviral Therapy
BMI: Body Mass Index
CAM: Confusion Assessment Method
DPA: Data Performance Assessment
DQ: Data Quality
DQA: Data Quality Assessment
EHR: Electronic Health Record
FiO2: Fraction of Inspired Oxygen
GAN: Generative Adversarial Network
HbA1c: Hemoglobin A1c
ICD: International Classification of Diseases
INR: International Normalized Ratio
LLM: Large Language Model
NLP: Natural Language Processing
NSAID: Non-Steroidal Anti-Inflammatory Drug

References

1. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" Research Program. *N Engl J Med*. 2019;381(7):668-676. doi:10.1056/NEJMsr1809937

2. Curtis MD, Griffith SD, Tucker M, et al. Development and Validation of a High-Quality Composite Real-World Mortality Endpoint. *Health Serv Res*. 2018;53(6):4460-4476. doi:10.1111/1475-6773.12872

3. Fu S, Wen A, Pagali S, et al. The Implication of Latent Information Quality to the Reproducibility of Secondary Use of Electronic Health Records. In: Otero P, Scott P, Martin SZ, Huesing E, eds. *Studies in Health Technology and Informatics*. IOS Press; 2022. doi:10.3233/SHTI220055

4. Engel N, Wang H, Jiang X, et al. EHR Data Quality Assessment Tools and Issue Reporting Workflows for the "All of Us" Research Program Clinical Data Research Network. *AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci*. 2022;2022:186-195.

5. Kapsner LA, Mang JM, Mate S, et al. Linking a Consortium-Wide Data Quality Assessment Tool with the MIRACUM Metadata Repository. *Appl Clin Inform*. 2021;12(4):826-835. doi:10.1055/s-0041-1733847

6. Mohamed Y, Song X, McMahon TM, et al. Tailoring Rule-Based Data Quality Assessment to the Patient-Centered Outcomes Research Network (PCORnet) Common Data Model (CDM). *AMIA Annu Symp Proc AMIA Symp*. 2022;2022:775-784.

7. European Health Data & Evidence Network. https://www.ehden.eu/

8. Aerts H, Kalra D, Sáez C, et al. Quality of Hospital Electronic Health Record (EHR) Data Based on the International Consortium for Health Outcomes Measurement (ICHOM) in Heart Failure: Pilot Data Quality Assessment Study. *JMIR Med Inform*. 2021;9(8):e27842. doi:10.2196/27842

9. Puttkammer N, Baseman JG, Devine EB, et al. An assessment of data quality in a multi-site electronic medical record system in Haiti. *Int J Med Inf*. 2016;86:104-116. doi:10.1016/j.ijmedinf.2015.11.003

10. Bian J, Lyu T, Loiacono A, et al. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data. *J Am Med Inform Assoc JAMIA*. 2020;27(12):1999-2010. doi:10.1093/jamia/ocaa245

11. 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;67(5):503-527. doi:10.1177/1077558709359007

12. 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 Wash DC*. 2016;4(1):1244. doi:10.13063/2327-9214.1244

13. 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;20(1):144-151. doi:10.1136/amiajnl-2011-000681

14. Wang Z, Penning M, Zozus M. Analysis of Anesthesia Screens for Rule-Based Data Quality Assessment Opportunities. *Stud Health Technol Inform*. 2019;257:473-478.

15. Kelley TA. International Consortium for Health Outcomes Measurement (ICHOM). *Trials*. 2015;16(S3):O4. doi:10.1186/1745-6215-16-S3-O4

16. Amisha null, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Fam Med Prim Care*. 2019;8(7):2328-2331. doi:10.4103/jfmpc.jfmpc_440_19

17. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-38. doi:10.1038/s41591-021-01614-0

18. Gardner A, Smith AL, Steventon A, Coughlan E, Oldfield M. Ethical funding for trustworthy AI: proposals to address the responsibilities of funders to ensure that projects adhere to trustworthy AI practice. *AI Ethics*. 2022;2(2):277-291. doi:10.1007/s43681-021-00069-w

19. Zhang Z, Yan C, Malin BA. Keeping synthetic patients on track: feedback mechanisms to mitigate performance drift in longitudinal health data simulation. *J Am Med Inform Assoc*. 2022;29(11):1890-1898. doi:10.1093/jamia/ocac131

20. Ozonze O, Scott PJ, Hopgood AA. Automating Electronic Health Record Data Quality Assessment. *J Med Syst*. 2023;47(1):23. doi:10.1007/s10916-022-01892-2

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

22. Lee RY, Kross EK, Torrence J, et al. Assessment of Natural Language Processing of Electronic Health Records to Measure Goals-of-Care Discussions as a Clinical Trial Outcome. *JAMA Netw Open*. 2023;6(3):e231204. doi:10.1001/jamanetworkopen.2023.1204

23. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467-473. doi:10.7326/M18-0850

24. García-de-León-Chocano R, Sáez C, Muñoz-Soler V, Oliver-Roig A, García-de-León-González R, García-Gómez JM. Robust estimation of infant feeding indicators by data

quality assessment of longitudinal electronic health records from birth up to 18 months of life. *Comput Methods Programs Biomed*. 2021;207:106147. doi:10.1016/j.cmpb.2021.106147

25. Sirgo G, Esteban F, Gómez J, et al. Validation of the ICU-DaMa tool for automatically extracting variables for minimum dataset and quality indicators: The importance of data quality assessment. *Int J Med Inf*. 2018;112:166-172. doi:10.1016/j.ijmedinf.2018.02.007

26. Toftdahl AKS, Pape-Haugaard LB, Palsson TS, Villumsen M. Collect Once - Use Many Times: The Research Potential of Low Back Pain Patients' Municipal Electronic Healthcare Records. *Stud Health Technol Inform*. 2018;247:211-215.

27. GADDE MA, WANG Z, ZOZUS M, TALBURT JB, GREER ML. Rules Based Data Quality Assessment on Claims Database. *Stud Health Technol Inform*. 2020;272:350-353. doi:10.3233/SHTI200567

28. Tian Q, Han Z, Yu P, An J, Lu X, Duan H. Application of openEHR archetypes to automate data quality rules for electronic health records: a case study. *BMC Med Inform Decis Mak*. 2021;21(1):113. doi:10.1186/s12911-021-01481-2

29. Tian Q, Han Z, An J, Lu X, Duan H. Representing Rules for Clinical Data Quality Assessment Based on OpenEHR Guideline Definition Language. *Stud Health Technol Inform*. 2019;264:1606-1607. doi:10.3233/SHTI190557

30. Estiri H, Stephens KA, Klann JG, Murphy SN. Exploring completeness in clinical data research networks with DQe-c. *J Am Med Inform Assoc JAMIA*. 2018;25(1):17-24. doi:10.1093/jamia/ocx109

31. Wang H, Belitskaya-Levy I, Wu F, et al. A statistical quality assessment method for longitudinal observations in electronic health record data with an application to the VA million veteran program. *BMC Med Inform Decis Mak*. 2021;21(1):289. doi:10.1186/s12911-021-01643-2

32. Mang JM, Seuchter SA, Gulden C, et al. DQAgui: a graphical user interface for the MIRACUM data quality assessment tool. *BMC Med Inform Decis Mak*. 2022;22(1):213. doi:10.1186/s12911-022-01961-z

33. Sengupta S, Bachman D, Laws R, et al. Data Quality Assessment and Multi-Organizational Reporting: Tools to Enhance Network Knowledge. *EGEMS Wash DC*. 2019;7(1):8. doi:10.5334/egems.280

34. Johnson SG, Speedie S, Simon G, Kumar V, Westra BL. Application of An Ontology for Characterizing Data Quality For a Secondary Use of EHR Data. *Appl Clin Inform*. 2016;7(1):69-88. doi:10.4338/ACI-2015-08-RA-0107

35. Lee K, Weiskopf N, Pathak J. A Framework for Data Quality Assessment in Clinical Research Datasets. *AMIA Annu Symp Proc*. 2018;2017:1080-1089.

36. Kiogou SD, Chi CL, Zhang R, Ma S, Adam TJ. Clinical Data Cohort Quality Improvement: The Case of the Medication Data in The University of Minnesota's Clinical Data Repository. *AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci*. 2022;2022:293-302.

37. Thuraisingam S, Chondros P, Dowsey MM, et al. Assessing the suitability of general practice electronic health records for clinical prediction model development: a data quality assessment. *BMC Med Inform Decis Mak*. 2021;21(1):297. doi:10.1186/s12911-021-01669-6

38. Terry AL, Stewart M, Cejic S, et al. A basic model for assessing primary health care electronic medical record data quality. *BMC Med Inform Decis Mak*. 2019;19(1):30. doi:10.1186/s12911-019-0740-0

39. Walker KL, Kirillova O, Gillespie SE, et al. Using the CER Hub to ensure data quality in a multi-institution smoking cessation study. *J Am Med Inform Assoc JAMIA*. 2014;21(6):1129-1135. doi:10.1136/amiajnl-2013-002629

40. Lewis AE, Weiskopf N, Abrams ZB, et al. Electronic health record data quality assessment and tools: a systematic review. *J Am Med Inform Assoc*. 2023;30(10):1730-1740. doi:10.1093/jamia/ocad120

41. IBM. AI Fairness 360 (AIF360). Published online September 20, 2023. Accessed September 21, 2023. https://github.com/Trusted-AI/AIF360

42. LinkedIn. The LinkedIn Fairness Toolkit (LiFT). Published online September 4, 2023. Accessed September 21, 2023. https://github.com/linkedin/LiFT

43. Microsoft. Responsible AI Toolbox. Published online September 21, 2023. Accessed September 21, 2023. https://github.com/microsoft/responsible-ai-toolbox

Supplementary Files

Figures

PRISMA 2020 flow diagram detailing study selection and reasons for exclusion for all articles considered for this scoping review.

