Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250439

Is Research Data Trustworthy? A Quality Comparison Between FHIR, Trinetx and Clinical Data Sources

Sude E. KOCMAN[a], Helene KÖSTER[b], Christian GULDEN[a], Lorenz A. KAPSNER[b,c], Igor ENGEL[b], Hans-Ulrich PROKOSCH[a,b], and Jonathan M. MANG[b,1]
[a] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute for Medical Informatics, Biometrics and Epidemiology, Medical Informatics, Erlangen, Germany*
[b] *Medical Center for Information and Communication Technology, Universitätsklinikum Erlangen, Erlangen, Germany*
[c] *Institute of Radiology, Universitätsklinikum Erlangen, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany*

Abstract. The use of electronic health records for clinical research offers access to large-scale real-world data, but it requires the accurate transformation of data across clinical data repositories. In this study, we evaluate the data quality and completeness in three repositories (DWH, FHIR, and TriNetX) at Erlangen University Hospital. Key data elements (diagnosis, procedure, and laboratory codes) were analyzed, alongside a specific research question. Our results show good overall consistency, but discrepancies arise due to differences in code systems, data filtering, and the mapping process. These findings highlight the importance of critically assessing data provenance and the transformation processes when conducting multicenter research. Understanding the strengths and limitations of each repository is essential for ensuring high-quality research outcomes.

Keywords. Data quality Assessment (DQA), Multicenter Research Collaboration, TriNetX, Clinical Data Repositories (CDRs)

1. Introduction

The use of electronic health records (EHR) for clinical research is rapidly expanding, providing access to large-scale data and valuable real-world evidence [1]. Efforts to use both intra- and inter-institutional patient data have led to the creation of clinical data repositories (CDRs) networked among organizations [2]. In Germany, the Medical Informatics Initiative (MII) has supported these efforts by funding data integration centers (DIC) at university hospitals since 2016 [3]. The goal is to enable multicenter research by integrating DICs into a nationwide network with a central research data portal (FDPG) [4]. Along with network integration, DICs provide various research services, including multiple CDRs to support local and networked research scenarios [5]. At Erlangen University Hospital (UKER) several platforms have been implemented, including a clinical data warehouse (DWH) [6], the MII FHIR server, and a local TriNetX CDR [2]. We are considering expanding TriNetX's use from clinical trials to

[1] Corresponding Author: Jonathan M. MANG; E-mail: jonathan.mang@uk-erlangen.de.

broader real-world data analysis, but before doing so, we aim to evaluate data completeness and quality, as transferring patient data across CDRs risks information loss.

The goal of this study is to assess how accurately data from clinical systems is reflected in three of our research repositories and to understand the impact of data provenance and CDR application scenarios.

2. Methods

The following briefly describes the ETL pipelines applied to transfer the patient data from the primary clinical data sources into the respective CDR.

The Oracle-based *DWH* is UKER's business intelligence application that supports hospital management and strategic planning and is the primary source of truth for all official administrative reporting for billing and quality assurance purposes. Since its initial implementation in 2004, integrating data from the billing system, surgical, cardiac surgery, anesthesia documentation, and laboratory information systems, it has expanded over 20 years to include nearly 50 administrative and clinical systems. Nightly ETL processes incrementally load data into the DWH, which remains integrated into a locally optimized, non-harmonized data model unsuitable for external data sharing projects but ideal for local use.

The *MII FHIR server* serves as the primary research repository for cross-site data sharing projects via the FDPG within the MII [4]. It comprises resources from various origins, reflecting the heterogeneous IT systems employed within the hospital. Data are integrated from both the clinical DWH and directly from primary systems via the clinical communication server.

In its early years, *TriNetX* used existing i2b2 instances as the sole upstream data source [2]. The data was first transferred into the DWH and then transformed into the i2b2 star schema format. The loading of data from i2b2 into the proprietary TriNetX data model, along with the mapping to international terminologies (e.g., SNOMED CT and LOINC), was then managed by the TriNetX semantic mapping team [7].

