Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230096

Cross-Registry Benchmarking of Data Quality: Lessons Learned

Jürgen STAUSBERG[a,1], Sonja HARKENER[a], Christoph ENGEL[b], Robert FINGER[c], Carsten HEINZ[d], Ekkehart JENETZKY[e,f], Patrick JERSCH[g], David MARTIN[e,h], Rüdiger RUPP[g], Martin SCHOENTHALER[i], Barbara SUWELACK[j], and Jeannine WEGNER[j]
[a] *University Duisburg-Essen, Faculty of Medicine, IMIBE, Essen, Germany*
[b] *Leipzig University, IMISE, Leipzig, Germany*
[c] *Department of Ophthalmology, University Hospital Bonn, Bonn, Germany*
[d] *Department of Ophthalmology, St. Franziskus-Hospital Münster, Münster, Germany*
[e] *Faculty of Health/School of Medicine, Witten/Herdecke University, Witten, Germany*
[f] *University Medical Center of the Johannes-Gutenberg-University, Mainz, Germany*
[h] *Spinal Cord Injury Center, Heidelberg University Hospital, Heidelberg, Germany*
[g] *Department of Pediatrics, Eberhard-Karls University Tübingen, Tübingen, Germany*
[i] *Department of Urology, University of Freiburg, Freiburg, Germany*
[j] *University Hospital Münster, Transplant Nephrology, Münster, Germany*
ORCiD ID: Jürgen Stausberg https://orcid.org/0000-0003-0348-5579

Abstract. Feedback of data quality measures to study sites is an established procedure in the management of registries. Comparisons of data quality between registries as a whole are missing. We implemented a cross-registry benchmarking of data quality within the field of health services research for six projects. Five (2020) and six (2021) quality indicators were selected from a national recommendation. The calculation of the indicators was adjusted to the registries' specific settings. Nineteen (2020) and 29 results (2021) could be included in the yearly quality report. Seventy-four per cent (2020) and 79% (2021) of the results did not include the threshold in their 95%-confidence-limits. The benchmarking revealed several starting points for a weak-point analysis through a comparison of results with a predefined threshold as well as through comparisons among each other. In the future, a cross-registry benchmarking might be part of services provided through a health services research infrastructure.

Keywords. Benchmarking, data quality, quality indicator, registry

1. Introduction

Medical registries provide an understanding about daily health care based upon an observational recording of health-related information. The understanding of daily health care is a prerequisite for advanced tasks as quality research, health economics, or benefit assessment. Having recorded data as its assets, the quality of the data determines the level of knowledge a registry is able to offer. Data should allow, on the one hand, the

---

[1] Corresponding Author: Prof. Dr. med. Jürgen Stausberg, IMIBE, Faculty of Medicine, University Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; E-mail: stausberg@ekmed.de.

answering of research questions defined in a registry protocol and, on the other hand, the execution of statistical procedures defined in an analysis plan. The data itself could be checked concerning inherent characteristics such as data completeness, consistency, or timeliness. Typically, a data management and monitoring facility is responsible for the quality control of the data. Recommendations about the organization and the operation of such a facility are published [1]. One element of quality control could be a feedback to study sites about the quality of the data for which a study site is responsible. Feedback is a well-known intervention of quality management in health care [2]. Current results of performance measures of one study site could be compared with results of other sites as well as with predefined target values. Missing is the possibility to compare the level of data quality of one registry with the level of data quality of another registry. This cross-registry benchmarking would allow learning from the best in the field, not only for study sites within a registry but also for the registry as a whole.

A funding initiative of the German Ministry of Education and Research brought together six medical registries and a supporting accompanying project. The registries cover different medical fields and objectives: ParaReg documents inpatient stays of people with spinal cord injury or disorder, families record fever episodes for the FeverApp-registry, TOFU is interested in treatment exit options for non-infectious, non-anterior uveitis patients with the lowest rate of recurrence, HerediCaRe interconnects health care and health research for women at risk of ovarian and breast cancer, RECUR integrates data of patients suffering from recurrent calculus of the urinary tract from a smartphone app with data from university clinics, living donors of kidney transplants are recruited by SOLKID-GNR (cf. [3] for details about the registries). The six projects started in 2019 with the realization of their registries. In parallel, the accompanying project implemented the cross-registry benchmarking in close cooperation with the registries. In this paper, we want to share our experiences and lessons learned so far.

