OPEN ACCESS

Research Article

Assessment of a data quality guideline by representatives of German epidemiologic cohort studies

Bewertung einer Leitlinie zur Datenqualität durch Vertreter epidemiologischer Kohortenstudien in Deutschland

Abstract

High data quality is a precondition for valid scientific conclusions. Indicators should therefore routinely be used to evaluate data quality within the life cycle of health studies. In this project, 15 representatives of seven German population-based cohort studies assessed 51 quality indicators that were proposed in a guideline for networked medical research. The applicability of the indicators to primary data collections was assessed. In addition, their importance was evaluated using a scale ranging from 1 (essential) to 4 (not important). Moreover, their implementation in data quality assessments in the participating studies was evaluated. Comments on potential improvements could be made. Forty-three indicators were rated as applicable. Of these, 29 received a mean importance score of 2 (important) or better, nine received a mean importance score of 1.5 or better. The latter represent a potential core set of data quality indicators for cohort studies. Most indicators that were rated as highly important were used in data quality assessments of the participating studies. Points of criticism regarding the guideline related to its structure and the understandability of some indicators. It was concluded that further improvement of the data quality indicator set will increase its usefulness and applicability in primary data collections. In practice, a small subset of data quality indicators may suffice to capture the most important aspects of data quality in cohort studies.

Keywords: data quality, cohort studies, data quality indicators, data monitoring

Zusammenfassung

Eine hohe Datenqualität ist wesentlich für valide wissenschaftliche Schlussfolgerungen. Indikatoren sollten daher routinemäßig angewendet werden, um die Datenqualität innerhalb des Lebenszyklus von Gesundheitsstudien zu beurteilen. In dem hier beschriebenen Projekt haben 15 Vertreter von sieben bevölkerungsbezogenen Kohortenstudien in Deutschland 51 Qualitätsindikatoren bewertet, die im Rahmen einer deutschen Leitlinie für die vernetzte medizinische Forschung vorgeschlagen wurden. Die Evaluation betraf die Anwendbarkeit der Indikatoren für primäre Datenerhebungen, deren Wichtigkeit auf einer Skala von 1 (essentiell) bis 4 (nicht wichtig) sowie deren Implementation in den teilnehmenden Kohorten. Verbesserungsvorschläge konnten gemacht werden. 43 Indikatoren wurden als anwendbar angesehen. Davon erhielten 29 eine durchschnittliche Wichtigkeit von mindestens 2 (wichtig), neun eine durchschnittliche Wichtigkeit von mindestens 1,5. Die als am wichtigsten bewerteten Indikatoren geben Hinweise auf einen für Kohortenstudien relevanten Kernsatz von Indikatoren zur Erfassung der Datenqualität. Die Mehrzahl der hoch bewerteten Indikatoren wurde in den teilnehmenden Kohorten im Rahmen der Datenqualitätssicherung betrachtet. Als Schwächen der Leitlinie wurden die Verständlichkeit einzelner Indikatoren sowie die Struktur der Leitlinie identifiziert. Das

Carsten Oliver Schmidt[1]
Adrian Richter[1]
Cornelia Enzenbach[2]
Herman Pohlabeln[3]
Christa Meisinger[4]
Jürgen Wellmann[5]
Sonja Selder[6]
Robin Houben[7]
Michael Nonnemacher[8]
Jürgen Stausberg[8]

1 University Medicine Greifswald, Institute for Community Medicine, Greifswald, Germany

2 University of Leipzig, Institute for Medical Informatics, Statistics, and Epidemiology, Leipzig, Germany

3 Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen, Germany

4 Ludwig-Maximilians-Universität München, Chair of Epidemiology at UNIKA-T Augsburg, Munich, Germany

5 University of Münster, Institute of Epidemiology and Social Medicine, Münster, Germany

6 University Munich, Clinics and Policlinics for Radiology, Munich, Germany

7 Robert Koch Institute, Berlin, Germany

8 University of Duisburg-Essen, Faculty of Medicine, IMIBE, Essen, Germany

Konzept von Indikatoren zur Datenqualität sollte weiter verbessert werden, um den Nutzwert und die Anwendbarkeit im Rahmen primärer Datenerhebungen zu erhöhen. In der praktischen Anwendung reicht eine Teilmenge der Indikatoren aus, um wesentliche Aspekte der Datenqualität in Kohortenstudien zu beschreiben.