To compare the different databases, key data elements of interest, such as diagnosis codes (ICD-10-GM), procedure codes (OPS, i.e. the German adaption of ICPM), and laboratory codes (LOINC), have been selected for analysis. For each resource type, the 10 most frequently used codes were selected for comparison [8–10]. In the case of diagnoses, the data was further differentiated by gender. Both the internal DWH and i2b2 utilized the ICD-10-GM for diagnosis codes, while TriNetX employed the international version, ICD-10-CM. Since OPS codes could not be retrieved from TriNetX, equivalent SNOMED codes were used for comparison. To ensure consistency across the CDRs, data from the years 2021 to 2023 was analyzed.

In addition, a research question ("How many patients had received a Lutetium-177 PSMA radiation therapy (OPS-Code: *8-530.d0*)? What time period is covered?") recently formulated by clinical researchers was analyzed to provide insight into the extent to which data discrepancies affect current research findings.

To efficiently count data elements and compare the different databases, we performed SQL queries on each CDR. Furthermore, we used our Data Quality Assessment (DQA) tool [11], which assesses the DQ dimensions 'Value Conformance'

and 'Completeness'. The same queries were then executed directly on the DWH and TriNetX systems. Data elements across FHIR, TriNetX, and the clinical DWH were compared to identify potential data quality issues arising from the transformation processes, emphasizing the importance of understanding data provenance and its impact on quality and research outcomes. Directly querying FHIR resources using the server's REST-based search turned out to be too inefficient. To improve analytical performance, the FHIR resources were additionally stored as Delta Lake tables [12], enabling faster SQL queries via Trino [13].

3. Results

For a detailed insight, the counts for the 10 most commonly used codes for diagnoses, lab results and procedures were analyzed (Figure 1). For the evaluated diagnosis codes, strong consistency was observed across systems for the first eight codes. However, almost no entries for code *Z38.0* ("Single liveborn infant") were found in TriNetX, despite its presence in i2b2.

In TriNetX, some procedure codes showed unusually high counts, and FHIR showed slightly increased values compared to the DWH for diagnoses and procedures.

ICD-10 & gender	DWH	i2b2 TriNetX	TriNetX	FHIR	OPS	DWH	i2b2 TriNetX	TriNetX	FHIR	LOINC	DWH	i2b2 TriNetX	TriNetX	FHIR
I25.1, m	9.160	9.160	9.160	9.222	5-469	3.336	3.336	27.190	3.400	26453-1	153.381	153.328	153.080	153.311
		→ 0%	→ 0%	↑ 0.7%			→ 0%	↑ 715.0%	↑ 1.9%			↓ 0.0%	↓ 0.2%	↓ 0.0%
I50.1, m	4.798	4.799	4.780	4.868	5-758	2.958	2.958	3.120	2.963	718-7	175.474	175.418	172.590	175.390
		↑ 0.0%	↓ 0.4%	↑ 1.5%			→ 0%	↑ 5.5%	↑ 0.2%			↓ 0.0%	↓ 1.6%	↓ 0.0%
I50.1, w	3.046	3.046	3.040	3.088	5-032	644	644	890	646	20570-8	175.440	175.384	172.560	175.339
		→ 0%	↓ 0.2%	↑ 1.4%			→ 0%	↑ 38.2%	↑ 0.3%			↓ 0.0%	↓ 1.6%	↓ 0.1%
I70.2, m	2.397	2.397	2.400	2.477	5-513	438	438	17.660	438	26464-8	153.382	153.328	153.080	153.331
		→ 0%	↑ 0.1%	↑ 3.3%			→ 0%	↑ 3932.0%	→ 0%			↓ 0.0%	↓ 0.2%	↓ 0.0%
K40.9, m	775	775	780	790	5-749	2.067	2.067	2.080	2.070	1988-5	129.560	147.283	133.040	129.525
		→ 0%	↑ 0.6%	↑ 1.9%			→ 0%	↑ 0.6%	↑ 0.1%			↑ 13.7%	↑ 2.7%	↓ 0.0%
M17.1, w	213	213	220	217	5-820	399	399	420	365	2093-3	48.313	52.105	52.070	48.288
		→ 0%	↑ 3.3%	↑ 1.9%			→ 0%	↑ 5.3%	↓ 8.5%			↑ 7.8%	↑ 7.8%	↓ 0.1%
O80, w	4.697	4.697	4.700	4.714	5-794	687	687	2.200	643	2085-9	31.847	33.939	33.920	31.830
		→ 0%	↑ 0.1%	↑ 0.4%			→ 0%	↑ 220.2%	↓ 6.4%			↑ 6.6%	↑ 6.5%	↓ 0.1%
S06.0, m	1.806	1.806	1.810	1.824	5-896	1.119	1.119	31.700	1.095	2089-1	31.853	33.486	33.460	31.832
		→ 0%	↑ 0.2%	↑ 1.0%			→ 0%	↑ 2732.9%	↓ 2.1%			↑ 5.1%	↑ 5.0%	↓ 0.1%
Z38.0, m	3.643	3.643	30	3.654	5-839	425	425	4.870	427	2571-8	55.017	53.635	53.580	51.416
		→ 0%	↓ 99.2%	↑ 0.3%			→ 0%	↑ 1045.9%	↑ 0.5%			↓ 2.5%	↓ 2.6%	↓ 6.5%
Z38.0, w	3.405	3.405	30	3.409	5-452	951	951	960	954	2339-0	2.629	2.629	2.430	2.611
		→ 0%	↓ 99.1%	↑ 0.1%			→ 0%	↑ 0.9%	↑ 0.3%			→ 0%	↓ 7.6%	↓ 0.7%