2. Material and Methods

2.1. Selection of quality indicators

The set of quality indicators used for the cross-registry benchmarking had to be based on an available national recommendation [4]. Two approaches were combined to reduce this set of 51 indicators to a feasible volume. On the one hand, 15 projects in a preceding phase of the funding initiative rated the indicators [5]. On the other hand, an estimation about the relevance of the quality indicators was adopted from the literature [6]. This selection led to an initial set of five indicators (keys in parenthesis): missing values in mandatory data elements (TMF-1014), outliers (TMF-1018), recruitment rate (TMF-1030), drop-out-rate (TMF-1034), completeness (TMF-1046). After the first quality report, two indicators were excluded. TMF-1018 was found to be not helpful due to optimal results for all registries. The application of TMF-1046 was too complex for the moment [7]. Therefore, the subsequent version of the set still included TMF-1014, TMF-1030 and TMF-1034. Three other indicators were added: consistency (TMF-1003), observational units with follow-up (TMF-1042), residual classes for qualitative data elements (REGISVF-1053). Most of the results were rates, having the enumerator as part of the denominator. The calculations of TMF-1034 and TMF-1042 led to ratios with a denominator population of a preceding period. TMF-1030 was a rate in case of registries aiming for case completeness, a ratio in case of registries with fixed target size.

2.2. Organization of the benchmarking

The available definitions of the quality indicators [4] were extended for the concrete use case. In order to achieve fair comparisons, an adjustment to the different settings was added. Following the definition of data quality in ISO 8000 [8], the adjustment could A) vary the applied quality indicators between the registries, B) define different requirements for each registry, or C) use registry-specific thresholds. We decided to apply a common set of quality indicators with only one general threshold for each indicator in the benchmarking. Adjustment was implemented by offering the registries option B, to adapt the calculation of the quality indicators to their conditions. For example, each registry determined the data elements included in the calculation of TMF-1014, the number and the content of integrity rules checked for TMF-1003, or the denominator used in TMF-1030. Threshold values for the indicators were taken from the literature [4] where possible.

Due to data protection reasons, the registries themselves calculated the results based on the specification of the quality indicators and a joint compilation of the registry-specific adaptations. The accompanying project received the results from each registry in a filled template with raw figures and rates or ratios at the beginning of the subsequent year. The accompanying project supplemented the results with 95%-confidence-intervals (95%-CI). A quality report covering all registries with the selected quality indicators was created including tabular and graphical elements. To support the registries in the interpretation of their results, remarks indicated noticeable results that could be the starting point of a weak point analysis. The quality reports were discussed together in workshops and additionally in bilateral meetings.

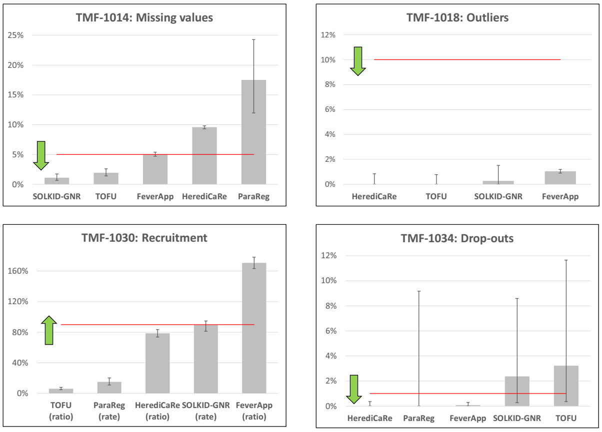

Figure 1. Results of the data year 2020. Arrows indicate the desired direction, the horizontal lines the predefined threshold, and the vertical lines the 95%-CI.

3. Results

Quality reports were created for the data years 2020 and 2021. Due to the Corona pandemic, recruitment of patients started delayed and slower as expected. One registry (RECUR) was not able to recruit any patient in these years. ParaReg did not record quantitative data elements with the possibility of outliers in 2020, in 2021 there was not any case with expected follow-up. Completeness (TMF-1046) was excluded from the quality report (cf. [7] for a discussion of the reasons). We ended up with 19 values for four quality indicators in the report 2020 (cf. figure 1) and 29 results for six quality indicators in the report 2021 (cf. figure 2).

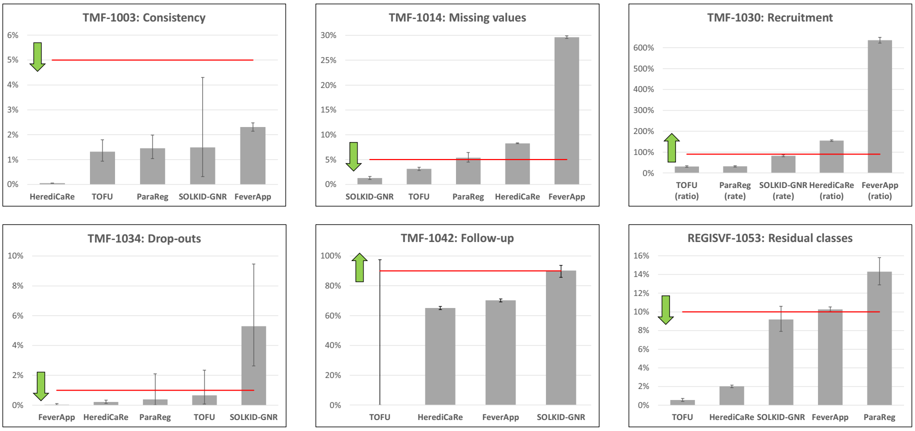

Figure 2. Results of the data year 2021. Arrows indicate the desired direction, the horizontal lines the predefined threshold, and the vertical lines the 95%-CI.