Schlüsselwörter: Datenqualität, Kohortenstudien, Indikatoren, Datenmonitoring

Introduction

High quality data is a necessary precondition for valid scientific conclusions. While numerous approaches to define data quality have been made, these works are heterogeneous with regards to the scope and type of data quality dimensions, as well as the definition and computation of indicators to describe data quality [1], [2], [3], [4], [5], [6]. There is a considerable difference between demands on data quality assessments in primary data collections compared to assessments in data obtained from electronic health records because only the former allow for a direct control of the data generation process. Data quality frameworks for secondary data sources, such as electronic health records, seem to have received more attention compared to primary data collections in the health sciences [1], [3], [4], [7]. This may be related to the fact that data provenance within secondary data collections is less clear compared to primary data collections. Yet, there is also a high demand to further harmonize data quality assessments in primary data collections, for example in epidemiologic cohort studies. There is neither a standard terminology nor a consensus on the conduct and reporting of data quality assessments. A better understanding of what is perceived as being important related to data quality indicators would therefore be helpful to further elaborate guideline recommendations.

One point of reference for data quality assessments in health studies was provided in Germany more than a decade ago based on work within the TMF (Technology, Methods, and Infrastructure for Networked Medical Research), an umbrella organization for networked medical research. The first edition of a data quality guideline was developed based on a systematic literature review [8]. It defined 24 data quality indicators and was structured into three levels: plausibility/integrity, organization, and trueness. These levels mirrored the approach of Donabedian. He distinguished quality in health care on the levels of structures, processes and outcomes [9]. Medical registries werethe initial focus of the guideline. In 2011, the guideline was revised, leading to a version 2.0 with 51 indicators [10], [11], [12]. It was expanded to cover demands for data quality indicators from different study types such as registries, cohort studies, and centralized research data repositories.

In this work, the applicability, importance and implementation of the TMF data quality indicators was assessed by representatives from established German population-based epidemiologic cohort studies. Based on this assessment, we aimed to provide guidance on indicators considered to be of high relevance and on how to further improve the guideline. Secondary data collections with their specific demands, including the integration of data from secondary sources within the participating cohort studies, were not in focus.

Methods

Representatives of BiDirect (1 principal investigator, 1 senior scientist/statistician (senior refers to at least 10 years of experience with cohort studies), the Heinz Nixdorf Recall Study (1 senior scientist, author of the TMF guideline data quality), KORA (1 senior scientist/head of examination centre), LIFE-ADULT (1 research associate), the MRI group of the NAKO Health Study (3 quality officers), the RKI health monitoring (1 senior scientist, 2 quality officers) as well as SHIP (1 co-principal investigator/senior scientist and project partner in the revision team of the TMF guideline, 2 quality officers) and the senior author of the TMF guideline data quality participated in the project.

The assessment of the guideline was conducted as part of a TMF-funded project on quality standards in cohort studies. To assess the guideline, an Excel spreadsheet to collect information on each indicator of the revised TMF guideline was created by the Greifswald coordination team. This spreadsheet contained all indicator names and descriptions as depicted in the original publication [12]. In a telephone conference in April 2015, a consensus was obtained on two key evaluation aspects. First, the perceived importance of the data quality indicators for primary data collections was assessed using the scale: essential (1), important (2), less important (3), and not important (4). Indicator importance was assessed with regard to the expected demands for primary data collections in general, rather than study-specific demands. Second, the implementation of these indicators for data quality assessments within the participating studies was assessed, differentiating between two degrees of implementation: “mostly implemented”, “partially implemented”. Subsequently, each cohort study received one spreadsheet for completion and returned it to the coordination team by June. Preliminary results were discussed in a one-day workshop in Berlin in June. The workshop targeted a catalogue of requirements for cohort studies [13] in the first part and subsequently the evaluation of the TMF data quality guideline. One result of the workshop was the addition of another item to the spreadsheet regarding the perceived applicability of each indicator in

