

I Einleitung

Der Nutzen krankheitsbezogener wissenschaftlicher Kohortenstudien und Register, wie sie von Kompetenznetzen in der Medizin, Deutschen Zentren der Gesundheitsforschung und zahlreichen anderen medizinischen Forschungsnetzwerken aufgebaut wurden und werden, hängt ganz wesentlich von der Qualität der dort erfassten Daten ab [Gliklich/Dreyer 2010, Müller et al. 2010]. Datenqualität wird vom Deutschen Institut für Normung in der DIN EN ISO 14050:2010-08 definiert als „Eigenschaften von Daten in Bezug auf ihre Eignung, festgelegte Anforderungen zu erfüllen“ [DIN EN ISO 14050 2010]. Zur effizienten Sicherstellung einer hohen Datenqualität sind wissenschaftlich abgesicherte Methoden und Verfahren vonnöten. Die vorliegende Leitlinie enthält daher Empfehlungen zum Management der Datenqualität von Registern, Kohortenstudien und Data Repositories. Sie wendet sich an Verantwortliche für das Datenmanagement und die Qualitätssicherung in Studien- oder Registerzentralen. Hierbei sollte es sich um Personen mit wissenschaftlichem Studium und relevanten Vorerfahrungen handeln. Bei der Anwendung der hier beschriebenen statistischen Verfahren zur Fallzahlplanung (s. Kap. II.5) kann die Hinzuziehung eines Statistikers oder Biometrikers sinnvoll sein.

Die Erstellung von Version 1.0 dieser Leitlinie vom 31.3.2006 wurde durch die TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. im Rahmen des Projektes „Entwicklung, Umsetzung und Evaluation von Leitlinien zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. Teil 1: Entwicklung und Umsetzung der Leitlinien“ gefördert. Im Jahre 2011 führten dann zwei Entwicklungen zu einer erneuten Auseinandersetzung mit dieser Version. Zum einen wurde in zwei Veröffentlichungen über eine erfolgreiche Anwendung der Leitlinie berichtet [Jacke et al. 2010, Prokein et al. 2010]. Zum anderen wurden – auch von TMF-Mitgliedsverbänden – komplementäre Ansätze zur Operationalisierung von Datenqualität vorgestellt, so dass die Aktualität der Leitlinie in Version 1.0 in Frage stand. Ein Workshop der TMF im April 2011 in Berlin bestätigte dann den

Bedarf einer Überarbeitung¹. Dort wurde in den Themenbereichen Methoden, Datenqualität in der Forschung und Datenqualität in der Versorgung eine Bestandsaufnahme vorgenommen. In dieser Situation förderte die TMF in den Jahren 2011 und 2012 ein Projekt, bei dem Überlegungen aus epidemiologischen Krebsregistern, einer Kohortenstudie sowie einem Data Repository bezüglich der Messung von Datenqualität mit der Version 1.0 der Leitlinie abgeglichen wurden². Als Ergebnis wurde deren konzeptioneller Ansatz bestätigt, das Set von Indikatoren zur Messung der Datenqualität jedoch deutlich erweitert [Stausberg et al. 2012].

Die Einbindung neuer Indikatoren stellt daher die wesentliche Änderung in Version 2.0 der Leitlinie dar. Weitere Teile der Leitlinie wurden entsprechend angepasst; die Literatursichtung wurde bis zum Jahr 2013 aktualisiert. Eine Erweiterung der Leitlinie findet sich mit Anwendungsempfehlungen zu Registern, Kohorten und Data Repositories. Die Erweiterung ergibt sich aus der Verbreiterung des Indikatorensets, bei dem in noch höherem Maß als in Version 1.0 ein Zuschnitt auf die Gegebenheiten des jeweiligen Vorhabens sinnvoll bzw. notwendig ist. Die Anwendungsempfehlungen sind dabei als beispielhafte Muster für jeweils einen Typus von empirischen Forschungsvorhaben zu verstehen. In Gliederung und Darstellung orientiert sich Version 2.0 der Leitlinie an der ersten Fassung. Damit wird Nutzern, die bereits mit Version 1.0 vertraut sind, ein reibungsloser Umstieg ermöglicht. Zur Abwärtskompatibilität wurden in Einzelfällen Gliederungspunkte ohne Inhalt belassen. Auch die Revision der Leitlinie wurde durch die TMF ideell und finanziell gefördert.

