

# VI Empfehlungen

## VI.1 Allgemeine Empfehlungen

Eine hohe Datenqualität ist für die Aussagekraft und Nutzbarkeit von Kohortenstudien und Registern von entscheidender Bedeutung. Die Leitlinie empfiehlt hierzu die Durchführung eines Originaldatenabgleiches (Source Data Verification, SDV) einerseits und die Implementierung eines Berichtswesens zur Datenqualität (Feedback) andererseits. Die Leitlinie sollte regelmäßig angewendet werden. Als Abstand zwischen zwei Zyklen werden sechs Monate empfohlen.

Mit dem Ziel der kontinuierlichen Verbesserung der Datenqualität wird eine SDV in Stichproben von Beobachtungseinheiten als ausreichend angesehen. Hierdurch können relevante Ressourcen im Monitoring eingespart werden. Umfang (Anzahl der Beobachtungseinheiten, für die eine SDV durchgeführt wird), Tiefe (Anzahl der Merkmale, die je Beobachtungseinheit abgeglichen werden) und Frequenz (Häufigkeit von Besuchen vor Ort) sind für jedes Zentrum an die Datenqualität anzupassen. Zur Bestimmung der Datenqualität definiert diese Leitlinie 51 Indikatoren und führt diese zu einem Score mit fünf Kategorien (sehr gut – gut – moderat – schlecht – sehr schlecht) zusammen. Über den Score sowie die Vorergebnisse bei der SDV werden Umfang und Tiefe der SDV für jedes Zentrum an die Datenqualität angepasst. Gemäß den Empfehlungen ist dann eine SDV durchzuführen, die einerseits qualitätslenkend auf die lokalen Erhebungszentren einwirkt und andererseits dem Aufdecken von Schwachstellen bei der Datenerfassung und -weitergabe aber auch im Studiendesign dient. Eine Korrektur von Daten, die bei der SDV als fehlerhaft erkannt wurden, ist nicht Ziel der Leitlinie. Diese wird nur dann empfohlen, wenn hierzu der gesamte Datenbestand – ggf. über die Verwendung zusätzlicher, für das gegebene Problem definierter Plausibilitätsprüfungen – herangezogen wird.

Das Berichtswesen dient ebenfalls der kontinuierlichen Verbesserung der Datenqualität, in dem es motivierend auf lokale Erhebungszentren einwirkt,

die im Vergleich mit anderen eine schwächere Datenqualität aufweisen. Gleichmaßen wie die SDV sind die erreichten Raten der Qualitätsindikatoren und der Qualitätsscore für jedes Zentrum, aber auch für ein Gesamtvorhaben, Grundlage einer Schwachstellenanalyse. Ein Berichtswesen sollte daher ebenfalls systematisch und regelmäßig durchgeführt werden.

## VI.2 Anwendungsempfehlungen für Register

Ulrike Bauer

### VI.2.1 Einleitung

Der Begriff des Registers wird im Forschungsbereich auf vielfältige Weise verwendet. Je nach verfolgtem Ziel gibt es unterschiedliche Arten von Registern:

- bevölkerungsbezogene, epidemiologische
- und klinische, krankheitsbezogene Register.

Insgesamt lassen sich Register beschreiben als eine „möglichst aktive, standardisierte Dokumentation von Beobachtungseinheiten zu vorab festgelegten, aber im Zeitverlauf erweiterbaren Fragestellungen, für die ein präziser Bezug zur Zielpopulation transparent darstellbar ist“ (Müller et al. 2010: S. 825). Bezüglich der Planung, Implementierung und Datenqualität liefern die „Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis“ der Deutschen Arbeitsgemeinschaft für Epidemiologie allgemein gehaltene Hinweise. Diese können als wertvoller Orientierungsrahmen genutzt werden. Die detaillierte Ausgestaltung von Maßnahmen zur Qualitätssicherung hängt sodann vom jeweiligen Register ab.

Bei der Planung eines Registers ist es zunächst unabdingbar, zu identifizieren, um welche Art von Register es sich handeln soll (s.o.).

Es soll beispielhaft am Nationalen Register für angeborene Herzfehler ein flächendeckendes, offenes klinisches Patientenregister dargestellt werden; dabei soll gezeigt werden, welche Maßnahmen zur optimalen Gestaltung der Datenqualität hierbei eingesetzt werden.

Im Nationalen Register für angeborene Herzfehler werden personenbezogen Patienten aller Altersgruppen mit jeglicher Art einer angeborenen Herzerkrankung ohne vorgegebene Endpunkte erfasst. Diese Fehlbildungen sind die häufigste angeborene Erkrankung beim Menschen. Es handelt sich um eine Vielzahl an klinischen Erscheinungsformen mit unterschiedlichsten Ausprägungen, die 70 ICD-10-Codes umfasst, welche pathomorphologisch mit 206 IPC-Codes abgebildet werden. Einzelne Entitäten treten deshalb zum Teil nur selten auf, sodass sich ausreichende Fallzahlen für aussagekräftige Studien oft nur schwer erreichen lassen.

Das seit 2000 bestehende und vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Kompetenznetzes Angeborene Herzfehler seit 2003 geförderte Nationale Register für angeborene Herzfehler e.V. erfasst deutschlandweit Daten und Biomaterial zu Personen mit angeborenen Herzfehlern. Bis 2013 wurden rund 50.000 Patienten erfasst.

Durch die zentrale Sammlung und Bündelung von Daten möchte das Nationale Register der Problematik zu geringer Fallzahlen in einzelnen Kliniken begegnen und so eine Grundlage für effiziente Forschung in verschiedenen Bereichen der klinischen, epidemiologischen und genetischen Forschung der angeborenen Herzfehler bereit stellen. Drauf aufbauend können auch spezifische Kohorten für einzelne Studienvorhaben gebildet werden.

### VI.2.2 Schwerpunkte der Qualitätssicherung in Registern

Damit sie als Grundlage für verschiedene Studien mit validen Ergebnissen herangezogen werden können, ist die Qualität der erfassten Daten in Registern von zentraler Bedeutung. Im Gegensatz zu Kohortenstudien werden in Registern keine Daten generiert, sondern vorhandene Daten erfasst. Wichtig ist deshalb in diesem Kontext, dass insbesondere geeignete Prozeduren der Erfassung und Codierung sowie der Verarbeitung der Daten Anwendung finden.

Nicht zu vernachlässigen sind als wesentliche Voraussetzungen für eine hohe Datenqualität ausreichendes und qualifiziertes Personal sowie geeignete räumliche und technische Ausstattung.

### VI.2.3 Beschreibung ausgewählter Indikatoren der Leitlinie in Registern

Die TMF-Indikatoren wurden gezielt auf das klinische Register zu den angeborenen Herzfehlern angewandt. Im Folgenden wird beispielhaft dargestellt, wie die Qualitätsindikatoren zum Einsatz kommen (s. Tab. 8).

Tab. 8 Liste von Qualitätsindikatoren mit besonderer Relevanz für Register

TMF-ID	Qualitätsindikator	Anwendungsbezogener Kommentar
<b>Ebene Integrität</b>		
TMF-1001	Übereinstimmung mit Vorwert	Neben unveränderlichen Werten wie Geburtsdatum sind über die Zeit Veränderungen von Werten möglich und damit plausibel (z.B. Veränderung der Körpergröße – Wachstum eines Kindes). Deshalb Bildung von Grenzwerten, die sich als nicht plausibel erweisen (Abnahme der Größe mit der Zeit)

TMF-ID	Qualitätsindikator	Anwendungsbezogener Kommentar
TMF-1002	Konkordanz	<p>Übereinstimmung der in der Datenbank erfassten Daten mit der Quelle; wichtig hierbei ist die Wahl der geeigneten Quelle, da diese als korrekt („wahr“) vorausgesetzt wird. (z.B. Angabe der durchgeführten Operation aus dem Operationsbericht und nicht aus dem Arztbrief).</p> <p>Ebenso Vollständigkeit der Datenerfassung (z.B. alle angeborenen Diagnosen erfasst und auch Ausschlussdiagnostik bestimmter sonst auftretender häufiger Begleitdiagnosen).</p>
TMF-1003 TMF-1035 TMF-1004 TMF-1005	Widerspruchsfreiheit	<p>Festlegungen von Plausibilitäten für sichere Widersprüche (z.B. geschlechtsspezifische genetische Syndrome).</p> <p>Warnung bei möglichen Widersprüchen (z.B. von Art des Herzfehlers abweichende Operationsmethode).</p> <p>Endlosüberlebender (schwierig bei sehr jungen und schwer kranken Patienten) – für sehr gute Datenqualität aufwendige Recherchen über regelmäßige Kontakte und evtl. Einwohnermeldeämter erforderlich.</p>
TMF-1013	Fehlende Einträge/ fehlende Werte bei Datenelementen	Unterscheidung zwischen fehlenden Werten (unvollständige Erfassung von Diagnosen und Behandlungen) und tatsächlich nicht vorhandenen Werten. (Ist in Bezug auf den jeweiligen Patienten zu werten: z.B. pränatale Diagnostik – Eintrag erst ab bestimmten Geburtsjahr relevant, da moderne Diagnostikform, bei Erwachsenen mit AHF irrelevant.)
TMF-1016	Fehlende Einträge/ Datenelemente mit Wert unbekannt	Entscheidung je nach Wert, ob unbekannt akzeptabel oder nicht; z.B. unbekannt akzeptabel bei Angabe zur eigenen Geburt und Schwangerschaft bei erwachsenen Patienten; Jedoch unbekannt möglichst vermeidbar bei Säuglingen.
<b>Ebene Organisation</b>		
TMF-1028	Aktualität der gespeicherten Daten	Unterscheidung in unveränderliche (angeborene Erkrankung) und veränderliche Daten (erworbene Erkrankungen, stattgehabte Operationen usw.). Bei veränderlichen Daten gleichzeitige Erfassung des Datums der Datenaktualisierung (Arztbriefdatum, Befunddatum usw.).
TMF-1029	Dubletten	Patientenbezogene Erfassung, keine fallbezogene Erfassung; Dubletten zu vernachlässigen.
TMF-1030	Rekrutierungsrate	<p>Nur sinnvoll für bestimmte Fragestellungen (z.B. Querschnittsuntersuchung wie Erfassung aller Neugeborenen eines Jahrganges oder einer bestimmten Untergruppe mit speziellen Einschluss- und Ausschlusskriterien) – Abgleich mit Geburtenzahlen und Versorgungsdaten</p> <p>Bezug auf Einrichtungen/Zentren – Beachtung des Zuweisungsradius (Einbeziehung des Wohnortes zum Abgleich).</p>
TMF-1034	Vorzeitig aus dem Register ausscheidende Personen	Rücknahme der Einverständniserklärung (sehr selten im Nationalen Register für angeborene Herzfehler, meistens nur bei sehr leichten Herzfehlern), Erfragen der Gründe des Ausscheidens.

