*German Medical Data Sciences: Shaping Change – Creative Solutions for Innovative Medicine*
*R. Röhrig et al. (Eds.)*

*doi:10.3233/SHTI190803*

# Towards a Core Set of Indicators for Data Quality of Registries

Sonja HARKENER [a], Jürgen STAUSBERG [a,1], Christiane HAGEL [b] and Roman SIDDIQUI [b]
[a] *Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, University Duisburg-Essen, Germany*
[b] *Technology, Methods, and Infrastructure for Networked Medical Research, Germany*

**Abstract.** Registries are a widely accepted method in health services research. Registry owners are faced with the challenge to document and assure data quality, vital for answering research questions and conducting quality research. Therefore a survey on indicators for data quality was conducted as part of a German funding initiative. A list of 51 pre-defined quality indicators was provided to 16 patient registry projects in a web based survey. The assessment included three criteria derived from the Rand Appropriateness Method (RAM), the application area, and three criteria representing a project-specific perspective. Considering the criteria adapted from RAM, a core set of 17 indicators could be identified. This core set covered important dimensions, such as case completeness, data completeness and validity. Adding importance as a criterion from a project-specific perspective led to a subset of six indicators. The selection of indicators identified through this survey may be applied on different use cases, e.g. a) benchmarking between registries, b) benchmarking of study sites, and c) value-based remuneration of study sites. Thus, the presented core set of indicators can be used as a basis to improve quality of registry data with a systematic approach.

**Keywords.** Data quality, health services research, quality indicators, registries

## 1. Introduction

Registries are an important method to gain insights into the health status of a population and into health care services in a systematic and supervised approach [1]. The US-American Agency for Healthcare Research and Quality (AHRQ) defined a patient registry as “an organized system that uses observational study methods to collect ... data ... to evaluate specified outcomes for a population” [2]. For health policy, epidemiological registries offer data concerning the prevalence of diseases for a specific population or information about rare conditions. In quality control, registries establish a basis for a performance benchmarking of health care providers. For health services research, quality-controlled registries can close the gap between the results of randomized controlled trials (RCTs) and health care practice [3].

Mostly, registries rely on data that has already been recorded by a health care professional while treating an individual patient [4]. This data is recorded manually a second time, when it is transferred into a registry database. Additionally, some data,

---

[1] Corresponding Author, Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Hufelandstraße 55, D-45147 Essen; Germany, E-mail: stausberg@ekmed.de.

such as quality of life, may be specifically collected for a registry. The registry is then in charge of data collection and data recording. The use of pre-existing data raises concerns about research results originating from patient registries and creates challenges for the registry owners [3, 5]. The concerns relate to the fact that data acquisition was not always been done in a standardized way and not controlled by the registry owners. In comparison with RCTs, data management in registries rarely undertake a source data verification to assure that data in the registry's database are the same as the data stored at the health care provider database. The challenge is not only to select adequate procedures for assuring data quality [6] but also to provide evidence that data contain sufficient information needed to answer the research questions.

Indicators for data quality play a major role in the funding initiative "Establishment of Exemplary Registries for Health Services Research" (REGISVF) published by the German Ministry for Education and Research. Sixteen projects were funded for the development of registry concepts (funding phase one). The projects covered different areas: rare diseases (Prader-Willi syndrome, systemic lupus erythematosus), oncology (hereditary breast and ovarian cancer, metastasizing breast cancer), acute conditions (acute respiratory distress syndrome, fever, heart attack, pulmonary embolism, recurrent calculus of the upper urinary tract), chronic diseases (non-infectious uveitis, non-tropical sprue, paraplegia), interventions (knee joint endoprosthesis, vaccination), and other conditions (death, living organ donors).

In parallel, a supporting project (REGISVF-AP) was launched to promote all registries, and as such to develop a core set of indicators for data quality. An existing set of 51 indicators dedicated to networked medical research was used as a starting point for the identification of a core set [7, 8]. The core set of REGISVF supports three main use cases during the implementation of the initiative (funding phase two): (1) the Ministry demanded a remuneration of study centers based on data quality, (2) benchmarking between the projects might be interesting, for example to initiate a competition between the registries, and (3) the individual projects could use the core set for the management of data quality within their registry. In order to determine the core set, the projects of phase one were invited for a survey.

## 2. Material and Methods

### 2.1. Design of the survey

The survey focused on 51 indicators being proposed in a guideline for adaptive management of data quality [8]. The indicators are divided into three levels of quality introduced for health care [9]: structure (renamed as "integrity"), process ("organization") and outcome ("trueness"). The evaluation of an indicator included its application area and three criteria that were developed with respect to the RAND Appropriateness Method (RAM) [10]: (1) the indicator's understandability, (2) the indicator's relevance, and (3) the indicator's feasibility. An additional project-specific application was rated according to the importance for the project, to the status of implementation and to the planned application in the registry.