Die Leitlinie bindet sich in ein vielfältiges Instrumentarium sinnvoller Maßnahmen im Datenmanagement von Kohortenstudien und Registern ein. Tabelle 1 zeigt ein Rahmenmodell, welches hierzu von Arts et al. [2002] erarbeitet wurde. In diesem Vorhaben wurden gezielt zwei Maßnahmen herausgegriffen, die sich im Umfeld der TMF als besonders relevant herauskristallisiert haben, ohne dass hierfür jedoch bereits klare Verfahrensanweisungen gegeben werden können. Dies sind der Originaldatenabgleich (Source Data Verification, SDV) und das Feedback.

Es wird davon ausgegangen, dass eine komplette Überprüfung der gemeldeten Daten mit der beim Datenlieferanten vorliegenden Primärdokumentation, also eine komplette SDV, zwar wünschenswert, aber nicht umsetzbar ist. Dies liegt zum einen an der fehlenden Finanzierbarkeit, zum anderen aber auch an der Unpraktikabilität dieses Vorgehens. Schwerpunkt der Leitlinie ist daher die Definition eines adaptiven Konzeptes zum Monitoring, welches eine dynamische Anpassung der Aktivitäten zur Qualitätsprüfung an die Datenquali-

1 Programm und Unterlagen der Vortragenden finden sich unter http://www.tmf-ev.de/Themen/Projekte/V020_01_Datenqualitaet.aspx.

2 Der Schlussbericht zu diesem TMF-Projekt (V020-04) ist unter www.tmf-ev.de/Leitlinie-Datenqualitaet zum Download verlinkt.

Tab. 1 Generisches Rahmenmodell zur Datenqualität in Registern aus Arts et al. [2002a]

Central coordinating centre	Local sites
Prevention during set up and organisation of registry	
At the onset of the registry <ul style="list-style-type: none"> ■ compose minimum set of necessary data items ■ define data & data characteristics in data dictionary ■ draft a data collection protocol ■ define pitfalls in data collection ■ compose data checks ■ create user friendly case record forms ■ create quality assurance plan In case of new participating sites <ul style="list-style-type: none"> ■ perform site visit ■ train new participants Continuously <ul style="list-style-type: none"> ■ motivate participants ■ communicate with local sites In case of changes (e.g., in data set) <ul style="list-style-type: none"> ■ adjust forms, software, data dictionary, protocol, training material, etc. ■ communicate with local sites 	At the onset of participating in the registry <ul style="list-style-type: none"> ■ assign a contact person ■ check developed software for data entry and for extraction ■ check reliability and completeness of extraction sources ■ standardise correction of data items Continuously <ul style="list-style-type: none"> ■ train (new) data collectors ■ motivate data collectors ■ make data definitions available ■ place date & initials on completed forms ■ keep completed case record forms ■ data collection close to the source and as soon as possible ■ use the registry data for local purposes In case of changes (e.g., in data set) <ul style="list-style-type: none"> ■ adjust data dictionary, forms, software, etc. ■ communicate with data collectors
Detection during data collection	
During import of data into the central database <ul style="list-style-type: none"> ■ perform automatic data checks Periodically and in case of new participants <ul style="list-style-type: none"> ■ perform site visits for data quality audit (registry data < > source data) and review local data collection procedures Periodically <ul style="list-style-type: none"> ■ check inter- and intraobserver variability ■ perform analyses on the data 	Continuously <ul style="list-style-type: none"> ■ visually inspect completed forms ■ perform automatic data checks ■ check completeness of registration
Actions for quality improvement	
After data import and data checks <ul style="list-style-type: none"> ■ provide local sites with data quality reports ■ control local correction of data errors After data audit or variability test <ul style="list-style-type: none"> ■ give feedback of results and recommendations ■ resolve causes of data errors 	After receiving quality reports <ul style="list-style-type: none"> ■ check detected errors ■ correct inaccurate data & fill in incomplete data ■ resolve causes of data errors After receiving feedback <ul style="list-style-type: none"> ■ implement recommended changes ■ communicate with personnel

tät ermöglicht. Diese Anpassung umfasst die Frequenz, den Umfang und die Tiefe der SDV. Die Prüfung der Datenqualität erfolgt dabei an zwei Stellen:

1. Zum einen werden Indikatoren der Datenqualität definiert, die aus den Daten eines Registers oder einer Kohortenstudie direkt erhoben werden können.
2. Zum anderen ist der Grad von Vollständigkeit, Vollständigkeit und Übereinstimmung zwischen gemeldeten Daten und der vor Ort existierenden Dokumentation ein Maß zur Anpassung des Monitorings.