Table 1: Data quality indicators at the level integrity

Indicator name	ID	Importance	Implemented (not applicable)
Agreement with previous values	TMF-1001	1.8	4/6 (1)
Concordance	TMF-1002	2.3	6/6 (0)
Consistency	TMF-1003	1.4	6/6 (0)
Certain contradiction/error	TMF-1004	1.3	6/6 (0)
Possible contradiction /warning	TMF-1005	2.0	6/6 (0)
Value distributions	TMF-1006	1.4	6/6 (0)
Last digit preferences	TMF-1007	2.2	4/6 (0)
Medical tests on weekends	TMF-1008	2.7	2/6 (3)
Distribution of parameters recorded by the examiner	TMF-1009	1.3	6/6 (0)
Distribution of parameters recorded by the device	TMF-1010	1.8	6/6 (0)
Distribution of findings recorded by a medical reader	TMF-1011	1.7	5/6 (1)
Missing modules	TMF-1012	1.6	5/6 (0)
Missing values in data elements	TMF-1013	1.3	6/6 (0)
Missing values in mandatory data elements	TMF-1014	1.4	6/6 (0)
Missing values in optional data elements	TMF-1015	2.5	5/6 (1)
Data elements with unknown values etc.	TMF-1016	1.8	4/6 (1)
Data elements with entries for all observational units	TMF-1017	2.3	3/5 (0)
Outliers (continuous data elements)	TMF-1018	1.6	6/6 (0)
Values that exceed measurement limits	TMF-1019	1.7	5/6 (1)
Values from standards	TMF-1020	2.5	3/6 (1)
Illegal values (IV) of qualitative data elements	TMF-1021	1.9	5/6 (0)
IV of qualitative data elements used for coding missing	TMF-1022	1.7	5/5 (0)
IV used for coding of missing modules	TMF-1023	2.0	4/6 (0)
IV of qualitative data elements used for the coding of results exceeding measurement limits	TMF-1024	1.7	6/6 (0)
Data elements with unspecific values	TMF-1025	2.6	1/6 (4)
Evidence of known correlations	TMF-1027	2.5	5/6 (1)
Coverage of metadata from investigations	TMF-1050	2.3	3/5 (0)
Distribution of parameters between study sites	TMF-1052	1.2	3/6 (3)

Importance: Mean importance score, based on ratings given by representatives from large established German population-based health cohort studies, reflecting a range between essential (1), important (2), less important (3), and not important (4).
Implemented: Displayed are the number of cohort studies stating that data quality reporting related to the indicator have been implemented out of 6, sometimes 5 due to an additional missing value.
Not applicable: The number of cohort studies stating that the indicator does not apply to the particular study.

primary data collections of cohort studies, because some indicators were generically considered to be unsuitable for primary data collections. Updates to the evaluation sheet were subsequently made until October. Each participating cohort study submitted one completed spreadsheet. These were subsequently integrated by the Greifswald coordination team into one result document.
Results were summarized based on counts and percentages. The results in Table 1, Table 2 and Table 3 have been organized by the levels plausibility/integrity, organization, and trueness of the TMF guideline. Displayed is the mean importance across all studies and the count of studies having implemented an indicator in their data quality reporting. A mean of "1" was the highest possible importance (essential), and a mean of "4" equals no importance. All seven cohort studies provided complete answers to the importance question. A data quality indicator was described as being implemented if it was applied at least in selected study variables. Six cohort studies participated in this assessment. In addition, the number of studies who considered the indicator to be inapplicable to their own context is displayed.

Results

The 51 indicators under evaluation in this study are briefly described in Table 1, Table 2, Table 3 and Table 4 by their internal TMF number ("TMF-ID."). A more detailed characterization of each indicator can be found in the published TMF-guideline [12].

Table 2: Data quality indicators at the level organization

Indicator name	ID	Importance	Implemented (not applicable)
Duplicates	TMF-1029	1.7	5/6 (1)
Recruitment rate	TMF-1030	1.7	6/6 (0)
Refusal rate of examinations	TMF-1031	1.9	5/6 (0)
Refusal rate of modules	TMF-1032	2.0	5/6 (0)
Refusal rate of single data elements	TMF-1033	2.1	5/6 (0)
Drop-out-rate	TMF-1034	1.7	3/6 (3)
Synonyms	TMF-1036	2.3	3/6 (3)
Homonyms	TMF-1037	2.3	3/6 (3)
Observational units with follow-up	TMF-1042	1.3	5/6 (1)

Importance: Mean importance score, based on ratings given by representatives from large established German population-based health cohort studies, reflecting a range between essential (1), important (2), less important (3), and not important (4).
Implemented: Displayed are the number of cohort studies stating that data quality reporting related to the indicator have been implemented out of 6, sometimes 5 due to an additional missing value.
Not applicable: The number of cohort studies stating that the indicator does not apply to the particular study.

