doi:10.3233/978-1-61499-852-5-566

Opportunities and Pitfalls in the Definition of Data Validity

Jürgen STAUSBERG[1], Nils KUKLIK, Karl-Heinz JÖCKEL
University of Duisburg-Essen, Faculty of Medicine, Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), Germany

Abstract. Several dimensions of data quality are described in the literature. One overriding aspect is considered to be the extent to which data represent the truth which is captured by data validity. Unfortunately, a common terminology, well defined concepts, and approved measures are missing in regard to data validity. In particular, there is a need to discuss the gold standard as reference for the data at hand and respective measures. Ultimate gold standard would be the state of the patient which itself is subjected to human and personal interpretations. Usually, an often diverse form of source data is used as gold standard. Based on the concept of the measure, it might be inappropriate differentiating between present and absent while calculating precision and recall. Due to the complexity and uncertainty of many health care related issues, a more sophisticated comparison might be necessary in order to establish relevant and general figures of data quality. Unfortunately, a harmonization in this field is not visible. Further research is needed to establish validated standards to measure data quality.

Keywords. Accuracy, correctness, data quality, validation

1. Introduction

Validity is the most important dimension of data quality. Usually, other dimensions of data quality like timeliness or plausibility are conditions that have to be fulfilled in order to obtain valid data. However, their fulfillment does not necessarily guarantee validity. Validity has to be assessed on its own. Terms and concepts that are related to validity are numerous [1, 2]. For example, Weiskopf and Weng assigned accuracy, corrections made, errors, misleading, positive predictive value, quality, and validity to the dimension correctness [3]. Thereby, they set correctness as the preferred denomination for a data element that is true, whereas, Botsis et al. used accuracy in defining inaccuracy as "non-specific, non-standards-based, inexact, incorrect, or imprecise information" [4]. Wang and Strong assigned accuracy together with believability, objectivity, and reputation to the high level dimension "intrinsic data quality" [5]. The data validation process within the regulation of clinical trials is more pragmatic by demanding trial data being "accurate, complete, and verifiable from source documents" [6], thereby reducing trueness as mentioned by Weiskopf and Weng to the congruence with data recorded elsewhere ("source documents"). The German guideline for the adaptive management of data

[1] Corresponding Author, Institute for Medical Informatics, Biometry and Epidemiology, Faculty of Medicine, University Duisburg-Essen, Hufelandstrasse 55, 45122 Essen, Germany; E-mail: stausberg@ekmed.de.

quality differentiates six indicators in the category trueness: accuracy, agreement with source data referring to data elements, agreement with source data referring to observational units, completeness, compliance with operating procedures, and representativeness [7].

In the context of a project on identification of adverse drug events with routine data, questions arose about the gold standard for data validity and its measurement [8]. In the following we will elaborate opportunities and pitfalls in data validation from the point of view of that project.

2. Gold standard

The truth is the patient's health status. However, using a second opinion concept for data validation is an uncommon practice. Even endpoint committees in clinical trials rely on recorded information while reviewing important events. Some studies evaluated a more basic concept of data validation. For instance, Pringle et al. in their validation study compared the documentation of physicians' practices with a gold standard established by an expert opinion based on videotaped material from the consultation process [9]. In a study on pressure ulcers a random sample of inpatients was examined by study nurses independently from the regular staff. During the visits, the study nurses examined the skin of the patients and recorded the pressure ulcer status [10]. The result was then compared against routine data used for reimbursement on the one hand and an additional documentation setup for quality management purposes on the other hand.

Hogan and Wagner described the information flow from the true state of the patient to the paper-based or electronic documentation (cf. figure 1) [11]. Paper-based documentation is being progressively replaced by electronic one, existing without a paper-based artifact, hence the term "source data" could be assigned to both. Consequently, a source data verification must deal with both media. However, if initial paper-based documentation has to be subsequently entered into the computerized patient record (CPR), the transfer process itself is vulnerable to mistakes, thus making the validity dependent on the choice of source.