TMF-ID	Qualitätsindikator	Anwendungsbezogener Kommentar
TMF-1036	Synonyme	Durch Verwendung des PID-Generators Wahrscheinlichkeit gering, Überprüfung durch Abgleich mit Diagnosen.
TMF-1037	Homonyme	
<b>Ebene Richtigkeit</b>		
TMF-1043	Genauigkeit	Unterschiedliche Verwendung der Nomenklatur für die Diagnose des Herzfehlers/Erkrankung von verschiedenen Ärzten/Kliniken. Vereinheitlichung der Nomenklatur durch Verwendung des IPC-Codes nach Wertung aller Angaben im Arztbrief und Befunden durch besonders geschultes fachspezifisches Personal (Vorteil: hohe Datenqualität bei Phänotypisierung für begleitende Biomaterialbank; Nachteil: sehr personal- und zeitaufwendig). Deshalb Einführung verschiedener Qualitätsstufen für die erforderliche Genauigkeit der Registerdaten.
TMF-1044	Übereinstimmung mit den Originaldaten bezogen auf die Datenelemente	Hoher Anspruch an die Übereinstimmung bei Angaben der Diagnosen von Patienten mit vorhandenem Biomaterial. 4-Augenprinzip, um Qualität zu optimieren. Nachteil: sehr personal- und zeitintensiv.
TMF-1045	Übereinstimmung mit den Originaldaten bezogen auf die Beobachtungseinheiten	Beobachtungseinheit ist ein Patient – Genauigkeit der Übertragung aller Angaben aus Arztbriefen. Hohe Anforderung bei Patienten mit komplizierten Krankheitsverläufen (vielen Interventionen, Komplikationen)
TMF-1046	Vollständigkeit	Bei einem breiten Register mit verschiedenen, teilweise sehr seltenen angeborenen Herzerkrankungen müssen verschiedene Subgruppen unterschiedlich in ihrer Vollständigkeit betrachtet werden, um den Dokumentationsaufwand machbar und bezahlbar zu halten: 1. maximale Vollständigkeit der Daten – Patienten mit Biomaterial 2. mittlere Vollständigkeit – komplexe, wissenschaftlich sehr interessante Fälle 3. variable Vollständigkeit – Patienten mit häufigen und wissenschaftlich weniger interessanten Herzfehlern – Vervollständigung nur nach Vorliegen bestimmter Forschungsfragen Bei epidemiologischem Ansatz: Bezug der Vollständigkeit auf Geburtenraten, Operationszahlen u.a. Angaben aus Erfassungsstatistiken.
TMF-1047	Übereinstimmung mit den Verfahrensregeln	Regelmäßige Aktualisierung der Registerdaten. Klare Verfahrensregeln für verschiedene Untergruppen im Register.
TMF-1048	Repräsentativität	Erfasste Population entspricht Stichprobe: Prozentuale Verteilung der Häufigkeiten der Herzfehler (seltene Fehler – selten im Register; häufige Fehler – häufig im Register). Ebenso Altersverteilung und Geschlechterverteilung – bestimmte Herzfehler haben eine Geschlechterwendigkeit.

### **VI.2.4 Beispielhafte Implementierung von Qualitätsindikatoren im Nationalen Register für angeborene Herzfehler e.V.**

Alle Daten werden personenbezogen erhoben, um verschiedene Datensätze eindeutig zuzuordnen und Einträge zu einzelnen Teilnehmern fortschreiben zu können. Aus datenschutzrechtlichen Gründen arbeitet das Nationale Register für angeborene Herzfehler deshalb mit zwei voneinander getrennten Datenbanken – derjenigen zur Speicherung der personenidentifizierenden Daten (IDAT-DB) und einer weiteren zur Speicherung und Verarbeitung der medizinischen Daten (MDAT-DB).

#### **VI.2.4.1 Personenidentifizierende Daten (IDAT)**

Bei der Eingabe der persönlichen Daten eines Teilnehmers in die IDAT-DB wird anhand der IDAT ein Pseudonym (Patientenidentifikator: PID) erzeugt, das für jede Person eindeutig ist und in allen Datenbanken als Zuordnungskriterium genutzt wird. Phonetische Algorithmen beachten die unterschiedlichen Schreibweisen wie z.B. „Müller“ und „Mueller“. Der PID bleibt ebenfalls für einen Patienten identisch, wenn er heiratet oder umzieht. So werden einerseits Dubletten vermieden; andererseits ermöglicht diese Art der Pseudonymisierung das kontinuierliche Fortschreiben einzelner Datensätze. Der PID selbst als alphanumerischer Code steht nicht in direktem Zusammenhang mit den zugehörigen IDAT, sodass ein „Rückrechnen“ zu den identifizierenden Daten nicht möglich ist. Für die Erzeugung des PID wird ein so genannter externer PID-Generator eingesetzt, auf den die Register-Software verschlüsselt über eine geschützte Verbindung zugreift. Durch die Sensibilität der personenidentifizierenden Daten, die sich in der Patientenliste des PID-Generators befinden, ist dieser Bereich stark geschützt und erlaubt nur einen sehr restriktiven Zugriff.

Der PID-Generator wurde im Kompetenznetz für Pädiatrische Onkologie und Hämatologie (Klaus Pommerening/Universität Mainz) als Werkzeug für die Erzeugung von pseudonymen Patientenidentifikatoren entwickelt (Reng et al. 2006). Er wird in vielen medizinischen Forschungsnetzen eingesetzt und hat sich durch seine hohe Sicherheit und das fehlertolerante Matchingverfahren als Kernmodul für das Identitätsmanagement etabliert. Durch die Eindeutigkeit der PID als Ordnungskriterium in allen Datenbanken des Netzes ist eine sichere und zuverlässige Arbeitsweise mit Sicherstellung des Datenschutzes gewährleistet.

Durch regelmäßiges Kontaktieren der Registerteilnehmer werden die Adressen gepflegt und bei Bedarf über das Einwohnermeldeamt aktualisiert. Dadurch wird auch bekannt, wenn Patienten versterben.

Als sehr effektiv hat sich bei der Datenpflege und Erreichung einer hohen Datenqualität erwiesen, dass die Erzeugung des PIDs anhand der Papiermel-

zung zentral erfolgt und die Patienten immer wieder kontaktiert bzw. auch über eine eigene Webseite informiert werden. So entsteht eine besondere „Kundenbindung“ und die Drop-outs sind sehr gering.

#### VI.2.4.2 Medizinische Daten (MDAT)

Bei Eintritt in das Register wird die Krankengeschichte zunächst retrospektiv verfolgt, indem alle angeborenen Erkrankungen und deren Verlauf bis zum Zeitpunkt der Anmeldung dokumentiert werden. Von diesem Zeitpunkt an wird die weitere Krankengeschichte mit eventuell auftretenden erworbenen Erkrankungen, medizinischen Maßnahmen und Eingriffen etc. sodann prospektiv erfasst. Als Datenquelle dienen dabei zum einen subjektive anamnestic Angaben, also Selbstangaben der Teilnehmer; zum anderen werden ärztliche Dokumente wie Arztbriefe, Katheterprotokolle, Operationsberichte oder Autopsiebefunde als Quelle genutzt. Für größtmögliche Genauigkeit und Transparenz wird die Datenquelle (Herkunft, Datum) dabei stets genau dokumentiert.

Die Daten werden ausschließlich zentral durch geschultes Personal in die Datenbank eingegeben. Alle genannten medizinischen Dokumente liegen dem Register-Personal in Papierform vor. Sind die verfügbaren Angaben nicht ausreichend, behält sich das Register die Möglichkeit vor, weitere Dokumente vom betreuenden Arzt/Zentrum anzufordern. Die Genehmigung hierfür wird bereits bei der Anmeldung neuer Teilnehmer mittels *Informed Consent* eingeholt.

Bei der Fülle der gesammelten Daten ist ein etabliertes Ordnungssystem von besonderer Bedeutung, um die Nutzbarkeit der Daten zu gewährleisten. Die Dokumentation der Diagnosen in der Datenbank folgt deshalb einer festgelegten Hierarchie. Dabei wird hinsichtlich der angeborenen Erkrankungen eine Einteilung unter anderem in führende Hauptdiagnosen und Nebendiagnosen vorgenommen. Zudem werden Prozeduren mit Angaben zu Zeit und Ort in chronologischer Reihenfolge erfasst.

Ebenso unabdingbar ist eine einheitliche Codierung der erfassten Daten. Als Nomenklatur nutzt das Nationale Register für alle Diagnosen und Prozeduren den IPC-Code; für erworbene Erkrankungen wird zusätzlich auf den ICD-10-Code zurückgegriffen. Auf Freitext wird weitestgehend verzichtet. Dadurch wird eine standardisierte Klassifizierung der erfassten Diagnosen und Prozeduren gewährleistet, wodurch Homonyme vermieden und Daten schnell und unproblematisch abgefragt bzw. genutzt werden können.

Die Eingabe sämtlicher medizinischer Daten in die Datenbank erfolgt durch geschulte Medizin-Studenten und wird durch einen verantwortlichen Arzt ständig kontrolliert. Liegt zu einem Teilnehmer neben den medizinischen Angaben eine Biomaterial-Probe vor, erfolgt zusätzlich eine zweite Stufe der Qualitätssicherung mittels Supervision. Der jeweilige Qualitätsgrad der Daten wird in der Datenbank vermerkt.

Für alle Arbeitsabläufe innerhalb des Nationalen Registers wurden spezifische *Standard Operating Procedures* (SOPs) entwickelt, welche kontinuierlich an rechtliche, technologische oder infrastrukturelle Änderungen angepasst werden. So sind die folgenden Vorgänge durch SOPs geregelt:

- Nutzung der IT-Plattform
- Organisatorische Fragen wie Zugangsrechte, regelmäßige Anleitung der Mitarbeiter, räumliche und organisatorische Trennung der Datenbanken (IDAT-DB, MDAT-DB)
- Registrierungsprozedur, Eingabe der IDAT, PID-Generierung, Teilnehmer-Feedback
- Beschriftungs- und Ablagesystem von Originaldokumenten
- Handling von originalen Arztbriefen und Eingabe der MDAT
- Qualitätskontrolle der MDAT-Eingabe
- Kontaktaufnahme mit Register-Teilnehmern (leitlinienbasierte Telefoninterviews)
- Abfragen in der MDAT-DB, Zusammenstellen von Daten für geplante Forschungsprojekte
- Herausgabe von Daten für Forschungsprojekte, Generieren einer doppelten Pseudonymisierung

Wie sich zeigt, besteht die besondere Herausforderung in einem Register wie dem Nationalen Register für angeborene Herzfehler in der Phänotypisierung, d.h. der genauen Beschreibung von Diagnosen und Prozeduren. Diese ist die Voraussetzung für die Bildung sinnvoller Kohorten z.B. mit Blick auf angeborene Herzfehler mit gemeinsamer Pathomorphologie, nach Schweregrad (bzw. derzeitigem Leistungszustand) oder, für genetische Untersuchungen, mit Blick auf angeborene Herzfehler mit bestimmten angeborenen Nebenerkrankungen, Syndromen o.ä. Je nach Studienvorhaben können dann für diese Kohorten spezielle Untersuchungsparameter gemäß den Qualitätskriterien für Kohorten erfasst werden.

### VI.2.5 Zusammenfassung und Ausblick

Das Management der Datenqualität in Registern verfolgt zwei wesentliche Ziele:

Zum einen hängen die Möglichkeiten der wissenschaftlichen Nutzung von Daten aus krankheitsbezogenen Registern vor allem von der Qualität der dort erfassten und gespeicherten Daten ab. Je nach Fragestellung der Nutzer können die Anforderungen an die Qualität der Daten sehr unterschiedlich sein. Die Entscheidung über die Verwendbarkeit der Daten kann nur der Nutzer treffen.

Zum anderen lassen sich bei der Betrachtung der Datenqualität wichtige Rückschlüsse auf die etablierten Prozesse der Datenerfassung und Datenspeiche-



rung ziehen. Darauf aufbauend lassen sich geeignete Maßnahmen treffen, um identifizierte Defizitbereiche zu beseitigen und effektiv die Datenqualität positiv zu beeinflussen.