The criteria derived from RAM were defined through a Likert scale with four options ("yes", "rather yes", "rather not", "no"). The criterion about the application was divided into "for primary and secondary data", "for primary data only", "for secondary data only" or for "no data at all". The importance of an indicator could be rated with “essential”, “important”, “less important”, and “not important”. The status of the implementation was related to any preliminary project with the options “the indicator was mainly implemented”, “the indicator was partly implemented”, “there was no preliminary project”, and “the indicator was not implemented”. Finally, the planned application was queried by “yes, systematically”, “yes, not systematically”, “undecided”, and “no”. The option “not specified” was always provided as fifth possibility.

The questionnaire included seven questions with five answer options for each indicator. All evaluations were requested at the same time and each project was requested to return only one single project-specific survey response within a period of five weeks. A detailed description of each of the indicators was provided within the survey. The questionnaire was conducted with the online survey tool SurveyMonkey.

### 2.2. Statistics and selection process

Relative frequencies were calculated for the seven criteria. The options “yes” and “rather yes” were merged as positive result for the RAM criteria. For the evaluation of the indicator’s importance a mean was calculated with the following transformation of the options into numerical values: essential - 1, important - 2, less important - 3, and not important - 4 (according to [11]). The resulting means were categorized into three overall assessments: very important (mean < 1.5), important (>= 1.5 and < 2.5), and less important (>= 2.5). An indicator was valuated as understandable if less than 25% of all answers were rated as “rather not” or “not understandable”. An indicator was valuated as relevant if at least 75% of all answers were in favor of it. Finally, a selection required that at least 75% of all answers concerning the feasibility were rated in favor of the respective indicators. The project specific perspectives required a rating of an indicator’s importance as being very important. The survey data were extracted as Microsoft Excel files; SAS was used to apply statistical calculations and Microsoft Excel to generate diagrams.

## 3. Results

### 3.1. General

Fifteen REGISVF-projects participated in the survey. All questions had been answered with a total of 5,355 answers. The option “not specified” was chosen 641 times. Considering the distribution of the answers to the individual criteria, most answers received positive votes. Considering the importance, more negative answers were encountered compared to the first four criteria. However, since some projects had no predecessor or a planned application has not yet been decided, the assessments of the implementation status in a preliminary project and the planned application differed.

### 3.2. General and project-specific perspective

In terms of understandability the description of eight indicators was classified as "rather not" or "not" understandable in at least 25% of the answers (cf. table 1). Twenty-five indicators were assessed as relevant according to the defined threshold including all indicators of the level trueness. Furthermore, 20 indicators were assessed as feasible on the same threshold. It is remarkable that there was only one indicator for the level trueness in this result set. None of the indicators whose description had been rated as not (or rather not) understandable was positively assessed with respect to relevance or feasibility. About half of the evaluations voted for the possibility of application to primary and secondary data, hardly any evaluation voted for application to secondary data only, and few evaluations saw no possibility of application at all.

**Table 1.** Indicators classified as "rather not" or "not" understandable in at least 25% of the answers.

| Title | Level | Percentage |
|---|---|---|
| Endless survivor | integrity | 26.7% |
| Distribution of parameters recorded by the investigator | integrity | 33.3% |
| Values from standards | integrity | 46.7% |
| Illegal values of qualitative data elements used for the coding of missings | integrity | 53.3% |
| Illegal values used for the coding of missing modules | integrity | 26.7% |
| Coverage of metadata from investigations | integrity | 40.0% |
| Synonyms | organization | 33.3% |
| Homonyms | organization | 26.7% |

Six indicators were rated as “very important”, 24 as “important” and 21 as “less important”. The assessment of the implementation status in a preliminary project revealed that in most projects a preliminary project did not exist or the indicators were not implemented. With regard to the intended application of the indicators, the answers "yes, systematically", "undecided" and "no" were distributed almost equally.

### 3.3. Selection of the core set

Applying all RAM conditions, 17 indicators qualified for the core set (cf. table 2). From the level trueness only one indicator qualified (17% from 6). The other five indicators of this level were excluded due to their minor results for feasibility. From the level integrity, nine indicators qualified (30% from 30) for the core set, and seven from the level organization (47% from 15 indicators). All indicators considered as being very important could be identified in the core set of 17 indicators establishing a smaller subgroup.