4. Discussion

We expected that the cross-registry benchmarking of data quality could be implemented like a cherry on top of carefully designed case report forms, a systematic and well-founded central monitoring, regularly scheduled reports for the registry office, audits at study sites with at least a partial source data verification, and a benchmarking of study sites using a wide range of quality indicators. However, we recognized that the basis for the cross-registry benchmarking was very heterogeneous. For example, some indicators were specifically implemented for the benchmarking and were not a by-product of an already implemented monitoring. For some projects, the cross-registry benchmarking produced pressure from the cherry on top, leading to extra efforts. Therefore, we assume that not all positive effects of the benchmarking are visible in the results of the quality indicators. The quality reports also highlighted a reverse correlation between data (TMF-1014) and case (TMF-1030) completeness. Both registries, that achieved outstanding results for the recruitment rate in 2021, had the highest rates of missing values.

Quality indicators should provide reliable and valid results. Good quality indicators support responsible parties in quality improvement [9]. In this respect, the quality reports of the cross-registry benchmarking provided a couple of starting points for further actions. All registries received a critical feedback in view of the thresholds differentiating between good and poor data quality. Additionally, the large span of results for most of the indicators allowed a head-to-head comparison of the registries, initiating a learning experience. Further quality-related actions are the rationale of using indicators for quality management to "identify events that merit further review" [9]. That does not mean that every critical result really illustrates a problem. We feel that a set of indicators should be tailored to the current needs. For example, using electronic systems for data capture in all registries avoided outliers through plausibility checks at the time of data recording.

The respective indicator did not provide any new insights and was skipped. The registries' repository of metadata (cf. https://www.toolpool-gesundheitsforschung.de/produkte/metadaten-bmbf-foerdermassnahme-modellhafte-register) disclosed the frequent use of residual classes as "other" for categorical data elements. A respective, newly added quality indicator was found to be helpful and motivated some actions immediately after sampling the data for the quality report.

5. Conclusions

Due to their heterogeneity, a quality indicator that makes sense in one registry may be nonsensical in another. Therefore it is very important to consider the quality indicators in their respective context. Our experience has also shown that the process of discussing quality indicators between those responsible for registry design and analysis can contribute to improving the respective registries. Being aware of the fact, that implementing a cross-registry benchmarking does not replace a thorough evaluation of the method, we are convinced that this approach is worthwhile to be further investigated. There might be a future with services that allow independent registries to compare their data quality with each other beyond the frame of a funding initiative.

Acknowledgements

Funded by the German Federal Ministry of Education and Research under contracts 01GY1901, 01GY1902, 01GY1903, 01GY1904, 01GY1905, 01GY1906, 01GY1917B.

References

[1] Gliklich RE, Leavy MB, Dreyer NA (sr eds). Registries for Evaluating Patient Outcomes: A User's Guide. 4th ed. AHRQ Publication No. 19(20)-EHC020. Rockville, MD: Agency for Healthcare Research and Quality, September 2020.

[2] Barnett GO, Winickoff R, Dorsey JL, Morgan MM, Lurie RS. Quality Assurance through Automated Monitoring and Concurrent Feedback Using a Computer-Based Medical Information System. Med Care 1978;11:962-70.

[3] Stausberg J, Harkener S, Semler S. Recent trends in patient registries for health services research. Methods of Information in Medicine 2021;60(S 01):e1-e8.

[4] Nonnemacher M, Nasseh D, Stausberg J. Datenqualität in der medizinischen Forschung. Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 2014.

[5] Harkener S, Stausberg J, Hagel C, Siddiqui R. Towards a core set of indicators for data quality of registries. Stud Health Technol Inform. 2019;267:39-45.

[6] Schmidt CO, Krabbe CEM, Schössow J, et al. Quality standards for epidemiologic cohort studies: An evaluated catalogue of requirements for the conduct and preparation of cohort studies. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2018; 61: 65-77 (in German).

[7] Stausberg J, Harkener S, Jenetzky E, Jersch P, Martin D, Rupp R, Schönthaler M. FAIR and Quality Assured Data - The Use Case of Trueness. Stud Health Technol Inform 2022;289:25-8.

[8] ISO 8000-2:2020(en) Data quality - Part 2: Vocabulary (s. https://www.iso.org/obp/ui/#iso:std:iso:8000:-2:ed-4:v1:en).

[9] Joint Commission on Accreditation of Healthcare Organisations (JCAHO). Primer on Indicator Development and Application. Measuring Quality in Health Care. JCAHO: Oakbrook Terrace, 1990.