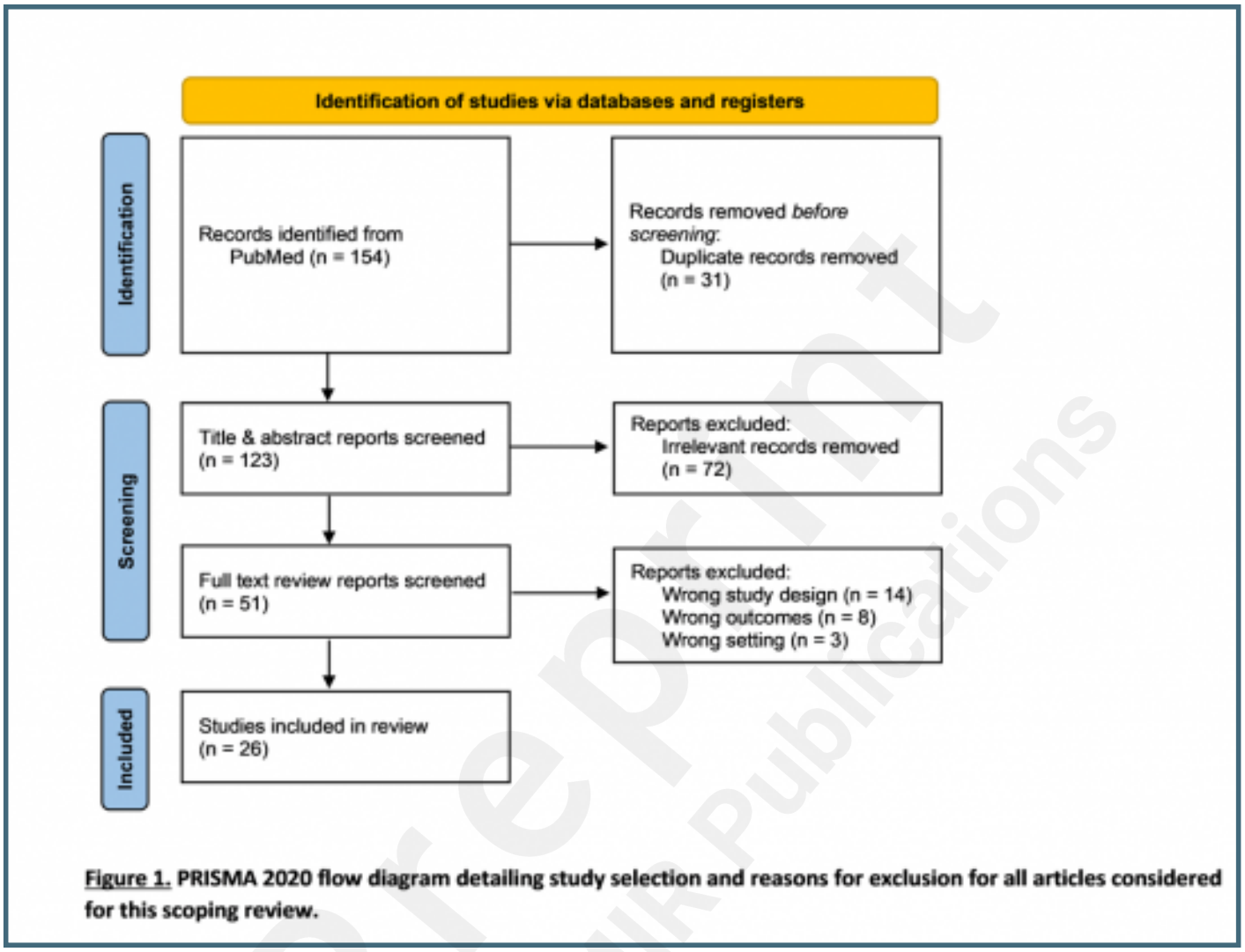

Figure 1. PRISMA 2020 flow diagram detailing study selection and reasons for exclusion for all articles considered for this scoping review.