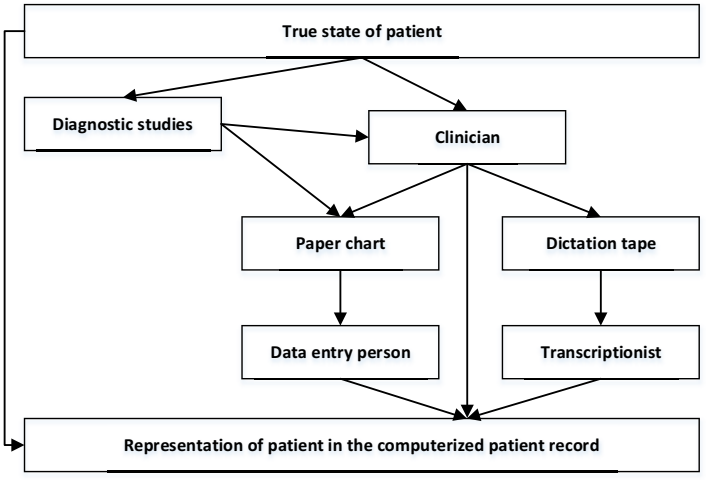

Figure 1. Relation between true state of the patient, paper-based and electronic documentation [11].

Several issues may come up by accepting source documents as reference for the true state of the patient. Firstly, although a source data verification in samples is not inferior to a complete source data verification in the benchmarking process [12], measures from samples might be prone to errors due to relationships between data validity and sampling strategies. For example, cross sectional samples of inpatients will preferably include inpatients with a longer hospital stay. Those patients will not only suffer from a higher risk of adverse events, but also their documentation pattern might differ from patients with a shorter length of stay. Secondly, identification of the truth in source documents is a challenge on its own. For example, in our study about validation of adverse drug events in routine data, the identification process starts with the decision about the parts of the patient's chart that should be consulted, e.g. lab reports and consultations. Specific tools could be implemented guiding through the review process, themselves focusing on selected documents and triggers (for example [13]). The level of evidence present in the chart has to be justified. Is it acceptable if the adverse drug event is explicitly named without any respective supporting findings? Is a recording of the disease term alone equivalent to the recording of the disease accompanied by its relationship to the application of a drug? Consequently, Stausberg et al. extended the figure of Hogan and Wagner and introduced new aspects of the documentation process (cf. figure 2) [14]. The figure shows that different truths exist even in the source data of a single institution (I to III).

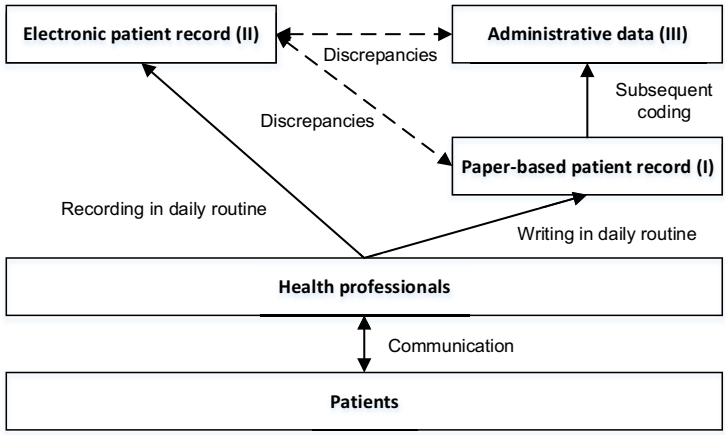

Figure 2. Different levels of truth established during the documentation process [14].

3. Measures

The measures of data validity are derived from the 2 by 2 table, the horizontal and vertical axes represent the gold standard and data under evaluation, respectively (cf. table 1). According to Hogan and Wagner, validity can be assessed by A/(A+C) and A/(A+B) [11]. The rate of recorded elements confirmed by the gold standard - A/(A+B) - is denoted as precision, correctness, or positive predictive value. The rate of true elements recorded in the data - A/(A+C) - is denoted as recall, completeness, or sensitivity. However, as learned from our project and also noted by Logan et al. these definitions are incomplete [15] and for the data under evaluation, the category "present" should be split

up into "present correct" and "present incorrect" (cf. table 2) with a different calculation of the measures for precision and recall. Hence, in the light of the importance of identifying adverse drug events recorded in routine data with a two-step approach covering the disease first and the relationship to a drug prescription second, it might be worthwhile to extend the concept introduced by Logan et al. for data validation studies.

Table 1. 2 by 2 table used for calculating measures of data validity [11].

		Gold standard		
		Present	**Absent**	
Data under evaluation	**Present**	A	B	A/(A+B)
	Absent	C	D	
		A/(A+C)		

Table 2. Extended contingency table used for calculating measures of data validity [15].

		Gold standard		
		Present	**Absent**	
Data under evaluation	**Present correct**	A1	B1	
	Present incorrect	A2	B2	A1/(A1+A2+B2)
	Absent	C	D	
		(A1+A2)/(A1+A2+C)		

4. Conclusions

To assess data validity, various gold standards and measures are available. Furthermore, same terms are applied differently. Therefore, it is quite difficult to compare the level of data validity between different data sources. One option might be the further standardization of the reporting of empirical analyses (cf. [16] for a respective quote in the field of routine data). Another option could be offering a methodology which provides the best knowledge about data validity and data quality [17]. However, as consent about such a methodology could not be expected in the foreseeable future, many more scientific efforts are needed to fully understand the complex issue of data validity.