Das Ziel der Sicherstellung einer hohen Datenqualität sollte immer im Zusammenhang mit der Machbarkeit und Finanzierbarkeit betrachtet werden. Die Leitlinien sind ein wichtiger Bezugspunkt, um bereits bei der Planung von Registern sinnvolle Entscheidungen hinsichtlich Datenqualität und -management zu treffen.

## VI.3 Anwendungsempfehlungen für Kohorten

Carsten Oliver Schmidt

### VI.3.1 Einleitung

Kohortenstudien sind in den Lebenswissenschaften ein unverzichtbares Studiendesign zur Beantwortung von Fragestellungen hinsichtlich der Verbreitung, des Verlaufes und Zusammenhangs von Erkrankungen, subklinischen Endpunkten und Risikofaktoren. Eine hohe Datenqualität ist eine notwendige Voraussetzung, um eine hohe Aussagekraft der Studien zu gewährleisten. Einen Orientierungsrahmen bieten die „Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis“, kurz GEP [DAE 2004]. Empfehlungen umfassen u. a. eine begleitende Qualitätssicherung aller relevanten Instrumente und Verfahren, eine Dokumentation für Verfahrensänderungen, eine Zweit- und Prüfeingabe für numerische Variablen sowie Plausibilitätskontrollen in einem geeigneten infrastrukturellen Rahmen zur Datenhaltung. In Hinblick auf konkrete Indikatoren zur Datenqualität und deren Operationalisierung bleibt die GEP jedoch unspezifisch. Solche praktisch anwendbaren Indikatoren bietet die vorliegende Fassung der TMF-Leitlinie Datenqualität. Sie wurde gegenüber der letzten Fassung um mehrere Indikatoren erweitert, die im Rahmen der Qualitätssicherung von Beobachtungsstudien relevant sind.

Der folgende Überblick erläutert spezifische Aspekte der Sicherung von Datenqualität in Kohortenstudien, führt ausgewählte Indikatoren der TMF-Leitlinie Datenqualität ein und beschreibt am Beispiel der Study of Health in Pomerania, SHIP [Völzke et al. 2011], eine exemplarische Umsetzung im Rahmen eines kontinuierlichen Datenmonitorings. Die SHIP-Studie ist eine bevölkerungsbezogene Gesundheitsstudie, bestehend aus zwei Kohorten mit insgesamt über 8800 Teilnehmern, die im Rahmen des Forschungsverbundes „Community Medicine“ an der Universitätsmedizin Greifswald durchgeführt wird. Die Studie erfasst erstens die Prävalenz und Inzidenz von Erkrankungen, subklinischen Auffälligkeiten und Risikofaktoren. Zweitens wird der Zusammenhang zwischen diesen im Quer- und Längsschnitt untersucht.

Insofern Kohortenstudien auch die Nutzung von sekundären Datenquellen vorsehen, etwa die Nutzung von Daten aus klinischen Tumorregistern, Krankenhäusern oder der kassenärztlichen Vereinigung, greifen primär Indikatoren zur Qualitätskontrolle bei Registern (s. Kap. VI.2), auf die hier nicht eingegangen wird. Umgekehrt bieten sich wichtige Möglichkeiten zur Validierung von Sekundärdaten auf Basis von Primärdaten aus Kohortenstudien, was im Ausblick beschrieben wird.

### VI.3.2 Schwerpunkte der Qualitätssicherung in Kohortenstudien

Im Gegensatz zu Registern und Datenrepositorien fokussiert die Qualitätssicherung in Kohortenstudien auf den Prozess der Datengenerierung. Daten sollen reliabel, valide und vollständig erhoben werden. Hierin besteht eine wesentliche Gemeinsamkeit mit klinischen Studien, wobei in Hinblick auf letztere ein wesentlich engerer, in Teilen gesetzlich geregelter Rahmen zur Qualitätssicherung besteht [Baigent et al. 2008, Eberhard/Herrlinger 2010, Macefield et al. 2013].

Primäre Ziele der Qualitätssicherung in Kohortenstudien sind die Minimierung von Messfehlern, die Gewährleistung einer hohen Generalisierbarkeit sowie der internen Vergleichbarkeit der Messungen innerhalb und zwischen den Erhebungswellen. Dazu sind zahlreiche Maßnahmen auf konzeptueller, struktureller und prozessualer Ebene notwendig [DAE 2004, Rajaraman/Samet 2005, Whitney et al. 1998, Lüdemann et al. 2000].

Konzeptuell ist ein geeignetes Studienprotokoll grundlegend. Dieses definiert u.a. Fragestellungen und Hypothesen, regelt die Stichprobenauswahl, Rekrutierungsmethoden, die Instrumentenauswahl und andere methodische Fragen. Viele der im Folgeabschnitt genannten Indikatoren liefern nur unter Verwendung standardisierter und vorab validierter Messverfahren sinnvolle Ergebnisse.

Hinsichtlich der strukturellen Ebene sind wesentliche Voraussetzungen zur Gewährleistung einer hohen Datenqualität u.a. geeignete Räumlichkeiten, die Qualifikation des Personals, Organisationsstrukturen mit einer nach Möglichkeit personellen Trennung von Untersucher- und Qualitätssicherungsfunktionalitäten, die IT-Infrastruktur zur Dateneingabe, u.a. elektronische Datenerfassungssysteme (Electronic Data Capture, EDC), Datenspeicherung und Datenmanagement.

Prozessual kann zwischen Maßnahmen unterschieden werden, die im Vorfeld der Datenerhebung stattfinden und solchen, die an den generierten Daten während und nach der Dateneingabe ansetzen. Ersteres umfasst u.a. Maßnahmen zur Qualifizierung des Personals, die Sicherstellung einer geeigneten Gerätefunktionalität und letzteres Maßnahmen zur Eingabe und zum Monitoring der generierten Daten. Einen Überblick wichtiger Maßnahmen bietet Tabelle 9.

Tab. 9 Prozessuale Elemente der Qualitätssicherung in Kohortenstudien

Maßnahmen vor der Datenerhebung
Studienhandbücher für Untersuchungen, Datenmanagement, Qualitätssicherung
Checklisten zur Untersuchungsdurchführung
Training/(Re-)Zertifizierung von Untersuchern, Befundern und ggfs. Trainern bei multizentrischen Projekten
Gerätekalibrierung, Gerätevergleiche, Gerätewartung
Phantommessungen bei bildgebenden Verfahren
Gespräche/Qualitätszirkel zu qualitätsrelevanten Aspekten der Untersuchung
Pilotierung der Studie/Prätest einzelner Untersuchungsmodule
Maßnahmen während und nach der Datenerhebung
Kontrolle der Teilnahmevoraussetzungen (Identität und Einverständnisse)
Standardisierte Dateneingaben (z.B. webbasiert mit automatischer Plausibilitätskontrolle)
Mehrfachbefundungen (u.a. bei bildgebenden Verfahren)
Mehrfacheingaben (Standard bei Eingabe von Papiervorlagen, z.B. Fragebögen)
On-Site Monitoring (Räumlichkeiten sowie Untersuchungsprozess)
Standardisierte und zentralisierte Erfassung von Auffälligkeiten (z.B. datenbankbasiert)
Qualitätsberichte
Monitoring der laufenden Datenerhebung (Fokus auf Rohdaten)
Datenmanagement, syntaxbasierte Datenbereinigungen
Externes Monitoring (Advisory Board)

Zur Kontrolle der laufenden Datenerhebungen ist ein zentrales Datenmonitoring, wie es auch in die auch in randomisierten Studien empfohlen ist [Baigent et al. 2008, De 2011], wesentlich, um Datenauffälligkeiten effizient detektieren zu können.

Die Implementation von Maßnahmen zur Qualitätssicherung und Qualitätskontrolle ist als Rückkopplungsschleife zu verstehen, bei der alle Maßnahmen während und nach der Datenerhebung alle Maßnahmen bedingen können, die vor der Datenerhebung genannt sind. Dies schließt auch Inhalte der konzeptuellen und strukturellen Ebene ein, die sowohl vor als auch während der Studie Gegenstand von Anpassungen sein können. Insgesamt entspricht dies dem in der Einleitung genannten Anspruch auf eine kontinuierliche Verbesserung der Datenqualität (s. Abb. 1).

Nach Abschluss der Datenerhebung folgt die finale Datenbereinigung, die vollständig skriptbasiert zu erfolgen hat, um alle Unterschiede zwischen Roh- und Auswertungsdaten nachvollziehen zu können. Anschließend greifen wei-

tere qualitätssichernde Prozesse zum geeigneten Umgang mit den bereinigten Ergebnisdaten. Beispiele sind ein standardisiertes, skriptbasiertes Datenübergabeprozedere oder ein Publikationskomitee. Durch einen Zugriff auf Auswertungssyntax und Manuskripte erhält Letzteres, wie z.B. in der SHIP-Studie, vor der Publikation die Möglichkeit, Autoren Hinweise zu geben.

Eine besondere Herausforderung in komplexen Kohortenstudien wie SHIP, der Nationalen Kohorte [Wichmann et al. 2012] oder dem WHO-MONICA Projekt [Tolonen et al. 2006] besteht darin, die Vergleichbarkeit der Messwerte zwischen verschiedenen Untersuchern, Zentren, Labors, Geräten über einen langen Zeitraum zu gewährleisten. Zentral ist also die interne Validität der Messergebnisse innerhalb einer Studie. Daher ist es nicht das primäre Ziel, die neuesten Gerätegenerationen oder Softwareupdates umgehend zu implementieren, sondern eine einmal gewählte Technologie so lange wie möglich beizubehalten und jeden Wechsel zwischen Technologien und Methoden durch Vergleichsuntersuchungen zu flankieren. Relativ kurze Lebenszyklen technischer Geräten erschweren dies, wenn alte Geräte nicht mehr zur Verfügung stehen. Durch eine erhöhte Sensitivität z.B. moderner bildgebender Technologien werden morphologisch ähnliche Endpunkte anders visualisiert, Detektionsraten klinischer Endpunkte verändert und dadurch die Vergleichbarkeit zwischen den Erhebungswellen gesenkt. Ein weiteres Beispiel sind zwischen und innerhalb von Erhebungswellen auftretende Veränderungen der laboranalytischen Untersuchungsmethoden, was kritisch ist, da die Variabilität der Messwerte zwischen Assays erheblich sein kann [Lai et al. 2012].

Ein Datenaudit, bei dem ein externes Team die Forschungsdaten mit den Prüfbögen vergleicht, spielt im Vergleich zu klinischen Studien aufgrund der anderen rechtlichen und finanziellen Rahmenbedingungen eine untergeordnete Rolle. Zudem wird durch ein solches Vorgehen nur ein Teil qualitätsrelevanter Fehler detektiert [Nahm et al. 2008]. Nichtsdestotrotz ist eine Quelldatenverifizierung auch in Kohortenstudien in Bezug auf ausgewählte Untersuchungen Routine. Ein Beispiel ist der Abgleich von Papiervorlagen mit den elektronischen Daten, etwa bei Fragebögen, Einverständniserklärungen oder Geräten, die keine automatische Übertragung von Messwerten erlauben [Lüdemann et al. 2000].