Table 3: Data quality indicators at the level trueness

Indicator name	ID	Importance	Implemented (not applicable)
Accuracy	TMF-1043	2.0	2/5 (2)
Agreement with source data referring to data elements	TMF-1044	2.3	1/6 (4)
Agreement with source data referring to observation units	TMF-1045	2.3	1/6 (4)
Completeness	TMF-1046	2.0	3/5 (2)
Compliance with procedural rule	TMF-1047	1.4	6/6 (0)
Representativeness	TMF-1048	1.7	5/6 (0)

Importance: Mean importance score, based on ratings given by representatives from large established German population-based health cohort studies, reflecting a range between essential (1), important (2), less important (3), and not important (4).
Implemented: Displayed are the number of cohort studies stating that data quality reporting related to the indicator have been implemented out of 6, sometimes 5 due to an additional missing value.
Not applicable: The number of cohort studies stating that the indicator does not apply to the particular study.

Table 4: Data quality indicators regarded as not being of importance for primary data collections

Indicator name	ID	Level
Endless survivor	TMF-1035	Integrity
Observational unit with unknown primary tumor	TMF-1026	Integrity
Currency	TMF-1028	Organization
DCO (death certificate only) rate	TMF-1051	Organization
Single notification per observational unit	TMF-1038	Organization
Single notification from pathologist	TMF-1039	Organization
Rejected notifications	TMF-1040	Organization
Data sources per observational unit	TMF-1041	Organization

The exclusion was based on a consensus among the participating cohort study representatives.

Applicability in cohort studies

In total, 43 out of the 51 indicators were perceived as being potentially applicable. Eight indicators (Table 4) were excluded as they were regarded as not being important for primary but for other types of data collections, such as registries. Examples are the proportion of observational units with unknown primary tumour (TMF-1026), data sources per observational unit (TMF-1038), or the Death Certificate Only-rate (TMF-1051), which reflects subjects with death certificate only but no clinical data. The majority of inapplicable indicators belong to the level "organization".

Importance of the indicators

In total, 29 out of the 43 remaining indicators received an average importance score of two or better, nine indicators received average evaluations of 1.5 or better. These nine indicators target different aspects of data quality. First, aspects of missing data are covered, e.g. missing values in data elements (TMF-1013), missing values in mandatory data elements (TMF-1014), and observational units with follow-ups (TMF-1042). The second group of indicators referred to value distributions (TMF-1006), study centre effects (TMF-1052), and examiner effects (TMF-1009). Furthermore, consistency (TMF-1003) was regarded as essential, particularly in terms of certain contradictions/errors in the data (TMF-1004). Finally, compliance with procedural rules (TMF-1047) received a very high importance rating.

Indicators with the lowest importance were those with a very narrow scope, vague implementation conditions, or uncommon application scenarios in cohort studies. An example of an indicator with a narrow scope is TMF-1008: 'Medical tests on weekends' because the realization of examinations in cohort studies depends on the design and may include examinations on weekends. Vague implementation conditions concern the indicator 'possible' contradictions (TMF-1015), or the coverage of metadata (TMF-1050), because metadata may comprise a highly heterogeneous scope of data. Examples of indicators with less common applications in cohort studies are missing values in optional data elements (TMF-1015), and data elements with unspecific values (TMF-1025).

Implementation in the cohort studies

Eight out of the nine indicators with an average evaluation of 1.5 or better were implemented, where applicable, by the participating cohort studies as part of their data quality assessments. For example "drop-out rate" was covered in all studies who have started follow-up assessments. However, in studies with a baseline collection only at the time of the survey there was no application for this indicator.

The rate of implementation was systematically lower for indicators which were considered less important. The main reason for non-implementation of highly important indicators was inapplicability to the respective study setting. This was for example the case for TMF-1052 "Distribution of parameters between study sites" in case of monocentric studies, or for TMF-1034 "Drop-out-rate" in case of lacking follow-up examination waves. In some cases, the wording of the indicator name rendered it unsuitable in some study settings. This was the case for TMF-1008: "Medical tests on weekends" if examinations were regularly conducted at weekends in a study.