Ein adaptives Monitoring verfolgt nicht das Ziel, eine vollständige Übereinstimmung der erhobenen Daten mit der Primärdokumentation sicherzustellen. Im Sinne einer Qualitätslenkung wird vielmehr eine Beeinflussung der meldenden Leistungserbringer angestrebt, um über das Konzept der kontinuierlichen Qualitätsverbesserung bereits bei der Erfassung eine hohe Datenqualität zu erzielen. Abbildung 1 verdeutlicht diesen Sachverhalt.

Derzeit gibt es zu einem adaptiven Monitoring weder Vorgaben für Register (z.B. Krebsregister) noch wissenschaftlich abgesicherte Verfahren. Alle medizinischen Forschungsverbünde mit zentraler Datenerhebung, ob als Kohorte, epidemiologische Studie oder Register, stehen daher vor der gleichen Aufgabe. Im Vorlauf zur Erstellung von Version 1.0 der Leitlinie wurde im Februar 2005 in Essen ein Workshop „Medizinische Daten in Forschung und Versorgung“ durchgeführt³. Dort wurde einerseits die Notwendigkeit eines adaptiven Monitorings aus ganz unterschiedlichen Sichten (Therapieoptimierungsstudien, wissenschaftliche Register, externe Qualitätssicherung) eindrücklich betont. Andererseits wurden die fehlende konzeptionelle Basis und der Bedarf nach einer Ausarbeitung von Leitlinien und deren konkrete Unterstützung durch Software-Werkzeuge deutlich.

Auch wenn die SDV weiterhin eine wichtige Rolle im Management von Datenqualität einnimmt, gewinnt die Datenerfassung ohne Verfügbarkeit von Originaldaten zunehmend an Bedeutung. Hierbei handelt es sich z.B. um Kohorten, bei denen die Erfassung der Daten ausschließlich zum Zwecke des Vorhabens unmittelbar durch studieneigenes Personal erfolgt. Daten werden auch bei ihrer Erhebung in Interviews oder Untersuchungen unmittelbar in zentralen Systemen erfasst, ohne einen Zwischenschritt über eine lokale, evtl. papierbasierte Dokumentation. Patienten und Angehörige geben selbst Daten zum funktionellen Outcome in zentralen Systemen zum Electronic Data Capture (EDC) ein oder führen dort Aufzeichnungen zu Schmerzen und anderen Symptomen. Daten werden aus medizintechnischen Systemen wie Sensoren automatisch übernommen, so dass das Studiensystem den einzigen permanenten Speicherort darstellt. In diesen Fällen bietet sich alternativ das Feed-