Figure 1. Patient counts for the most frequently used diagnosis (ICD-10), procedure (OPS), and laboratory (LOINC) codes. Green: Exact match of counts; blue: increasing; purple: decreasing values. Green 0% indicates exact match of counts, while colored 0.0% indicates slight differences.

Furthermore, the analysis of a specific research question — examining the number of patients who had ever received Lutetium-177 PSMA therapy — showed identical patient counts and matching time periods between i2b2, TriNetX and the DWH.

4. Discussion and Conclusions

The absence of *Z38.0* entries in TriNetX is likely due to a privacy filter in TriNetX, which restricts access to information about patients within their first 30 days of life. As

a result, birth codes registered during this period are not displayed on the platform, explaining the discrepancy between TriNetX and i2b2 in this context.

For TriNetX, the inflated counts of certain procedure codes result from the platform's international focus, where SNOMED CT is used instead of OPS codes. The broader mapping between SNOMED CT and OPS during the internal conversion process led to the merging of multiple OPS codes into a single SNOMED CT code, inflating the counts for certain procedures. Mapping German code systems, like OPS, to international ones, such as in TriNetX, is challenging — even between similar systems like ICD-10 GM and CM. In addition, a slight difference between i2b2 and TriNetX can be explained by the fact that TriNetX uses data masking to hide exact counts to protect privacy and meet international compliance requirements. Discrepancies between the DWH and FHIR can be attributed to the FHIR server's use of the communication server, which provides more real-time data based on the HL7 v2 stream, in contrast to the DWH.

In general, the focus was on identifying larger discrepancies, as a perfect match between the analyzed databases is not anticipated.

Overall, there was good accordance across the analyzed data repositories, but some of the differences mentioned emerged due to the complexity of transforming data between systems. Data undergoes multiple transformation steps, which can introduce intentional or unintentional variations. To ensure usable research data, some filtering is required. For example, only laboratory results with valid LOINC assignments are transferred to the harmonized FHIR server, while the raw data for internal evaluations remains stored in internal CDRs. In this analysis, laboratory results without mapped LOINC assignments as well as internal tests, data from test patients, material data without measured values or results without clinical approval were filtered out.

Since direct FHIR search queries were too inefficient, the analysis was performed on FHIR data stored as Delta Tables using Trino as a query engine. It cannot be excluded that additional transferring errors occurred. In addition, the queries were executed sequentially rather than simultaneously, which may have led to minor variations due to ongoing updates or deletions in the data sources.

By analyzing diagnosis, procedure, and laboratory codes, some of the most important resources were covered. Additionally, evaluating a more complex resource, such as medication data, is expected to provide further valuable insights. While our current analysis compares the number of most common diagnoses, procedures, and laboratory values, it does not yet include DQ dimensions such as (computational/relational) conformance and plausibility [14]. Future assessments should include a broader range of DQ dimensions to provide a more comprehensive understanding of the quality and reliability of datasets across CDRs.