### VI.3.3 Beschreibung ausgewählter Indikatoren der Leitlinie in Kohortenstudien

Die im Rahmen dieser TMF-Leitlinie beschriebenen Indikatoren der Datenqualität fokussieren auf die Ergebnisqualität vorliegender Daten. Nicht im Fokus stehen also die beschriebenen konzeptuellen und strukturellen Rahmenbedingungen sowie qualitätssichernde Maßnahmen vor und während der Datenerhebung. Dies war mit der vorliegenden Version der Leitlinie auch nicht intendiert. Damit spielen die TMF-Indikatoren vor allem im Datenmonitoring eine Rolle.

Methodisch lassen sich die drei Ebenen der TMF-Leitlinie vereinfacht wie folgt charakterisieren: Die Ebene Integrität fokussiert auf die Reliabilität der Messwerte und das Fehlen einzelner Datenelemente (Vollständigkeit) sowie im geringeren Maße auf die interne Validität. Die Ebene Organisation fokussiert auf Indikatoren für das Fehlen kompletter Beobachtungseinheiten (Vollzähligkeit), was für die Generalisierbarkeit wesentlich ist, und die Ebene Richtigkeit auf die externe Validität. Allerdings bestehen zwischen den Indikatoren der drei Ebenen der TMF-Leitlinie teils enge Bezüge, weshalb die eindeutige Zuordnung nur begrenzt möglich ist.

Gegenüber der alten TMF-Leitlinie Datenqualität [Nonnemacher et al. 2007] bestehen Veränderungen mit Relevanz für Kohortenstudien vor allem in der Verfeinerung und Ergänzung von Indikatoren auf der Ebene Integrität [Stausberg et al. 2012]. Ein Beispiel ist die Differenzierung verschiedener Ursachen von Messfehlern. Eine kommentierte Übersicht ausgewählter Indikatoren mit Relevanz für Kohortenstudien wird in Tabelle 10 gegeben.

Vergleicht man die drei Ebenen der TMF-Qualitätsindikatoren zwischen Registern und Kohortenstudien, zeigt sich deren unterschiedliche Gewichtung in der praktischen Anwendung. Diese ist in unterschiedlichen Vorgehensweisen in beiden Anwendungsfeldern begründet, Qualitätsindikatoren zu bilden. Kohortenstudien bieten bessere Möglichkeiten, Indikatoren zur Reliabilität und fehlenden Werten zu generieren als zur internen oder externen Validität. Meistens fehlt eine externe Quelle, die zur Überprüfung der Validität der Daten dienen kann. Bei Registern besteht hingegen ein starker Fokus auf die Ebene Richtigkeit im Sinne einer Quelldatenverifizierung. Sofern dies in Kohortenstudien nicht möglich ist, stellt dies keine qualitative Überlegenheit der Register dar. Im Gegenteil: Inwieweit die einem Register zu Grunde liegenden Quellen ihre Daten reliabel und valide erfassen, bleibt weitgehend offen, während genau dies in Kohortenstudien zentraler Gegenstand der Qualitätskontrolle ist und durch geeignete Qualitätssicherungsmaßnahmen gewährleistet werden kann.

Bei der Anwendung von Schwellenwerten der Leitlinie sind im Rahmen der Kohortenstudie zwei Ebenen zu unterscheiden: Erstens Schwellen zur Entscheidung über Maßnahmen der Qualitätssicherung im Rahmen einer laufenden Datenerhebung. Diese fokussieren auf Rohdaten, eine Implementierung wird beispielhaft im folgenden Abschnitt beschrieben. Zweitens Schwellen zur abschließenden Bewertung der erreichten Datenqualität auf der Ebene des Auswertungsdatensatzes. Letzteres kann für die Planung von Datenanalysen genutzt werden, etwa durch Stratifizierung der Analysen über verschiedene Qualitätsstufen bei multizentrischen Studien [Tolonen, et al. 2006]. Ein weiteres Anwendungsgebiet ist die Nutzung der Indikatoren zur retrospektiven Harmonisierung vorliegender Daten, um die Qualität von Analysen aus gepoolten Studien zu verbessern [Fortier et al. 2011].

Tab. 10 Liste von Qualitätsindikatoren mit besonderer Relevanz für Kohortenstudien

Qualitätsindikator	Anwendungsbezogener Kommentar
Übereinstimmung mit Vorwert (TMF-1001)	Die Bildung von Grenzwerten für Werte, die sich unplausibel von ihren Vorwerten unterscheiden, ist zur Beurteilung der internen Validität wichtig. Da intraindividuell oft große Veränderungen über die Zeit plausibel sind, kommt Lagemaßen (z.B. Mittelwert) für ein Untersuchungskollektiv eine wichtige Rolle zu.
<b>Werte Verteilung (TMF-1006)</b>	
Bevorzugung bestimmter Endziffern (TMF-1007)	Anfällig sind Messungen mit nicht-digitaler Anzeige, bei denen manuell gerundet werden muss. Entsprechend wird das Auftreten durch geeignete Gerätewahl mit möglichst automatischer Datenübertragung minimiert.
Werte Verteilung der durch Untersucher erfassten Parameter (TMF-1009)	Hauptursache für Messfehler in Kohorten, der Messfehler schwankt stark zwischen Untersuchungstypen. Eine randomisierte Zuweisung zwischen Untersuchern und Probanden ist wichtig, damit im Durchschnitt keine Messunterschiede zwischen Untersuchern auftreten. Darauf beruht die Bildung statistischer Kennwerte.
Werte Verteilung der durch Geräte erfassten Parameter (TMF-1010)	Wird in der Praxis auch bei geeichten Geräten regelmäßig beobachtet und bedarf fortwährender Kontrolle. Eine randomisierte Zuweisung, sowohl gegenüber Probanden als auch Untersuchern ist für die Bildung von statistischen Kennwerten wichtig.
Werte Verteilung von Befunden (TMF-1011)	Quantitative Befunde sind im Allgemeinen weniger anfällig für relevante systematische Fehler als qualitative. Insbesondere die qualitative Beurteilung bildgebender Ergebnisse erweist sich in der praktischen Umsetzung als problematisch.
Werte Verteilung von Parametern zwischen Zentren (TMF-1052)	Ist eine Variante der TMF-Indikatoren 1009 und 1010, bei der mittlere Unterschiede zwischen Untersuchungszentren fokussiert werden. Die Interpretation ist dadurch erschwert, dass Populationsunterschiede mit Messfehlern konfundiert sind, dem nur in Teilen durch eine adjustierte Analyse begegnet werden kann.
<b>Fehlende Einträge</b>	
Fehlende Module (TMF-1012)	Bezeichnet die Nichtrealisierung von kompletten Untersuchungsmodulen und kann sowohl designbedingt sein als auch durch Selbstselektion. In Hinblick auf letzteres siehe TMF-1032.
Fehlende Werte bei Datenelementen (TMF-1013)	Siehe 1014
Fehlende Werte bei mandatorischen Datenelementen (TMF-1014)	Fast alle Datenelemente in Kohortenstudien sind mandatorisch. Sofern geeignete Eingabekontrollen vorliegen, etwa durch EDC mit Plausibilitätskontrollen, sind fehlende Werte mit 0% Toleranzschwelle zu bearbeiten. Kritischer sind Fragebögen, da Probanden bei bestimmten Fragen mit erhöhter Wahrscheinlichkeit keine Angaben machen.

Qualitätsindikator	Anwendungsbezogener Kommentar
Fehlende Werte bei optionalen Datenelementen (TMF-1015)	Spielen in Kohortenstudien eine untergeordnete Rolle. Sind eher bei Nutzung von Daten aus der klinischen Routine relevant.
Datenelemente mit Wert unbekannt o.ä. (TMF-1016)	Sollten bei elektronischen Eingaben in geschlossenen Ergebnisformaten nicht aufgetreten, daher 0% Toleranz
Ausreißer bei stetigen Datenelementen (TMF-1018)	Jeder Ausreißer bei vordefinierten Plausibilitätsgrenzen ist Gegenstand von Qualitätssicherungsmaßnahmen, 0% Toleranz.
Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten (TMF-1019)	Im Kontext von Laboranalysen bei Bevölkerungskohorten relevant, da durch den hohen Anteil gesunder Probanden Messbarkeitsgrenzen systematisch unterschritten sein können.
<b>Unerlaubte Werte</b>	
Unerlaubte Werte bei qualitativen Datenelementen (TMF-1021)	Sollten bei EDC in geschlossenen Ergebnisformaten nicht auftreten, daher 0% Toleranz. Jeder Fall ist Gegenstand von Qualitätssicherungsmaßnahmen.
Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings (TMF-1022)	Sollten bei EDC in geschlossenen Ergebnisformaten nicht auftreten, daher 0% Toleranz. Jeder Fall ist Gegenstand von Qualitätssicherungsmaßnahmen.
Unerlaubte Werte zur Kodierung von fehlenden Modulen (TMF-1023)	Sollten durch Definition von geeigneten Sprungregeln nicht auftreten, daher 0% Toleranz. Jeder Fall ist Gegenstand von Qualitätssicherungsmaßnahmen.
Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen (TMF-1024)	Sollten bei EDC in geschlossenen Ergebnisformaten oder bei automatischem Übertrag aus Messgeräten nicht auftreten, daher 0% Toleranz. Jeder Fall ist Gegenstand von Qualitätssicherungsmaßnahmen.
Nachweis bekannter Korrelationen (TMF-1027)	Kann als Element der Qualitätskontrolle unter Nutzung von verfügbaren Informationen aus Vorwellsen oder anderen Studien mit ähnlicher Methodik genutzt werden. Problematisch ist die Definition von zulässigen Grenzen für eine Bewertung qualitätsrelevanter Abweichungen.

### VI.3.4 Beispielhafte Implementierung von Qualitätsindikatoren in der SHIP-Studie

SHIP ist nicht auf die Erfassung bestimmter vordefinierter Erkrankungen, subklinischer Auffälligkeiten oder Risikofaktoren ausgelegt [Völzke et al. 2011]. Vielmehr werden gesundheitsrelevante Endpunkte in einer maximal möglichen Breite gemessen. Entsprechend umfasste SHIP in der letzten Erhebungswelle mehrere dutzend Untersuchungen mit einer Gesamtdauer von über 20 Stunden. Diese reichten von einem Interview und Fragebögen über eine Blutdruckmessung, kardiovaskuläre Untersuchungen, bildgebende Verfahren incl. Ultraschall und MRT bis hin zu einer dermatologischen Untersuchung und Schlaflabor. Dadurch entstehen sehr hohe Anforderungen an die Imple-

mentation einer leistungsfähigen Qualitätssicherung, im Rahmen derer ein begrenzter Mitarbeiterstab vielfältige Untersuchungen mit heterogenen Anforderungen betreuen muss.

Strukturell ist das Organisationszentrum für die Studienplanung verantwortlich, während das Untersuchungszentrum die SHIP-Kernuntersuchungen durchführt. Hinzu kommen assoziierte Kliniken, die spezielle Untersuchungsmodule durchführen, z.B. die Radiologie oder Dermatologie. Für jede Untersuchung sind auf Seiten des Untersuchungszentrums Untersuchungs- und Geräteverantwortliche definiert, auf Seiten des Organisationszentrums ein Qualitätsverantwortlicher. Dieser ist für die geeignete Implementierung und Durchführung der Untersuchungen verantwortlich.