Acknowledgements

The work was supported by the Federal Ministry of Education and Research of Germany (Bundesministerium für Bildung und Forschung, BMBF, grant number 01GY1328, www.bmbf.de).

References

[1] M. Nonnemacher, D. Weiland, M. Neuhäuser, J. Stausberg, Adaptive management of data quality in cohort studies and registers: proposal for a guideline, *Acta Informatica Medica* **15** (2007), 225-230.
[2] 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.
[3] N.G. Weiskopf, C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, *J Am Med Inform Assoc* **20** (2013), 144-151.
[4] T. Botsis, G. Hartvigsen, F. Chen, C. Weng, Secondary use of EHR: data quality issues and informatics opportunities, *AMIA Jt Summits Transl Sci Proc 2010* (2010), 1-5.

[5] R.Y. Wang, D.M. Strong, Beyond accuracy: what data quality means to data consumers, *Journal of Management Information Systems* **12** (1996), 5-33.
[6] International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH), *ICH Harmonized Guideline. Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2)*, Current Step 4 version dated 9 November 2016.
[7] M. Nonnemacher, D. Nasseh, J. Stausberg, *Datenqualität in der medizinischen Forschung. Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern, 2., aktualisierte und erweiterte Auflage*, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, 2014.
[8] N. Kuklik, J. Stausberg, K.H. Jöckel, Adverse drug events in German hospital routine data: A validation of International Classification of Diseases, 10th revision (ICD-10) diagnostic codes, *PLoS One* **12** (2017), e0187510.
[9] M. Pringle, P. Ward, C. Chilvers, Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer, *Br J Gen Pract* **45** (1995), 537-41.
[10] J. Stausberg, G. Bartoszek, B. Lottko, K. Kröger, W. Niebel, H. Schneider, I. Maier, Problem focused Integration of Information, Quality and Process Management with Empirical Research: the Example of the Essen Interdisciplinary Pressure Ulcer Project. In: H.A. Park, P. Murray, C. Delaney, eds., *Consumer-Centered Computer-Supported Care for Healthy People. Proceedings of NI2006*, IOS, Amsterdam, 2006, 609-612.
[11] W.R. Hogan, M.M. Wagner, Accuracy of data in computer-based patient records, *J Am Med Inform Assoc* **4** (1997), 342–355.
[12] O. Brosteanu, G. Schwarz, P. Houben, U. Paulus, A. Strenge-Hesse, U. Zettelmeyer, A. Schneider, D. Hasenclever, Risk-adapted monitoring is not inferior to extensive on-site monitoring: Results of the ADAMON cluster-randomised study, *Clin Trials* (2017), doi: 10.1177/1740774517724165 [Epub ahead of print].
[13] L. Carnevali, B. Krug, F. Amant, D. Van Pee, V. Gerard, X. de Bethune, A. Spinewine, Performance of the Adverse Drug Event Trigger Tool and the Global Trigger Tool for Identifying Adverse Drug Events: Experience in a Belgian Hospital, *Annals of Pharmacotherapy* **47** (2013), 1414-1419
[14] J. Stausberg, D. Koch, J. Ingenerf, M. Betzler, Comparing paper-based with electronic patient records: lessons learned during a study on diagnosis and procedure codes, *Journal of the American Medical Informatics Association* **10** (2003), 470-477.
[15] J.R. Logan, P.N. Gormann, B. Middleton, Measuring the quality of medical records: a method for comparing completeness and correctness of clinical encounter data. In: S. Bakken, ed., *A Medical Odyssey: Visions of the Future and Lessons from the Past. Proc AMIA 2001 Annu Symp* (2001) 408-412.
[16] H. Quan, T. Williamson, Guiding the reporting of studies that use routinely collected health data, *CMAJ* **188** (2016), 559-60.
[17] M.G. Kahn, T.J. Callahan, J. Barnard, A.E. Bauck, J. Brown, B.N. Davidson, H. Estiri, C. Goerg, E. Holve, S.G. Johnson, S.-T. Liaw, M. Hamilton-Lopez, D. Meeker, T.C. Ong, P. Ryan, N. Shang, N.G. Weiskopf, C. Weng, M.N. Zozus, L. Schilling, A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* **4** (2016), Iss. 1, Article 18.