3 Programm, Unterlagen der Vortragenden und eine Zusammenfassung finden sich unter <http://www.ekmed.de/routinedaten/>.

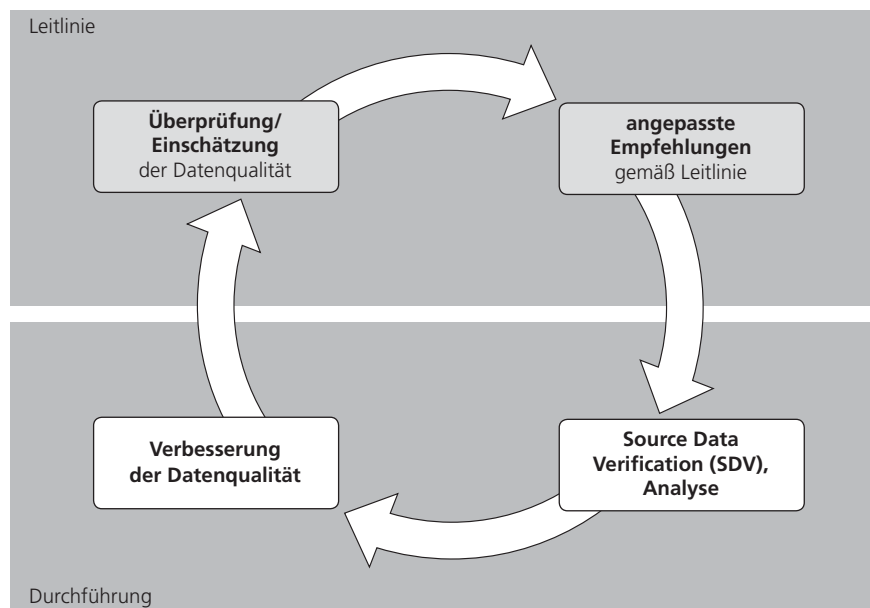


Abb. 1 Kontinuierliche Verbesserung der Datenqualität

back als Maßnahme im Qualitätsmanagement an. Die Messung der Qualität der Daten ist aber auch erforderlich, um ihre Eignung für den jeweiligen Anwendungszweck zu prüfen. Wird hier ein bestimmtes Niveau unterschritten, können Daten zur Beantwortung bestimmter Fragestellungen nicht mehr geeignet sein [Malin/Keating 2005]. Auch für die elektronische Akte in der Versorgung wurden differenzierte Anforderungen an die Datenqualität abhängig vom Anwendungszweck der Dokumentation beschrieben [Baxter et al. 2007].

Der erste Schritt bei der Entwicklung der Leitlinie in Version 1.0 war die Identifikation und Analyse der bestverfügbaren Evidenz über eine Sichtung der Literatur und eine Befragung von Experten⁴. Daran schlossen sich die Konsensusfindung zur inhaltlichen Zusammenführung und Ergänzung der Ergebnisse und die Ableitung der Leitlinie an. Ein Zwischenstand wurde der interessierten Fachöffentlichkeit im Februar 2006 in einem Workshop „Qualität medizinischer Daten in Forschung und Versorgung“ präsentiert⁵. Auf diesem Workshop wurden unter anderem in vier Beiträgen Evidenz, Empfehlungen und Anwendung der Leitlinie erläutert und diskutiert. Die Vorgehensweise orientiert sich an Empfehlungen zur Entwicklung von Leitlinien, die national erarbeitet [AWMF/ÄZQ 2001] und wissenschaftlich begründet wurden

4 Die Literatursichtung bis 2005/2006 steht unter www.tmf-ev.de/Leitlinie-Datenqualitaet zum Download zur Verfügung.

5 Programm, Unterlagen der Vortragenden und eine Zusammenfassung finden sich unter <http://www.ekmed.de/routinedaten/>.

[Bundesärztekammer/Kassenärztliche Bundesvereinigung 2005]. Das Vorgehen bei der Literatursichtung entspricht den Schritten zur Quellensuche und Quellenbewertung, wie es die BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH für die externe vergleichende Qualitätssicherung nach § 137 SGB V beschreibt [Mohr et al. 2005]. Die Fortschreibung der Leitlinie zur Version 2.0 erfolgte hingegen vor dem Hintergrund konkreter Beispiele für das Management von Datenqualität. Die gleichzeitig vorgenommene Aktualisierung der Literaturbasis ist daher komplementär zu verstehen und war nicht Ausgangspunkt zur Einführung der neuen Indikatoren.

Unabhängig von den hier gewonnenen Ergebnissen wurde festgelegt, durchgehend das Glossar der Arbeitsgruppe Qualitätsmanagement in der Medizin der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) als Referenz für die Definition qualitätsrelevanter Begriffe zu Grunde zu legen [GMDS-AG Qualitätsmanagement in der Medizin 2003]. Dieses 1996 erschienene und 2003 sowie 2007 [Sens et al. 2007] grundlegend überarbeitete und aktualisierte Glossar zu Begriffen und Konzepten des Qualitätsmanagements findet sowohl in der Fachöffentlichkeit als auch in der Praxis durchgehend Anerkennung und kann derzeit in Deutschland als De-facto-Standard gelten.



Hinweis

Bei geschlechtsspezifischen Bezeichnungen wird in der Leitlinie nur eine Form genannt. Damit sind beide Geschlechter gemeint, solange nicht ausdrücklich anders vermerkt.