Soweit das Monitoring von Daten im Vordergrund steht, kommt teilautomatisch generierten Qualitätsreports bei SHIP eine herausgehobene Rolle zu. Diese werden für alle Untersuchungsmodule generiert und operationalisieren zahlreiche Indikatoren der TMF-Leitlinie. Die vierteljährlich, bei Bedarf auch öfter generierten Qualitätsreports haben folgende Merkmale und Voraussetzungen:

- In den verschiedensten Untersuchungsmodulen sind ähnliche Indikatoren sinnvoll einsetzbar. Im Fokus der Reports stehen die Anzahl fehlender Werte (TMF-1012–1016), Extremwerte (TMF-1018), die Variabilität von Messwerten zwischen Untersuchern (TMF-1009), Geräten (TMF-1010) und Befundern (TMF 1011), die Bevorzugung bestimmter Endziffern (TMF-1007), sowie Zeittrends (TMF-1009–1011). Unerlaubte Werte (TMF-1021–1026) werden bereits vorab im Datenmanagement erfasst.
- Hauptzielsetzung der Berichte ist die kontinuierliche Verbesserung der Datenqualität während der laufenden Studie. Daher ergeben sich strenge Grenzwerte für die Einleitung qualitätssichernder Maßnahmen, die bereits bei Verdacht auf Probleme ansetzen.
- Der statistische Berichtsteil wird inkl. der Tabellen und Grafiken durch Skripte automatisch generiert, manuell erfolgt lediglich die Interpretation durch die Qualitätsverantwortlichen, die in einer Datenbank über eine Webeingabemaske eingetragen wird. Die automatische Zusammenstellung ist Voraussetzung zur breiten Abdeckung der erhobenen Variablen. Der Report wird über LaTeX erstellt und als PDF-Dokument ausgegeben.
- Alle erhobenen Variablen unterliegen einem Screening auf Auffälligkeiten, dies umfasst die o.g. Aspekte exklusive Endziffern und Zeittrends. Ein Teil der Variablen wird mit Fokus auf die Variabilität der Messwerte und Zeittrends vertieft analysiert und graphisch dargestellt. Über die Zuordnung zu einer vertieften Analyse entscheiden die Bedeutung einer Variable für spätere Analyseprojekte und das erwartete Risiko für Datenauffälligkeiten.
- Die Zuweisung von Probanden zu Untersuchern erfolgt annähernd randomisiert, da die Terminvergabe Probandenmerkmale nicht berücksich-



tigt. Aus dieser quasi-experimentellen Situation ergibt sich, dass der Erwartungswert für beobachtete Probandenmerkmale zwischen Untersuchern identisch sein sollte. Abweichungen von Messwerten zwischen Untersuchern weisen unter der Voraussetzung ausreichender Fallzahlen auf systematische Fehler hin, die einer Abklärung bedürfen. Hinsichtlich der Annahme gleicher Erwartungswerte besteht eine wichtige Einschränkung: Probandenmerkmale verändern sich im Verlauf einer Erhebungswelle, ein wichtiger Einflussfaktor sind Rekrutierungsmethoden [Haring et al. 2009]. Daher gilt die Annahme gleicher Erwartungswerte nicht für Untersucher, die zu verschiedenen Zeiträumen tätig sind.

- Alle Indikatoren werden hinsichtlich der Entscheidung über qualitäts-sichernde Maßnahmen unabhängig voneinander betrachtet. Daher wurde bislang kein übergreifender Score entwickelt und genutzt.

Weitere qualitätsrelevante Indikatoren werden außerhalb der Qualitätsreports gebildet, da sie für keine einzelne Untersuchung spezifisch sind. Dazu zählen die Rekrutierungsrate (TMF 1030), Dubletten (TMF-1029), Verweigerungsraten (TMF-1031) oder vorzeitig ausscheidende Beobachtungseinheiten (TMF-1034).

Am Beispiel der Anwendung von Schwellen zur Entscheidung über Maßnahmen der Qualitätssicherung im Rahmen einer laufenden Datenerhebung in SHIP zeigen sich sehr deutliche Unterschiede zu Registern. Laut TMF-Leitlinie liegen empfohlene Schwellen für den Anteil von Werten mit Auffälligkeiten bei fehlenden Werten bei mandatorischen Datenelementen bei 5%, der Grenzwert für Auffälligkeiten Widerspruchsfreiheit liegt bei 5% und der für Extremwerte bei 10%. Im Gegensatz hierzu liegen die entsprechenden Grenzwerte bei SHIP für diese Felder bei 0%. Der Grund ist, dass aufgrund des EDC weder fehlende noch unbekannte Werte im Rahmen der Datenerhebung auftreten dürfen. Jeder Einzelfall führt zur Überprüfung, um eine Wiederholung zu vermeiden. Auch bei Extremwerten führen vordefinierte Plausibilitätsgrenzen zu einem Routinecheck aller außerhalb liegender Werte mit den Untersuchern und ggfs. den Fachverantwortlichen für die Untersuchung.

Komplizierter gestaltet sich die Bewertung auffälliger Werteverteilungen. Die Variabilität der Messwerte zwischen Untersuchern, Befundern und Geräten wird über Intraklassenkorrelationen (ICC) und Design-Effekte bestimmt sowie im Verlauf graphisch auf Basis von nichtparametrischen Regressionsmodellen visualisiert. Weil verschiedene Untersucher unterschiedliche Probanden untersuchen, indiziert eine niedrige Intraklassenkorrelation wenig systematische Abweichungen. Da sowohl graphische als auch statistische Ergebnisse genutzt werden, ist kein strikter Grenzwert vorgegeben. Üblicherweise werden Variablen mit  $ICC > 0,02$  Gegenstand weiterer Betrachtungen.

Zusätzliche Grenzwerte spielen bei der Auswertung von Zertifizierungsergebnissen eine Rolle, auf deren Basis sich sowohl Maße zur Inter- wie Intra-Rater-Reliabilität bilden lassen. Visualisiert werden Zertifizierungsergebnisse im

Fälle kontinuierlicher Variablen über Bland und Altman Plots [Bland/Altman 1999]. Der Standard in SHIP für das Bestehen einer Zertifizierung ist u.a. weniger als 5% durchschnittliche Abweichung (mean bias) im Vergleich zum Goldstandard [Völzke et al. 2011]. Bei dichotomen Outcomes sollte ein Kappa-Wert  $> 0,8$  erreicht sein.

Die erhebliche Komplexität der Implementierung der Studie zur Gewährleistung eines aussagekräftigen Monitorings kann am Beispiel bildgebender Verfahren demonstriert werden. Mögliche Einflussfaktoren für die Ergebnisse sind

1. das Messgerät zur Erstellung der Bilder,
2. der Untersucher, der diese Bilder aufnimmt,
3. der Befunder, der die Bilder auswertet und
4. die Technik, die der Befunder verwendet (z.B. Monitor, Software).

Genaugenommen sind auch die räumlichen Verhältnisse, z.B. angemessene Verdunkelung, unter denen die Bilder gemacht und befundet werden, relevant. Alle Ebenen voneinander zu trennen ist statistisch komplex und kann im Falle von Studien mit mehreren Untersuchern, Geräten und Befundern nur dadurch adäquat gelöst werden, dass erstens eine vollständige Dokumentation dieser Faktoren erfolgt und zweitens diese Einflussfaktoren unabhängig voneinander sind. Verhindert werden muss also, dass z.B. Mitarbeiter A bevorzugt Raum 2 mit Gerät 2 bedient und dies in der Frühschicht, während Mitarbeiter B Raum 1 mit Gerät 1 nimmt und vorwiegend Wochenenden und Nachmittage betreut und beide dazu ihre eigenen Bilder befunden. In diesem Fall ist die Beurteilung der Fehlerquellen untrennbar konfundiert. Letztlich sind also aussagekräftige Metadaten erforderlich, um die Belastbarkeit von statistischen Routinen und Ergebnissen besser deuten zu können, was im Indikator TMF-1050 abgebildet werden kann.

### VI.3.5 Zusammenfassung und Ausblick

Indikatoren zur Messung der Datenqualität in Kohortenstudien sind unverzichtbar und sind in der aktuellen TMF-Leitlinie Datenqualität aufgeführt. Dass diese Indikatoren nicht alle Aspekte der Qualitätssicherung in Kohortenstudien umfassen, liegt am Fokus auf die Bewertung vorliegender Daten. Das Implementationsbeispiel in SHIP verdeutlicht, dass eine praktikable Bildung von Indikatoren in einer komplexen Kohortenstudie IT-Lösungen erfordert. Ein offener Punkt ist die Aggregation verschiedener Indikatoren zu globalen Kennwerten der Studienqualität. Diese Aggregation gestaltet sich kompliziert, da Kohortenstudien ein sehr breites Spektrum von Endpunkten und Methoden umfassen. Nur wenige Indikatoren sind für eine Kohortenstudie von globaler Bedeutung, z.B. die Rekrutierungsrate, während die meisten anderen spezifisch für einzelne Untersuchungsmodule und darin enthaltene Variablengruppen sind. Mit der gelungenen Implementation eines Interviews wird nichts über die Qualität bildgebender Verfahren in der gleichen Studie ausgesagt, da

sich Techniken und Personal unterscheiden. Daher ist z.B. die Bildung von Mittelwerten der Indikatoren über alle Untersuchungsmodule wenig aussagekräftig. Zielführender ist eine Aggregation auf der Ebene einzelner Untersuchungsmodule. Daraus lassen sich Hinweise zu potenziellen Problematiken bei konkreten Analyseprojekten gewinnen.

Geeignete Studiendesigns sind eine wesentliche Voraussetzung, um aussagekräftige Indikatoren zur Datenqualität zu generieren. Einige Designelemente sollten dabei eine größere praktische Bedeutung erhalten als bisher. Beispielsweise würden Validierungsstudien innerhalb einer Kohortenstudie vielfältige Optionen eröffnen, Indikatoren der Ebene Richtigkeit zu generieren. Sofern Sekundärdatenquellen in Kohortenstudien genutzt werden, bieten sich leistungsfähige Möglichkeiten zur Bildung von Indikatoren der Ebene Richtigkeit für die Register selbst. Standardisierte und zumeist besser qualitätsgesicherte Messungen der Kohortenstudien könnten zur Validierung von Angaben aus Sekundärdatenquellen dienen, beispielsweise in Hinblick auf chronische Erkrankungen. Ein Beispiel ist die Abbildung von Hypertonie von Patienten in Daten der Kassenärztlichen Vereinigung im Vergleich zu den standardisierten Messungen einer Kohortenstudie.

## VI.4 Anwendungsempfehlungen für Data Repositories

Thomas Schrader

### VI.4.1 Einleitung

Unter einem klinischen Forschungsdatenregister wird ein Repository verstanden, welches klinische Daten aus unterschiedlichen Quellen für Forschungszwecke zugänglich macht [Eminaga et al. 2013, Hruby et al. 2013, Piwowar/Vision 2013, Sinaci et al. 2013]. Dabei können unterschiedliche Strategien umgesetzt werden:

1. Die Datenhaltung erfolgt in einem zentralen Register als sog. Centralized Research Data Repository [Hruby et al. 2013].
2. Es wird ein föderierter Ansatz verfolgt, der die eigentlichen Daten in den Quellen hält. Die Abfrage erfolgt via Metadaten über die Einzelquellen [Ngouongo et al. 2013, Sinaci et al. 2013, Weber 2013].

Allen gemeinsam ist die Tatsache, dass der Aufbau und die Datenhaltung an keine spezifische Forschungsfrage gebunden sind. Vielmehr steht die Erschließung von klinischen Daten im Vordergrund unter dem Aspekt der Wiederverwendung dieser Daten (secondary use of clinical data).