**Table 2.** Indicators qualified for the core set (logical order). Six indicators of the subset in italic type.

| Title | Level | Relevance | Feasibility | Importance |
|---|---|---|---|---|
| | | (positive votes) | | (mean) |
| Agreement with previous values | integrity | 86.7% | 86.7% | 2.07 |
| Consistency | integrity | 86.7% | 93.3% | 1.57 |
| Distribution of values | integrity | 80.0% | 86.7% | 2.15 |
| Distribution of parameters between sites | integrity | 86.7% | 93.3% | 2.29 |
| Missing modules | integrity | 93.3% | 93.3% | 1.57 |
| *Missing values in data elements* | *integrity* | *93.3%* | *93.3%* | *1.36* |
| *Missing values in mandatory data elements* | *integrity* | *93.3%* | *93.3%* | *1.29* |
| Data elements with value unknown etc. | integrity | 86.7% | 80.0% | 1.92 |
| Outliers (continuous data elements) | integrity | 100.0% | 93.3% | 1.50 |
| Currency | organization | 80.0% | 80.0% | 2.31 |
| *Duplicates* | *organization* | *100.0%* | *93.3%* | *1.33* |
| *Recruitment rate* | *organization* | *93.3%* | *86.7%* | *1.47* |
| Refusal rate of modules | organization | 80.0% | 80.0% | 2.54 |
| Refusal rate of single data elements | organization | 80.0% | 80.0% | 2.38 |
| *Drop-out-rate* | *organization* | *93.3%* | *100.0%* | *1.47* |

| | | | | |
|---|---|---|---|---|
| Observational units with follow-up | organization | 100.0% | 100.0% | 1.73 |
| *Completeness* | *trueness* | *86.7%* | *80.0%* | *1.43* |

Considering the distribution of the answers over the seven criteria for the core set (cf. figure 1), almost all indicators were applicable (at least) on primary data. The importance criterion was rated higher in comparison to the set of all 51 indicators. The implementation status of the core set was evaluated in a similar way to all indicators and the proportion of positive answers regarding the planned application was rated higher.

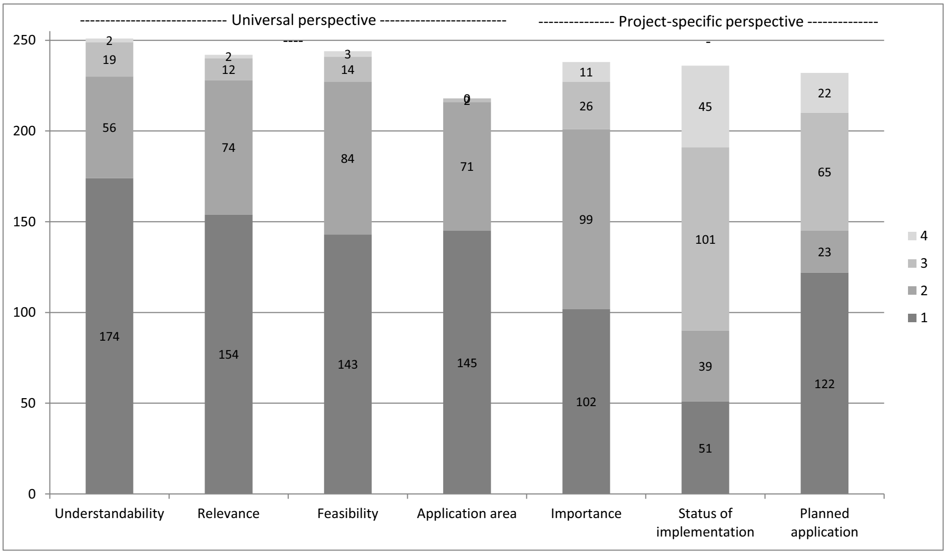

**Figure 1.** Distribution of the votes for the core set. From positive (1) to negative (4). The horizontal axis shows the 7 criteria. The vertical axis shows the absolute numbers of votes. The category "not specified" is not colored.

## 4. Discussion and conclusions

The identified core set covers the main requirements of an "ideal" registry, such as case completeness [12], data completeness [13], and validity [14]. Furthermore, the consideration of follow-up information is crucial in registries that observe the course of a disease or a condition over time [15]. The respective success is covered in the core set with the measure "Observational units with follow-up".

The AHRQ does not offer a formal list of indicators for data quality in its recommendations for patient registries [2]. However, in the context of data tracking, three comparable indicators are mentioned exemplarily: the listed expected to observed rates of patient enrolment (cf. the recruitment rate in the current set), case report form completion (cf. indicators of group missing entries), and the rate of queries. The latter could be seen as a composite calculated as the sum of events addressed by several specific indicators resulting from all levels: integrity, organization, and trueness. Currently, the TMF guideline for data quality indicators [8] offers comparably a weighted sum of all 51 or less indicators as composite. Although the European Crossborder Patient Registries Initiative (PARENT) did not collect any specific indicators [16], data quality was described with six dimensions: accuracy, completeness, interpretability and accessibility, relevance, timeliness, and coherence. Beside the fact that those terms are less standardized and overlapping [17], some further indicators of the core set become relevant. In conclusion, the core set covers many of the measures for data quality mentioned in international recommendations [28].