Potentials for improvement of the TMF guideline

Several potentials for the improvement of the guideline were discussed. First, the organization of data quality indicators into the three levels plausibility/integrity, organization, and trueness was considered by respondents to be unclear in the context of primary data collections. Second, some indicators should receive better explanations. This concerns for example "coverage of metadata" (TMF-1050). The current guideline presumes an existing metadata concept to compute the related indicator, yet guidance is lacking on the scope or type of metadata. Third, the indicators have a very heterogeneous level of abstraction. The indicator TMF-1008: "Medical tests on weekends" is highly specific and only applicable if no examinations are made on weekends by design. A more general formulation such as "inadmissible examination date" would allow for more flexibility to assess the consistency of recorded examination dates. In contrast, "accuracy" (TMF-1043) is very broad and might require more guidance for an appropriate implementation. Another issue concerned varying demands on data quality indicators depending on the time point at which they are computed. Data monitoring during an ongoing data collection is primarily concerned with corrective actions to improve data collection processes and data curation. In contrast, the final evaluation of the data quality after completion of a study forms the basis for decisions on the feasibility of subsequent scientific analyses [14]. In the former case, evidence of a single technical failure to import a data field commonly triggers corrective data management actions while a single missing data field might be considered of irrelevant impact for statistical analyses in a large data set.

Finally, none of the participating studies directly used the TMF guideline as a basis for their data quality assessments. Rather, their targets for data quality assessment resembled indicators mentioned in the TMF guideline.

Discussion

The assessment of the TMF 2.0 data quality guideline by representatives of major German population-based health cohort studies revealed that 43 out of 51 indicators are regarded as being potentially applicable. In total 29 received a mean importance score of at least 2 (important), nine received a mean importance score of 1.5 or better with a score of 1 indicating essential importance. These highest rated indicators cover missing values, value distributions, and the absence of contradictions. They may form the basis of a useful core set of indicators to assess data quality in cohort studies. In practice, a small subset of data quality indicators may suffice to capture the most important aspects of data quality. Most of them have already been targeted by data quality assessments in the participating studies. There was a consensus on the necessity to further improve the guideline. Issues regarding

understandability may also explain why the guideline has not been directly used as a basis for data quality assessments in the participating studies.

The TMF guideline was selected for this assessment because it was and is the only available German framework to assess data quality with relevance for primary data collections and because it is promoted by one of the most important German network organizations for medical research, the TMF. Its relation to other data quality concepts have been described elsewhere [11], [12]. Another German work on quality standards in epidemiologic cohort studies [13] defined a catalogue of requirements for the preparation and conduct of cohort studies. It contained requirements about the general necessity of a data quality reporting without elaborating on relevant aspects of data quality. This was a key aspect of the current work. Insofar it complements the catalogue of requirements.

Our results can best be conceived as one step towards a better understanding of demands for a harmonized assessment of data quality in epidemiologic cohort studies. Most indicators were considered potentially important for related primary data collections. Yet, this does not necessarily extend to the individual study setting. This emphasizes the need for tailored approaches to assess data quality in line with specific design aspects of cohort studies. Despite of substantial differences in the wording there is considerable overlap to aspects targeted in data quality frameworks which have been designed for electronic health records [4], [5]. Indicators related to data completeness and data correctness form the indispensable core of data quality assessments. Yet, the approaches differ considerably because of the substantially different data structures.

While the guideline targets important data quality aspects, further developments seem indispensable. Beyond a better description of some indicators, a process-oriented approach to data quality assessments seems necessary. For example, it is necessary to distinguish demands related to data monitoring and data curation from requirements regarding the final evaluation of data quality after all data have been collected [14]. Furthermore, extensions are recommended regarding the use of metadata, which should be defined as precisely as possible. This would provide important additional guidance for the calculation of indicators. The classification of some indicators as inapplicable for cohort studies reflected the original design of the guideline, which was oriented toward registries without primary data collections [8] and cannot be considered a weakness. However, for some of these indicators wording could be adapted to improve applicability outside registries. This has been illustrated above in the case of medical assessments on weekends. The definition of indicators should also safeguard a harmonized level of abstraction to improve their usability.

One strength of our assessment was the involvement of representatives from established German population-based health cohort studies. However, the number of involved studies/persons in the evaluation was small. The evaluation of the importance of data quality indicators might have been biased by the degree to which these aspects were already implemented in data quality assessments. The interpretation of the presented results must take into account that the focus was on primary data collections in epidemiologic cohort studies, and relates to the specific background under which these studies have been implemented.

Conclusion

The TMF guideline 2.0 addresses many areas of importance for data quality assessments within epidemiologic cohort studies. The nine highest rated indicators cover missing values, value distributions, as well as the absence of contradictions and may be conceived as a useful core set of indicators to be targeted in data quality assessments. However, improvements should to be made by improving the structure of the guideline, the wording of the indicators, and by adding a stronger focus on the implementation of analyses. Hands-on guidance regarding the setup of metadata would be helpful to increase the usefulness and understandability of the guideline in the context of cohort studies.