Frequency of A) clinical specialties among all articles and B) study limitations among all limitations identified by reviewers in this analysis.

Figure 2. Frequency of A) clinical specialties among all articles and B) study limitations among all limitations identified by reviewers in this analysis

Elements of data A) quality and B) performance commonly referenced by articles included in this review.

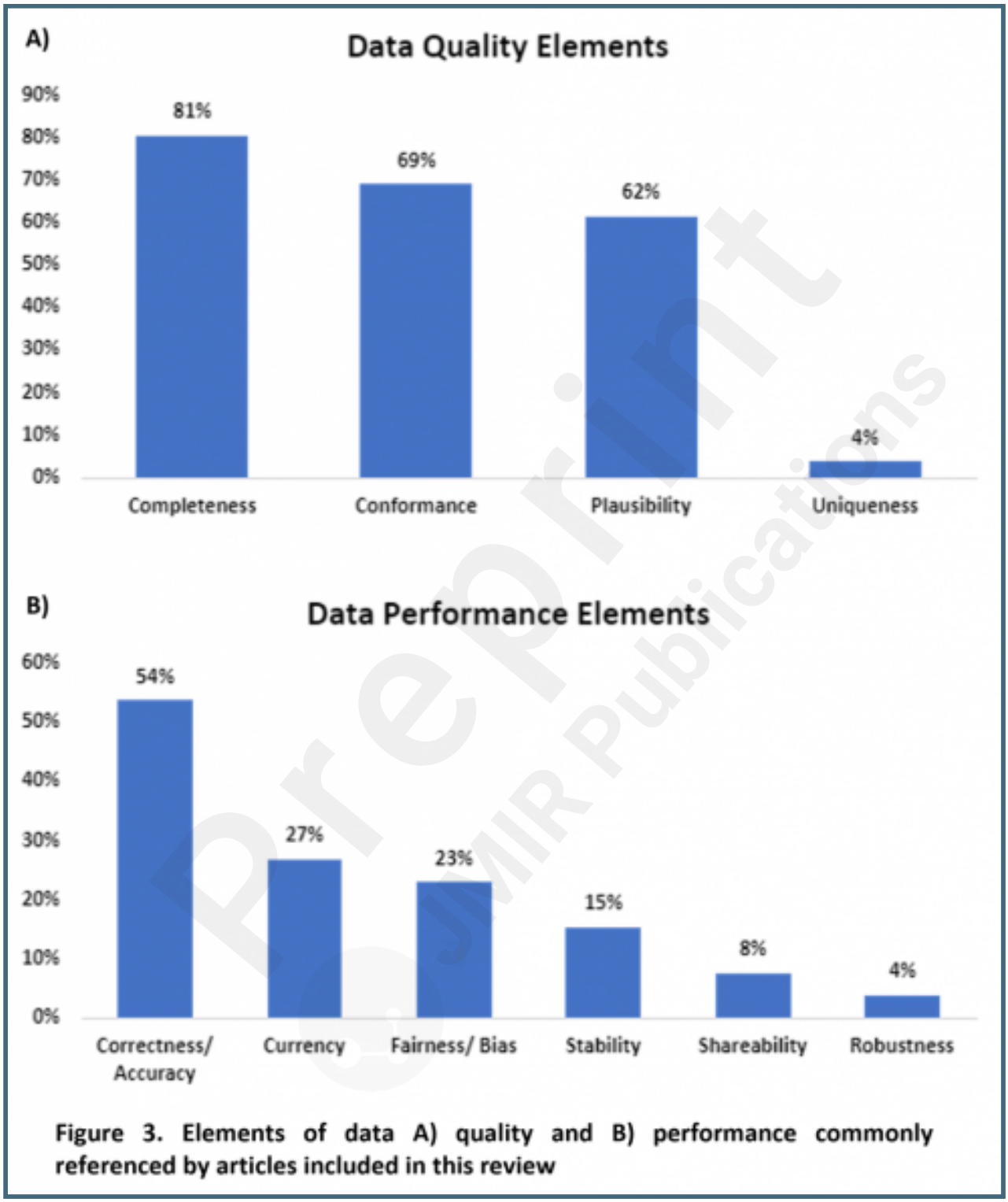

Figure 3. Elements of data A) quality and B) performance commonly referenced by articles included in this review

Multimedia Appendixes

Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.
URL: http://asset.jmir.pub/assets/b9020be62ef837615340b9265270eeec.docx

Search terms.
URL: http://asset.jmir.pub/assets/bdadc9e3967c9b00df009bae16fce995.docx

Study characteristics.
URL: http://asset.jmir.pub/assets/6e31ec14c96ae753032712cf4a8b6090.xlsx