

15.08.2019

Gemeinsame Stellungnahme der Deutschen Gesellschaft für Epidemiologie (DGEpi) und der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) zum Umgang mit Messwerten in der epidemiologischen und klinischen Forschung

Eine valide und reliable Messung und Klassifizierung von Beobachtungen ist grundlegend für wissenschaftliches Arbeiten und verlässlichen wissenschaftlichen Erkenntnisgewinn. Eine Beobachtung ist objektiv in dem Sinne, dass sie mit Hilfe von transparenten, beschreibbaren Routinen entsteht (Chalmers, 2009) und reproduzierbar ist:

„...observations suitable for constituting a basis for scientific knowledge are both objective and fallible. They are objective insofar as they can be publicly tested by straightforward procedures, and they are fallible insofar as they may be undermined by new kinds of tests made possible by advances in science and technology.“ (Seite 25)

Die Leitlinien für Gute Epidemiologische Praxis (Hoffmann et al., 2018) setzen diese Objektivität von Beobachtungen als Grundlage der Definition von Variablen wie folgt voraus:

„Alle interessierenden Variablen sollen präzise definiert und möglichst standardisiert operationalisiert werden. Für die Bestimmung sind möglichst valide und reliable Mess- und Erhebungsinstrumente einzusetzen ... Die Validität und Reliabilität der eingesetzten Instrumente sollte differenziert ... beschrieben bzw. geprüft werden.“

Es gibt eine Vielzahl von Methoden, wie in epidemiologischen und klinischen Studien Probandinnen(Patientinnen)informationen erhoben werden können. Unter anderem entstehen Messwerte durch die direkte Ableitung physiologischer Parameter und durch den Vergleich dieser Parameter mit standardisierten Skalen („Messung“, z.B. des Blutdrucks mit Hilfe eines geeigneten Apparats), oder auch durch Befragung von Personen und die Einordnung in bestimmte Klassen aufgrund der Antwort („Klassifizierung“, z.B. der gesundheitsbezogenen Lebensqualität anhand einer Gruppe standardisierter Fragen). Die Körpergröße kann zum Beispiel durch Befragung der Person („wie groß sind Sie?“), aber auch durch direkten Vergleich der Scheitelhöhe einer stehenden Person mit einer vertikal angeordneten metrischen Messlatte ermittelt werden. Die Wortwahl der Frage kann hierbei einen Einfluss auf die Ergebnisse nehmen. Ebenso kann das Messprotokoll zur Bestimmung der Körpergröße, Einfluss auf die Ergebnisse nehmen, in dem die Körpergröße z.B. mit oder ohne Schuhe gemessen wird.

Die präzise Definition der Mess- oder Klassifizierungsvorschrift ist daher notwendiger Teil des Studienprotokolls. Ebenso ist die möglichst genaue Beschreibung des tatsächlichen Mess- oder Klassifizierungsvorgangs notwendiger Teil der Berichterstattung über die Studie (üblicherweise im Methodenteil).

Damit die Validität und Reliabilität des Verfahrens und damit der Ergebnisse beurteilt werden kann, muss beschrieben werden, wie die Messwerte zustande gekommen und mit welchen Messfehlern sie behaftet sind, sofern es sich nicht um sehr etablierte, allseits bekannte Verfahren handelt. Maßnahmen der Validierung, sofern solche unternommen wurden, sind ebenso zu beschreiben. So sollte zum Beispiel eine klinische Beurteilung durch eine Untersucherin durch eine zweite Untersucherin bestätigbar sein. Auch sollte beschrieben werden, ob, wie und warum von einem festgelegten Messprotokoll abgewichen wurde. Ebenso ist die genaue Beschreibung der Präzision und Begleitumstände der Messung, der Schulung der Untersucherinnen, und der Entscheidungen, die im Lauf der Messung getroffen werden, aufzuzeichnen und zusammen mit den Ergebnissen zumindest in dem Abschlussbericht eines Projektes und gegebenenfalls in der peer-reviewed Publikation zu berichten.

In vielen Situationen ist es nicht möglich, von allen Probandinnen einer Studie Messwerte zu erhalten. Zum Beispiel kann die Beantwortung einer bestimmten Frage vergessen oder verweigert werden, eine Messung kann technisch nicht machbar oder fehlerhaft sein. Da viele statistische Verfahren Vollständigkeit der Daten verlangen, kann der Ausschluss fehlender Werte zu einem Ausschluss von Beobachtungseinheiten führen, was die Ergebnisse in Abhängigkeit des Mechanismus der Entstehung der fehlenden Werte verzerren kann.

Es kann daher in manchen Situationen wünschenswert sein, fehlende Daten durch einen statistischen Wert bzw. durch eine statistische Imputationsprozedur („Multiple Imputation“) zu ersetzen (Harel & Zhou, 2007, Harel et al., 2018). Die verwendete Methode zur Ersetzung von fehlenden Daten muss jedoch reproduzierbar berichtet werden, damit beurteilbar ist, ob das Ersetzungsverfahren valide ist.

Der Umgang mit fehlenden Werten (missing values) hängt vom Charakter des Mechanismus ab, der zum Fehlwert führt. Es gibt prinzipiell drei Aspekte, die beachtet werden müssen: (1) Das Auftreten eines Fehlwertes ist völlig unabhängig vom Studiengeschehen; (2) Der Fehlwert ergibt sich aufgrund eines Prozesses zu dem Informationen im Rahmen der Studie gesammelt werden können; (3) Der Fehlwert wird durch einen Mechanismus erzeugt, der mit dem Studienverlauf zusammenhängt, über den aber keine Informationen vorliegen. Mit Situation (1) und (2) kann gut umgegangen werden (etwa durch „Multiple Imputation“). Die Situation stellt den Analytisten vor besondere methodische Herausforderungen. Um hierzu eine klare Bewertung der wissenschaftlichen Validität von Analyse und Resultaten bewerten zu können, muss die verwendete Methode zur Ersetzung von fehlenden Daten reproduzierbar berichtet werden.

Zusammenfassend ist bei der Generierung und Behandlung von Messwerten stets auf Transparenz und Nachvollziehbarkeit zu achten, damit beurteilt werden kann, wie in der jeweiligen Studie mit den Daten umgegangen wurde und welche Konsequenzen dieser methodischen Entscheidungen auf die Ergebnisse zu erwarten sind.

gez.
Prof. Dr. med. Dietrich Rothenbacher, MPH
für den Vorstand der Deutschen Gesellschaft für Epidemiologie (DGEpi) e. V.

gez.
Prof. Dr. PH Eva Grill, MPH

gez.
Prof. Dr. Andreas Stang, MPH
Präsident bzw. 1. Vizepräsident der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) e. V.

gez.
Prof. Dr. Alfred Winter

A.F. Chalmers „What is this thing called science?“ Maidenhead: Open University Press, 2009
W. Hoffmann et al. „Leitlinien für Gute Epidemiologische Praxis“ 2018,
<https://www.dgepi.de/assets/Leitlinien-und-Empfehlungen/6074a4e7b8/Leitlinien-fuer-Gute-Epidemiologische-Praxis.pdf>, download 25.07.2019

Harel O, Zhou XH. Multiple imputation: review of theory, implementation and software. Stat Med 2007;26:3057-3077

Harel O, Mitchell EM, Perkins NJ, Cole SR, Thcetgen Tchetgen EJ, Sun B, Schisterman EF. Multiple imputation for incomplete data in epidemiologic studies. Am J Epidemiol 2018;187:576-584