Das Open European Nephrology Science Center (OpEN.SC) ist ein Forschungsdaten-Repository für klinische Daten aus dem Bereich der Nephrologie. Es war ein von der Deutschen Forschungsgemeinschaft gefördertes Projekt aus dem

Bereich der Digitalen Bibliotheken, welches 2013 abgeschlossen wurde. Entstanden ist eine internetbasierte Plattform mit einer serviceorientierten Architektur (SOA), die in der Lage ist, innerhalb von kurzer Zeit klinische Daten aus unterschiedlichen Einrichtungen und dementsprechend verschiedener Infrastruktur in einem Repository zusammenzuführen und für Forschungsfragen zugänglich zu machen. Aus drei nephrologischen Kliniken sowie dem Institut für Pathologie der Charité, Universitätsmedizin Berlin, wurden ca. 20.000 Fälle für die Forscher erschlossen und zugänglich gemacht. Dabei wurden Krankheitsverläufe von mehr als 15 Jahren ebenso eingeschlossen wie auch vollständig digitalisierte histologische Schnitte als sogenannte Whole Slide Images (WSI).

OpEN.SC verstand sich als eine Einrichtung der Datenhaltung und des Datenmanagements. Die Aufgabe ist mit der einer „Research Data Stewardship“ zu vergleichen [Marco 2000]. Aus dieser Rolle im Lebenszyklus wissenschaftlicher Daten ergaben sich zum einen eine besondere Verantwortung und zum anderen eine besondere Sichtweise auf Daten und Datenqualität.

### **VI.4.2 Schwerpunkte der Qualitätssicherung in einem Daten-Repository**

2007 wurde die vorliegende Leitlinie in Version 1.0 publiziert [Nonnemacher et al. 2007]. Darin wurden zum ersten Mal systematisch für den medizinischen Bereich Datenqualitätskriterien vorgestellt und in den Kontext von Tumoregistern und Kohortenstudien gebracht. 2012 wurden einige Qualitätsparameter ergänzend aufgenommen, die sich als bedeutungsvoll für andere Arten von Datenbeständen wie Kohorten (Study of Health in Pomerania – SHIP, Greifswald) und Forschungsdatenregistern wie OpEN.SC erwiesen haben [Stausberg et al. 2012].

### **VI.4.3 Beschreibung ausgewählter Indikatoren der Leitlinie für Forschungsdaten-Repositories**

Im Rahmen des OpEN.SC Projektes wurde sehr frühzeitig begonnen, das Thema Datenqualität aufzugreifen. Dabei wurde in der Diskussion festgestellt, dass die Meinungen und das Verständnis von Qualität sehr stark divergierten. Es wurde klar, dass klinisch tätige Kolleginnen und Kollegen andere Anforderungen an die Daten haben als beispielsweise Pathologinnen und Pathologen. Aus der Sicht des Systemadministrators wurden wiederum andere Aspekte betrachtet. Diese Unterschiede lassen sich dadurch erklären, dass die Daten inhaltlich anders strukturiert sein müssen, wenn eine Verlaufsuntersuchung retrospektiv durchgeführt werden soll oder wenn nach Mustern und Zusammenhängen in Daten gesucht wird.

Für ein Forschungsdaten-Repository ist kennzeichnend, dass die Daten zunächst ohne konkrete Aufgabenstellung gesammelt werden. Die klinischen

und Bild-Daten werden aus unterschiedlichen Quellen, in unterschiedlichem Umfang und mit unterschiedlichem Format im Register abgelegt. Die Analyse von Datenqualität verfolgt deshalb zwei Ziele:

1. Anwender und Nutzer der Daten sollen einen Überblick erhalten, mit welchen Daten und Eigenschaften der Datensammlung sie es zu tun haben. Dazu können Parameter verwendet werden, die das Repository beschreiben.
2. Die Frage nach der Datenqualität soll vor dem Hintergrund einer spezifischen Aufgabenstellung (z.B. Studienrekrutierung) untersucht werden. Zentrale Frage dabei ist, ob die Daten hinsichtlich Inhalt und Form hinreichend gut sind, um verlässliche Aussagen zu erhalten. Dabei können für einige Parameter konkrete Anforderungen formuliert und damit auch bewertet werden.

Um ein allgemeines Verständnis von Datenqualität zu entwickeln, wurde begonnen, eine Ontologie zu entwickeln, die versucht, die unterschiedlichen Sichten in Beziehung zu bringen. Diese Ontologie wurde auf Grund der Integration von WSI in das OpEN.SC-Repository um Aspekte der Bildqualität erweitert (s. <http://bioportal.bioontology.org/ontologies/IDQA>) (s. Abb. 6).

In der Ontologie werden die Qualitätsparameter Dimensionen zugeordnet und ein Zusammenhang zum Daten-Lebenszyklus sowie zu den Akteuren im jeweiligen Anwendungskontext hergestellt. Insgesamt wurden 18 verschiedene Qualitätsdimensionen aus der Literatur identifiziert [Wand/Wang 1996, Loshin 2001, Pipino et al. 2002, Sebastian-Coleman 2013] und für das OpEN.SC-Repository spezifiziert (s. Abb. 7).

In die Ontologie wurden alle Qualitätsparameter der TMF-Leitlinie in der ersten und dann modifizierten Version aufgenommen. Jeder Parameter wurde auf eine Dimension der Ontologie abgebildet. In der aktuellen Version der TMF-Leitlinie werden alle dort genannten Qualitätsparameter drei Ebenen zugeordnet: Integrität, Organisation und Richtigkeit.

In den Tabellen 11 bis 13 werden aus der Leitlinie die für Forschungsdaten-Repositories interessanten Qualitätsparameter herausgegriffen, in die Bild- und Daten-Qualitätsontologie eingeordnet und deren Bedeutung diskutiert. Die

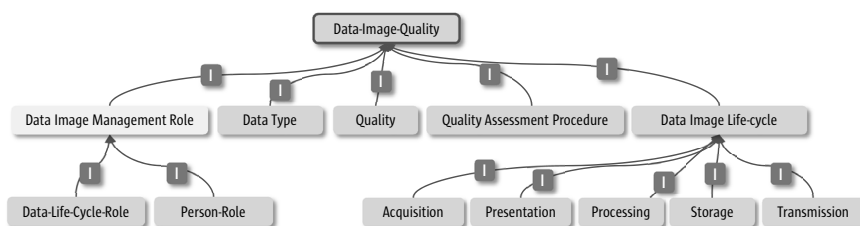


Abb. 6 Übersicht Daten- und Bildqualitätsontologie

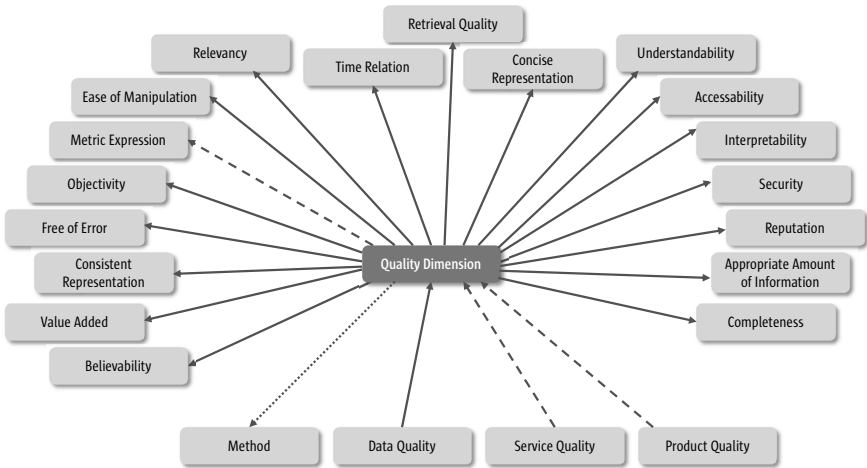


Abb. 7 Dimensionen von Datenqualität in der Ontologie

hier genannten Qualitätsparameter erlauben einen wesentlichen Einblick in die Datenqualität und stellen zwischen den Datenzentren eine Vergleichbarkeit her, die in der gemeinsamen Betrachtungs- und Berechnungsweise begründet ist. Dadurch, dass die möglichen Anfragen nach Daten in Forschungsregistern sehr unterschiedlich sein können, werden weitere Parameter benötigt, die den Datenbestand beschreiben und dabei helfen, dass die Nutzer die Qualität auch bewerten können.

Tab. 11 Qualitätsindikatoren der Ebene Integrität für ein Forschungsdaten-Repository

TMF-ID	Name des Parameters	Einordnung in die Ontologie
	Erläuterung	
TMF-1003, TMF-1004, TMF-1005	Widerspruchsfreiheit, sicherer Widerspruch, möglicher Widerspruch	Consistent Representation
TMF-1007	Bevorzugung bestimmter Endziffern	Believability, Interpretability
	Das Auftreten von bestimmten Endziffern schränkt sowohl die Glaubhaftigkeit der Daten als auch deren Interpretation ein. In dem Forschungsdaten-Repository kann nicht geklärt werden, warum die Ziffern vergeben wurden. Aus den Häufigkeiten kann lediglich eine Rückkopplung an die Klinik erfolgen, dieses Problem eingehender zu beobachten und ggf. die Datenqualität zu verbessern.	
TMF-1035	Endlosüberlebende	Completeness, Consistent Representation
	Die Anzahl der Endlosüberlebenden lässt sich relativ schnell ermitteln. Bei einer hohen Anzahl der Endlosüberlebenden in einem Datensatz ist nicht gleich von einer schlechten Datenqualität auszugehen. Es lässt lediglich die Aussage zu, dass die Sterbedaten in den Datenbestand nicht oder nur wenig nachgetragen werden. Bei Daten aus einer elektronischen Krankenakte ist das eher die Regel.	

TMF-ID	Name des Parameters	Einordnung in die Ontologie
	Erläuterung	
TMF-1013	Anteil fehlender Werte bei Datenelementen	Completeness
	Dieser Wert kann für eine Datenquelle erhoben werden, der dann weiter in mandatorisch (TMF-1014) und optional (TMF-1015) differenziert wird. Die Bewertung dieses Parameters muss allerdings differenziert erfolgen. Da es sich um klinische Daten aus dem normalen Behandlungsprozess handelt, bestehen die Daten aus einer Datenquelle auch aus unterschiedlichen Diagnosen und Diagnosegruppen. Jede Erkrankung hat ihr spezifisches Muster an erhobenen Daten. Fehlende Werte in Datenelementen können auch bedeuten, dass diese Werte für Patienten oder Erkrankungen nicht relevant waren und deshalb nicht erhoben wurden, obwohl das Datenelement an das Forschungsregister mit übertragen wurde. Hier muss im Einzelfall entschieden werden, ob der Wert fehlt obwohl dieser prinzipiell vorhanden und/oder für die Erkrankung interessant war.	
TMF-1014	Anzahl fehlender Werte bei mandatorischen Datenelementen	Completeness
	Die Menge der mandatorischen Datenelemente ist im OpEN.SC-Register relativ klein und bezieht sich im Wesentlichen auf die Stammdaten wie Alter, Geschlecht und Diagnosen.	
TMF-1015	Anzahl fehlender Werte bei optionalen Datenelementen	Completeness
	Siehe auch Kommentar zu TMF-1013.	
TMF-1016	Anteil von Datenelementen mit dem Wert unbekannt o.ä.	Believability, Interpretability
	Dieser Parameter ist verwandt mit den Parametern TMF-1007 und TMF-1025.	
TMF-1017	Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten	Completeness
	Dieser Parameter ist wichtig für die Beurteilung, in wie weit in allen Datensätzen aus den unterschiedlichen Quellen gesucht werden kann. Im Forschungsregister von OpEN.SC wurden mit den einzelnen Datenquellen separate Vereinbarungen getroffen, welche Daten zum Register übertragen wurden. Für jede Quelle können also unterschiedliche Datenelemente vorliegen. Die Frage ist nun, mit welchem Datenelement erfolgt eine Suche in allen Daten, weil dieses Datenelement in jeder Quelle zumindest geführt wurde. Ob das jeweilige Datenelement auch einen Wert hat, ist damit noch nicht beantwortet und kann nur über die Erhebung von TMF-1013 bis TMF-1015 erfolgen.	
TMF-1018	Ausreißer in stetigen Datenelementen	Believability
	Prinzipiell ist eine Analyse von Daten nach Ausreißern sinnvoll. Sie ist allerdings an die Bedingung geknüpft, ob für den betrachteten (Labor-)Wert auch Metainformationen über das Untersuchungsverfahren vorliegen. Nur mittels dieser Metadaten kann ein geeigneter Grenzwert für Ausreißer festgelegt werden. Deshalb ist das Vorliegen von Metadaten zu Laborwerten essentiell (siehe TMF-1050).	
TMF-1019	Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten	Believability
	Dieser Qualitätsparameter verhält sich analog zu TMF-1018. Um für die Qualitätsanalyse Grenzwerte festzulegen, muss das Messverfahren genau bekannt sein.	