In conclusion, all three CDRs evaluated (DWH, FHIR, and TriNetX) are suitable for research, each excelling in different contexts. The DWH is ideal for internal evaluations with identifiable data, FHIR supports standardized international collaboration using pseudonymized data, and TriNetX facilitates global clinical study collaboration with pseudonymized data. However, the discrepancies noted above highlight the importance of closely tracking data sources and transformations. The discrepancy between what researchers perceive about data and the way we should critically assess its quality before analyzing it is essential to be aware of. The quality of data mapping is crucial, and while

it is important not to blindly trust unknown systems, they may still offer valuable, unexpected insights.

Declarations

Funding: This work was partly funded by the German Federal Ministry of Education and Research (BMBF) within the MII (MIRACUM Consortium): Funding Number (FKZ) 01ZZ2301A and within the NUM-DIZ project: 01KX2121. **Authors' contributions**: SK: Main research, review and editing. HK: Validation, writing original draft, review and editing. CG: Technical infrastructure. LAK: Local DQ expert. IE: Local TriNetX expert. HUP: Funding acquisition, project supervision, resources, writing, review and editing. JMM: Conceptualization, project supervision, writing original draft, review and editing. **Competing interests**: None declared.

References

[1] Nordo AH, Levaux HP, Becnel LB, Galvez J, Rao P, Stem K, et al. Use of EHRs data for clinical research: Historical progress and current applications. Learn Health Syst. 2019;3:e10076. doi:10.1002/lrh2.10076.

[2] Palchuk MB, London JW, Perez-Rey D, Drebert ZJ, Winer-Jones JP, Thompson CN, et al. A global federated real-world data and analytics platform for research. JAMIA Open. 2023;6:ooad035. doi:10.1093/jamiaopen/ooad035.

[3] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018;57:e50–e56. doi:10.3414/ME18-03-0003.

[4] Kleinert P, Gebhardt M, Buckow K, Gruendner J, Ganslandt T, Prokosch HU, Semler SC. Germany's approach for the secondary use of federated real-world data in the German Portal for Medical Research Data. Stud Health Technol Inform. 2024;316:1704–1708. doi:10.3233/SHTI240751.

[5] Albashiti F, Thasler R, Wendt T, Bathelt F, Reinecke I, Schreiweis B. Die Datenintegrationszentren – Von der Konzeption in der Medizininformatik-Initiative zur lokalen Umsetzung in einem Netzwerk Universitätsmedizin. Bundesgesundheitsbl. 2024;67:629–636. doi:10.1007/s00103-024-03879-5.

[6] Ganslandt T, Kunzmann U, Diesch K, Pálffy P, Prokosch HU. Semantic challenges in database federation: lessons learned. Stud Health Technol Inform. 2005;116:551–556.

[7] Millan-Fernandez-Montes A, Perez-Rey D, Hernandez-Ibarburu G, Palchuk MB, Mueller C, Claerhout B. Mapping clinical procedures to the ICD-10-PCS: The German operation and procedure classification system use case. J Biomed Inform. 2020;109:103519. doi:10.1016/j.jbi.2020.103519.

[8] DESTATIS. Die 20 häufigsten Operationen insgesamt (OPS 5). 2024. Available from: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Krankenhaeuser/_inhalt.html

[9] DESTATIS. Statistischer Bericht Diagnosedaten. 2024. Available from: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Krankenhaeuser/Publikationen/Downloads-Krankenhaeuser/statistischer-bericht-diagnosedaten-5231301227015.xlsx

[10] Trappe N. Laborbefund lesen: Das sind die 10 wichtigsten Blutwerte. 2024. Available from: https://www.adac.de/gesundheit/gesund-unterwegs/vorsorge/laborwerte/

[11] Mang JM, Seuchter SA, Gulden C, Schild S, Kraska D, Prokosch HU, Kapsner LA. DQAgui: A graphical user interface for the MIRACUM data quality assessment tool. BMC Med Inform Decis Mak. 2022;22:213. doi:10.1186/s12911-022-01961-z.

[12] Grimes J, Szul P, Metke-Jimenez A, Lawley M, Loi K. Pathling: Analytics on FHIR. J Biomed Semantics. 2022;13:23. doi:10.1186/s13326-022-00277-1.

[13] Trino. Distributed SQL query engine for big data. 2024. Available from: https://trino.io/

[14] Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. eGEMs. 2016;4:18. doi:10.13063/2327-9214.1244..