During the second phase of the funding initiative, the subset with six indicators could be used to adjust the level of remuneration of the study centers to their data quality and to offer a cross-project benchmarking. The six indicators already cover case completeness, data completeness, and validity. The bigger core set with 17 indicators may be applied as part of the quality management procedures of each registry. Thus, the findings reported here - towards a core set of data quality indicators for registries - can stimulate an independent evaluation in the future, addressing registries with a broader range of topics within health services research.

## Acknowledgements

REGISVF-AP was funded by the German Federal Ministry of Education and Research under contracts 01GY1720A and 01GY1720B.

## References

[1] E.A.M. Neugebauer, J. Stausberg, What can and cannot be achieved by registries: Perspective of the registry working group of the German Network of Health Services Research, *Der Unfallchirurg* **119** (2016), 493-500.

[2] R. Gliklich, N. Dreyer, M. Leavy, eds., *Registries for Evaluating Patient Outcomes: A User's Guide,* Third edition. Two volumes, Agency for Healthcare Research and Quality, Rockville, MD, April 2014.

[3] N.A. Dreyer, S. Garner, Registries for Robust Evidence, *JAMA* ***302*** (2009), 790-791.

[4] D. Müller, M. Augustin, N. Banik, W. Baumann, K. Bestehorn, J. Kieschke, R. Lefering, et al., Memorandum registry for health services research, *Das Gesundheitswesen* **72** (2010), 824-839.

[5] J.-P.W.R. Roovers, Registries: what level of evidence do they provide? *Int Urogynecol J Pelvic Floor Dysfunct* **18** (2007), 1119-1120.

[6] D.G. Arts, N.F. de Keizer, G.J. Scheffer, Defining and improving data quality in medical registries: a literature review, case study, and generic framework, *J Am Med Inform Assoc* **9** (2002), 600-611.

[7] M. Nonnemacher, D. Weiland, M. Neuhäuser, J. Stausberg, Adaptive management of data quality in cohort studies and registers: proposal for a guideline, *Acta Inform Med.* **15** (2007), 225-230.

[8] J. Stausberg, R. Pritzkuleit, C.O. Schmidt, T. Schrader, M. Nonnemacher, Indicators of data quality: revision of a guideline for networked medical research, *Stud Health Technol Inform* **180** (2012), 711-715.

[9] A. Donabedian, Evaluating the quality of medical care, *Milbank Memorial Fund Quarterly* **44** (1966), 166-206.

[10] K. Fitch, S.J. Bernstein, M.D. Aguilar, B. Burnand, J.R. LaCalle, P. Lazaro, M. van het Loo, et al., *The RAND/UCLA Appropriateness Method User's Manual*, RAND, Santa Monica, 2001.

[11] C.O. Schmidt, C.E.M. Krabbe, J. Schössow, K. Berger, C. Enzenbach, P. Kamtsiuris, G. Schöne, et al., Quality standards for epidemiologic cohort studies: An evaluated catalogue of requirements for the conduct and preparation of cohort studies, *Bundesgesundheitsblatt* **61** (2018), 65-77.

[12] A.D. McDonald, Mesothelioma registries in identifying asbestos hazards, *Ann N Y Acad Sci* **330** (1979), 441-454.

[13] D.D. Gladman, P. Rahman, G.G. Krueger, P.J. Mease, A.A. Qureshi, J.P. Dutz, U. Lindqvist, S. Carneiro, P.S. Helliwell, I. Ujfalussy, A.B. Gottlieb, F. Behrens, M. Ståhle, Clinical and genetic registries in psoriatic disease, *J Rheumatol* **35** (2008), 1458-1463.

[14] W.R. Hogan, M.M. Wagner, Accuracy of data in computer-based patient records, *J Am Med Inform Assoc* **4** (1997), 342-355.

[15] E.M.M. Brooke & World Health Organization, *The current and future use of registers in health information systems*, World Health Organization, Geneva, 1974.

[16] M. Zaletel, M. Kralj, eds., *Methodological guidelines and recommendations for efficient and rational governance of patient registries*, National Institute of Public Health, Slovenia, Ljubljana, 2015.

[17] J. Stausberg, D. Nasseh, M. Nonnemacher, Measuring Data Quality: A Review of the Literature between 2005 and 2013, *Stud Health Technol Inform* **210** (2015), 712-716.