Notes

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the TMF grant V114-01M and by the German Research Foundation (DFG, SCHM 2744/3-1).

References

1. Stausberg J, Nasseh D, Nonnemacher M. Measuring Data Quality: A Review of the Literature between 2005 and 2013. Stud Health Technol Inform. 2015;210:712-6.
2. Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002 Nov-Dec;9(6):600-11. DOI: 10.1197/jamia.m1087
3. Lee K, Weiskopf N, Pathak J. A Framework for Data Quality Assessment in Clinical Research Datasets. AMIA Annu Symp Proc. 2017;2017:1080-9.
4. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A Data Quality Assessment Guideline for Electronic Health Record Data Reuse. EGEMS (Wash DC). 2017 Sep;5(1):14. DOI: 10.5334/egems.218

5. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, Estiri H, Goerg C, Holve E, Johnson SG, Liaw ST, Hamilton-Lopez M, Meeker D, Ong TC, Ryan P, Shang N, Weiskopf NG, Weng C, Zozus MN, Schilling L. 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

6. Chen H, Yu P, Hailey D, Wang N. Methods for assessing the quality of data in public health information systems: a critical review. Stud Health Technol Inform. 2014;204:13-8.

7. Kahn MG, Brown JS, Chun AT, Davidson BN, Meeker D, Ryan PB, Schilling LM, Weiskopf NG, Williams AE, Zozus MN. Transparent reporting of data quality in distributed data networks. EGEMS (Wash DC). 2015;3(1):1052. DOI: 10.13063/2327-9214.1052

8. Nonnemacher M, Weiland D, Stausberg J. Datenqualität in der medizinischen Forschung: Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. Berlin: Medizinisch Wissenschaftliche Verlagsgeselschaft; 2007. (Schriftenreihe der TMF).

9. Donabedian A. The quality of care: How can it be assessed? JAMA. 1988;260(12):1743-8. DOI: 10.1001/jama.1988.03410120089033

10. Stausberg J, Pritzkuleit R, Schmidt CO, Schrader T, Nonnemacher M. Indicators of data quality: revision of a guideline for networked medical research. Stud Health Technol Inform. 2012;180:711-5.

11. Stausberg J, Bauer U, Nasseh D, Pritzkuleit R, Schmidt CO, Schrader T, Nonnemacher M. Indicators of data quality: review and requirements from the perspective of networked medical research. GMS Med Inform Biom Epidemiol. 2019;15(1):Doc05. DOI: 10.3205/mibe000199

12. Nonnemacher M, Nasseh D, Stausberg J. Datenqualität in der medizinischen Forschung: Leitlinie zum Adaptiven Datenmanagement in Kohortenstudien und Registern. Berlin: Medizinisch Wissenschaftliche Verlagsgeselschaft; 2014. (Schriftenreihe der TMF).

13. Schmidt CO, Krabbe CEM, Schössow J, Berger K, Enzenbach C, Kamtsiuris P, Schöne G, Houben R, Meisinger C, Bamberg F, Hendel T, Selder S, Nonnemacher M, Moebus S, Stausberg J. Qualitätsstandards für epidemiologische Kohortenstudien: Ein bewerteter Anforderungskatalog zur Studienvorbereitung und Studiendurchführung [Quality standards for epidemiologic cohort studies: An evaluated catalogue of requirements for the conduct and preparation of cohort studies]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2018 Jan;61(1):65-77. DOI: 10.1007/s00103-017-2658-y

14. Huebner M, Le Cessie S, Schmidt CO, Vach W. A contemporary conceptual framework for initial data analysis. Obs Stud. 2018;4:171-92.

Corresponding author:
Prof. Dr. Dr. Carsten Oliver Schmidt
University Medicine Greifswald, Institute for Community Medicine, SHIP-KEF, Walther Rathenau Str. 48, 17475 Greifswald, Germany, Phone: +49 3834 867713
carsten.schmidt@uni-greifswald.de

Please cite as
Schmidt CO, Richter A, Enzenbach C, Pohlabeln H, Meisinger C, Wellmann J, Selder S, Houben R, Nonnemacher M, Stausberg J. Assessment of a data quality guideline by representatives of German epidemiologic cohort studies. GMS Med Inform Biom Epidemiol. 2019;15(1):Doc09.
DOI: 10.3205/mibe000203, URN: urn:nbn:de:0183-mibe0002033

This article is freely available from
https://www.egms.de/en/journals/mibe/2019-15/mibe000203.shtml

Published: *2019-12-06*