TMF-ID	Name des Parameters	Einordnung in die Ontologie
	Erläuterung	
TMF-1020	Werte aus Standards	Understandability
	Die Verwendung von Standardterminologien und Begriffen gemäß internationaler Klassifikationen erhöht das Verstehen z.B. von Diagnosen. Berücksichtigt werden muss hierbei, das für bestimmte medizinische Domänen einige Klassifikationen zu wenig differenziert den Sachverhalt wiedergeben. Zum Beispiel: Diagnostische Begriffe der Pathologie lassen sich sehr schlecht mit den Begriffen der ICD-10 abbilden, die sich eher an klinischen Entitäten orientiert. Umso wichtiger ist es dann, dass die Begriffe aus der Pathologie gemäß einer vereinbarten Terminologie verwendet werden z.B. SNOMED CT. Problematisch wird es insbesondere dann, wenn Spezialbegriffe verwendet werden, die jeweils nur in einer Datenquelle auftreten.	
TMF-1021	Unerlaubte Werte bei qualitativen Datenelementen	Interpretability
TMF-1050	Anteil der von Untersuchungen übermittelten Metadaten	Completeness, Interpretability
	Es ist davon auszugehen, dass in einem Forschungsdatenregister Laborwerte aus unterschiedlichen Laboren mit verschiedenen Messverfahren und -bedingungen vorliegen. Diese Laborwerte können nur dann interpretiert und verglichen werden, wenn die Informationen über die Verfahren vorliegen und darüber eine Vergleichbarkeit wenn möglich hergestellt werden kann. Die Bedeutung von Metadaten für Untersuchungsverfahren und -bedingungen kann hier nur unterstrichen werden und gilt auch für alle bildgebenden Verfahren einschließlich der virtuellen Mikroskopie. Eine differenzierte, vergleichende Untersuchung von Laborwerten und Bildern über verschiedene Institutionen hinweg kann nur erfolgen, wenn sichergestellt ist, dass die Werte oder Bilder auch wirklich vergleichbar sind. Deshalb wurde dieser Qualitätsparameter in der Ontologie auch unter Completeness eingeordnet. Für jeden Laborwert bzw. auch für jede Bildkategorie kann genau festgelegt werden, welche Metadaten vorhanden sein müssen, damit die Interpretierbarkeit sichergestellt werden kann. Die Verwendung von LOINC-Kodes ist dabei sehr sinnvoll.	

Tab. 12 Qualitätsindikatoren der Ebene Organisation für ein Forschungsdaten-Repository

TMF-ID	Name des Parameters	Einordnung in die Ontologie
	Erläuterung	
TMF-1028	Aktualität der gespeicherten Daten	Timerelation/Timelines
	Im OpEN.SC Forschungsdaten-Repository wird dieser Parameter anders angegeben als in der Leitlinie: Das Forschungsdaten-Repository hat einen regelmäßigen Update-Zyklus (z.B. wöchentlich). Gemäß dieser Frequenz liegen die aktuellen Werte im Register vor. Der Bezug auf einzelne Werte mit Zeitdifferenz und akzeptabler Zeitdifferenz ist so in einem Forschungsdatensrepositorium nicht umzusetzen. In der Literatur wird die Verwendung der Begriffe Timelines und Currency diskutiert [Wand/Wang 1996], wobei hier unterschieden wird zwischen der Aktualität der Daten selbst und der Aktualisierung eines Datensatzes. In einem Forschungsregister kann lediglich zur Häufigkeit der Updates aus den Datenquellen Stellung genommen werden. Die Bewertung dieses Parameters ist dann sehr davon abhängig, welche Anforderungen die Nutzerin der Daten hat.	



TMF-ID	Name des Parameters	Einordnung in die Ontologie
	Erläuterung	
TMF-1030	Rekrutierungsrate	Completeness/Population completeness
	Dieser Parameter ist der Qualitätsdimension Vollständigkeit zugeordnet, wobei hier der Aspekt der vollständigen Erfassung der gewünschten Personen im Vordergrund steht. Das kann in einem Forschungsdaten-Repository bei der Festlegung der übertragenen Daten und Metadaten bestimmt werden. Später lässt sich dieser Parameter aus dem Datenbestand selbst nicht mehr bestimmen, da keine Relation zur Grundgesamtheit hergestellt werden kann.	
TMF-1036	Anzahl Synonyme	Understandability
	Aus Sicht eines Forschungsdaten-Repositories ist es wichtig, die Anzahl möglicher Synonyme bzw. die Synonyme für einen Begriff selbst zu kennen. Eine hohe Anzahl von Synonymen spricht nicht zwangsläufig für eine schlechte Datenqualität. Verwendete, aber nicht bekannte Synonyme erschweren das Verständnis eines Datensatzes. Wenn Synonyme in Klassifikationen und Terminologien benannt sind und entsprechend eingeordnet werden, sind Synonyme unproblematisch. Deshalb ist hier auch der Qualitätsparameter TMF-1020 (Werte aus Standards) interessant. Die Bewertung von Synonymen in einem Datenbestand ist nicht einfach: Das Management von Synonymen ist häufig eher ein technisches Problem und spielt in der Lebenszyklus-Phase der Datenverarbeitung eine Rolle. Die Verarbeitung von Synonymen ist dann kompliziert, wenn sie sich nicht eindeutig auf einen Begriff zuordnen lassen. Somit ist die Anzahl der nicht eindeutig referenzierbaren Synonyme ein interessanter Qualitätsparameter (klassisches Beispiel: HWI als Hinterwandinfarkt und als Harnwegsinfekt), weil dieser auch die Retrievalqualität stark beeinflusst.	

Tab. 13 Qualitätsindikatoren der Ebene Richtigkeit für ein Forschungsdaten-Repository

TMF-ID	Name des Parameters	Einordnung in die Ontologie
	Erläuterung	
TMF-1043	Genauigkeit der Registerdaten	Free of Error
	Im OpEN.SC Forschungsdaten-Repository ist diese Frage insbesondere wichtig für Diagnosen. Problematisch dabei aber ist, dass Diagnosen stets Abstraktionsvorgänge sind, die von Ärztin zu Ärztin unterschiedlich sein können. Die Frage nach der wirklichen Diagnose lässt sich nicht beantworten. Hier hat sich das Vorgehen bewährt, dass zu Diagnosen und Fällen insgesamt eine Zweitmeinung eingeholt wird. Damit kann ein Maß angegeben werden, in wie weit Erstmeinung und Zweitmeinung übereinstimmen. Aus diesem Vorgehen heraus lassen sich weitere Maße ableiten: Anzahl der Fälle, für die eine Referenzmeinung eingeholt wurde, Anzahl der übereinstimmenden Fälle, Anzahl der Abweichungen, Anzahl der Fälle mit groben Abweichungen u.a.	

TMF-ID	Name des Parameters	Einordnung in die Ontologie
	Erläuterung	
TMF-1047	Übereinstimmung mit Verfahrensregeln	Reputation
	In einem Forschungsdaten-Repository erfolgt die Auswertung retrograd. Da die Daten aus verschiedenen Quellen kommen und auch der Nutzungskontext offen ist, stellt sich nach der Suche von Fällen die Frage, ob die Qualität der Daten dieser Fälle geeignet ist, um eine Studie durchzuführen (siehe Anwendungsbeispiel). Dieses Maß ist aus Sicht der Forschungsregister ein zusammengesetztes Maß, weil die Verfahrensregeln sehr unterschiedlich sein können. Sie können die Vollständigkeit aller Dateneinträge ebenso fordern (siehe TMF-1013 bis TMF-1015) wie die Kontrolle der Daten und/oder Diagnosen von einer beauftragten Stelle. Die Erfüllung dieser Kriterien ist auf jeden Fall ein gutes Zeichen für Datenqualität und erhöht deren Reputation.	

### VI.4.4 Beispielhafte Implementierung von Qualitätsindikatoren im OpEN.SC-Forschungsdaten-Repository

#### VI.4.4.1 Einführung

Im Rahmen des OpEN.SC-Projektes wurden verschiedene Aspekte der Datenqualität in eigenständigen Forschungsvorhaben untersucht und einige Indikatoren evaluiert. Zunächst wurden die Metadaten von radiologischen Bildern analysiert, die stellvertretend für Labordaten betrachtet werden können (TMF Qualitätsparameter TMF-1050: Anteil der von Untersuchungen übermittelten Metadaten). Für die Entwicklung eines Studyadvisors wurden das Qualitätskriterium TMF-1020: Werte aus Standards eingehender untersucht [Streit et al. 2012; Streit 2013].

#### VI.4.4.2 Die Anwendung des Qualitätsfaktors: Umfang der Metadaten bei Untersuchungen

Metadaten enthalten Informationen über die Daten selbst. Bei Laboruntersuchungen sind dem reinen Laborwert selbst weitere Information zugeordnet wie zum Beispiel:

- Welche Labormethode wurde zur Bestimmung des Wertes herangezogen?
- Was sind die Grenzwerte für diese Laboruntersuchung?<sup>7</sup>
- Unter welchen Bedingungen wurde die Untersuchung durchgeführt, z.B. nüchtern oder nach einer Mahlzeit?

<sup>7</sup> Bei einigen Laborverfahren werden von Labor zu Labor unterschiedliche Grenzwerte angegeben. Zur Beurteilung eines Wertes ist es allerdings notwendig, die Grenzen zu kennen.

Diese Informationen können mittels eines Codes der Logical Observation Identifiers Names and Codes (LOINC) als Metadatum übermittelt und gespeichert werden (s. <http://loinc.org/>). Für radiologische Bilder gibt es ebenfalls Metadaten, die die Untersuchungsverfahren und -bedingungen näher beschreiben, wenn die Bilder im DICOM-Format abgespeichert werden. DICOM kann alle notwendigen Metainformationen zu einem Bild aufnehmen und für die spätere Wiederverwendung bereitstellen (s. <http://medical.nema.org/>).

In 3.461 anonymisierten DICOM-Bildern aus unterschiedlichen Einrichtungen und von verschiedenen Geräten wurde untersucht, wie viele Metadaten-Items für jedes Bild abgespeichert wurden. Weiterhin wurde dann bestimmt, wie viele Metadaten-Items von allen Bildern verwendet wurden. Die Anzahl gemeinsamer Elemente ist ein wichtiger Indikator dafür, ob eine Suchanfrage auf diesem Bilddatensatz auch wirklich alle Bilder erreicht. In den untersuchten Bildern wurden zwischen 60 und 205 Metadaten-Items gefunden. Am häufigsten wurden die Bilder mit 96 Metadaten versehen (s. Abb. 8).

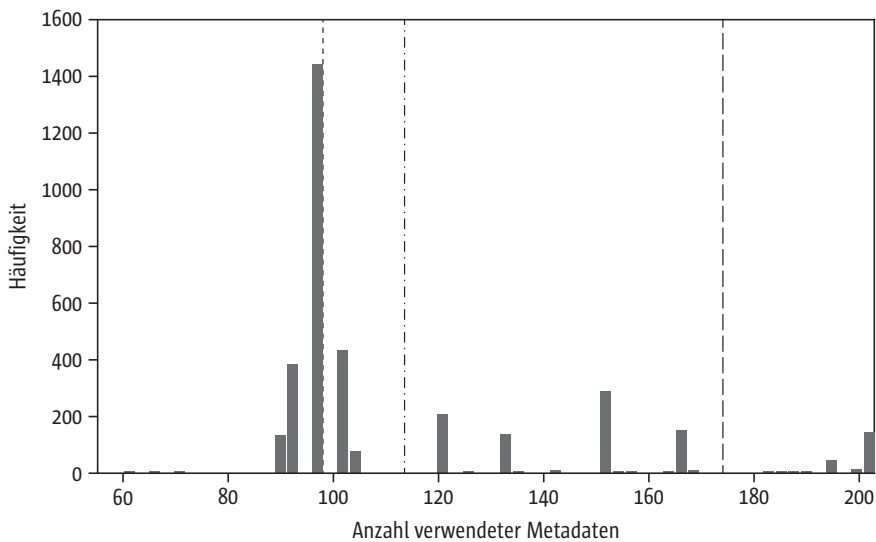


Abb. 8 Verteilung der Anzahl von verwendeten Metadaten in den radiologischen Bildern

Die Anzahl der gemeinsam verwendeten Metadaten war sehr gering: nur 5 Metadaten wurden von allen Bildern verwendet und erlauben damit eine Suche in allen Bildern. Mit nur 15 Metadaten werden 50% der Bilder in einer Suchanfrage erreicht (s. Abb. 9). Mit dieser Untersuchung konnte gezeigt werden, dass selbst die Verwendung eines Standards (hier DICOM) nicht sicherstellt, dass alle Elemente in einem Suchraum abgefragt werden können.

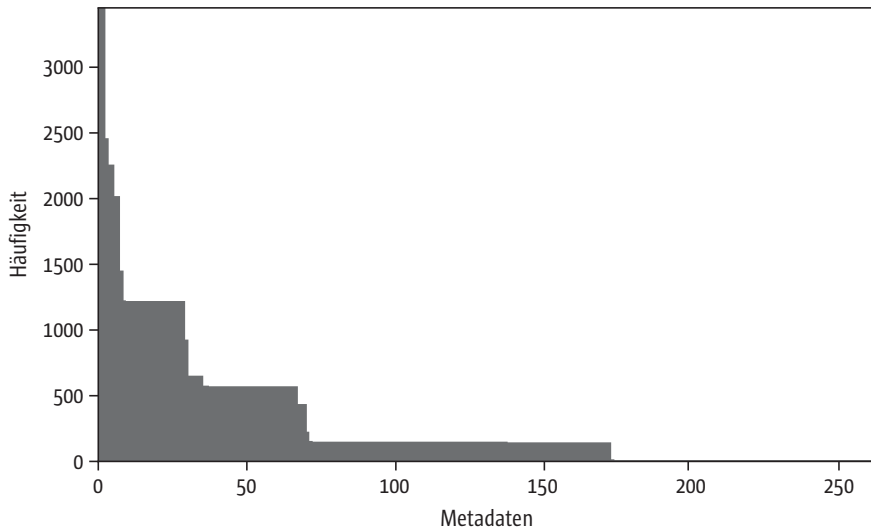


Abb. 9 Häufigkeit gemeinsam verwendeter Metadaten-Items

### VI.4.4.3 Der StudyAdvisor der OpEN.SC-Plattform

Die Durchführung einer empirischen Studie stellt einen komplexen Prozess dar. Das Forschungsdaten-Register des OpEN.SC bietet die Möglichkeit, retrospektiv Untersuchungen an den Daten durchzuführen. Allerdings sind für diese Studien wichtige Kriterien der Datenqualität zu erfüllen.

Der Prozess der Vorbereitung einer retrospektiven Studie schließt die Suche nach geeigneten Patienten und Patientinnen an Hand von Einschluss- und Ausschlusskriterien ein. Am häufigsten wird nach einer bestimmten Diagnose bzw. einer Diagnosekombination, Alter und Geschlecht gesucht. Für das Finden geeigneter Fälle ist es daher sehr wichtig, dass entweder die Diagnosen einen einheitlichen Schlüssel verwenden, z.B. die ICD-10, oder die diagnostischen Begriffe des ICD-10. Werden Synonyma verwendet und lassen sich diese nicht auf die terminologischen Begriffe abbilden, werden diese Fälle nicht gefunden oder/und falsch zugeordnet.

Der StudyAdvisor des OpEN.SC-Projektes ist ein Werkzeug zur Unterstützung der Durchführung von Studien, welches eine Analyse der Datenqualität einschließt (s. Abb. 10). Neben dem Kriterium der Verwendung von Standardbegriffen, die die Interpretierbarkeit von Befunden beeinflussen, wurden Aktualität der Daten und zeitliche Häufigkeit von Untersuchungen (Timedistance aus der Dimension Timerelation) bestimmt. Das Tool integriert sich in die SOA des OpEN.SC-Systems. Über eine Zwischenschicht Hibernate wird auf die OpEN.SC-Datenbank zugegriffen. Die Analyse selbst wird von verschiedenen Modulen ausgeführt (Patientenstammdatenanalyse, Diagnoseanalyse, Laboruntersuchungsanalyse). Die Analyse erfolgt via Zugriff auf die Ontologie zur Bild- und Datenqualität.

Im untersuchten Datenpool von ca. 10.000 Patienten und Patientinnen wurden 90.000 ICD-Kodes in die entsprechenden Felder eingetragen. Demgegenüber stehen aber 110.000 diagnostische Bezeichner. In der weiteren Analyse wurden folgende Probleme gefunden:

- In den ICD-Kodes wurden häufig Xxx.9 („nicht näher bezeichnet“) gefunden, was eine ungenaue diagnostische Bezeichnung darstellt und die Interpretierbarkeit deutlich herabsetzt (siehe auch TMF-1007, TMF-1016, TMF-1025). Vor allem aber wird die Glaubwürdigkeit der Daten herabgesetzt, weil dieser Parameter darauf hindeutet, dass die Kodierqualität eingeschränkt ist.
- Beispielsweise wurden bei einem Patienten neben der Hauptdiagnose einer chronischen Pyelonephritis folgende ICD-10 Kodes vergeben:
  - J06.9 – Begriff des Standards: akute Infektion der oberen Atemwege, nicht näher bezeichnet. Bei diesem Patienten wurde stets ein „Infekt, grippal“ angegeben.
  - T86.9 – Begriff des Standards: Versagen und Abstoßung/Rejektion eines nicht näher bezeichneten transplantierten Organes oder Gewebes. Hier wurden verschiedene Bezeichner angegeben: Transplantatruptur, Rejektion oder Rejektion (Banff IIb).
  - B18.2 – Begriff des Standards: Chronische Virushepatitis. Hier wurde immer „Hepatitis C, chronisch“ als nachvollziehbares Synonym angegeben.

Für die Durchführung einer Studienrekrutierung sind einheitliche Bezeichner bzw. eindeutige Verschlüsselungen notwendig, um sicher potenzielle Kandidaten und Kandidatinnen zu identifizieren. Auch bei der Musteranalyse führt die Verwendung einer Vielzahl von verschiedenen Begriffen dazu, dass keine lokalen Häufigkeiten und Zusammenhänge detektiert werden können.

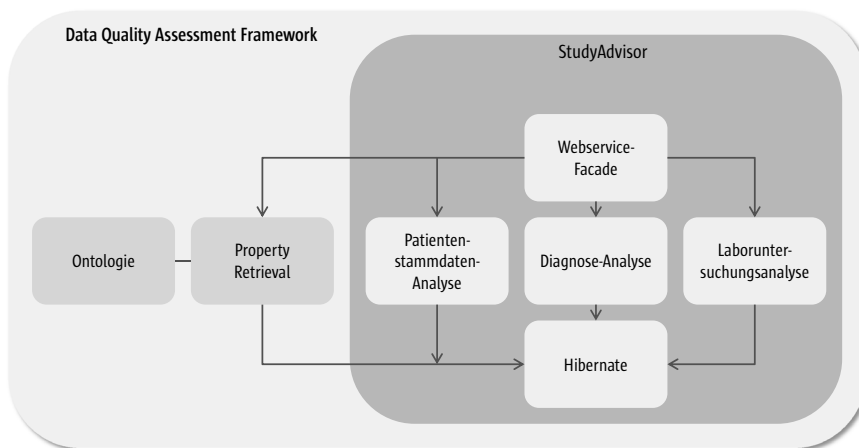


Abb. 10 Systemarchitektur des StudyAdvisors für die Analyse von Datenqualität

### VI.4.5 Zusammenfassung und Ausblick

Für die sekundäre Nutzung von klinischen Daten im Rahmen von Forschungsregistern ist die Darstellung der Datenqualität aus zwei Gründen sehr wichtig:

1. Die Datenqualitätsparameter vermitteln den Nutzern einen Eindruck darüber, mit welchen Daten und in welcher Qualität sie rechnen können. Die Bewertung darüber, ob die Daten nützlich oder für eine bestimmte Studie nicht geeignet sind, hängt sehr stark von der Aufgabenstellung ab und muss den Nutzern überlassen werden.
2. Für eine Reihe von Parametern lässt sich jedoch die Datenqualität unabhängig von deren Nutzung bewerten. So lassen sich daraus Rückschlüsse ziehen, an welchen Stellen Einfluss genommen werden kann. Dabei ist die Unterscheidung von Prozess-, Struktur- und Ergebnisqualität sehr wichtig.

Die Verantwortung für Datenqualität lässt sich nicht immer nur auf den Erzeuger der Daten beschränken. Einige Parameter werden in den anderen Phasen des Datenlebenszyklus beeinflusst. Mit der Leitlinie der TMF ist den Nutzern wie Administratoren ein wichtiges Werkzeug in Hand gegeben worden, um Aussagen über die Datenqualität zu treffen. Für die Forschungsdaten-Repositories und in den aktuellen Entwicklungen von Big Data, wo sehr große Datenmengen (im Terra-, Penta- und Exabyte-Bereich) gesammelt, analysiert und wieder verwendet werden sollen, muss davon ausgegangen werden, dass weitere Qualitätsparameter notwendig sind.