

Indikatoren zur Datenqualität in der TMF: Leitlinie Version 2

TMF-Projekt V020-04

Jürgen STAUSBERG^a, Ulrike BAUER^b, Johannes DREPPER^c, Daniel NASSEH^a, Michael
NONNEMACHER^d, Ron PRITZKULEIT^e, Carsten O. SCHMIDT^f, Thomas SCHRADER^g

^aIBE, Ludwig-Maximilians-Universität München, Germany

^bKompetenznetz Angeborene Herzfehler

^cTMF, Berlin

^dIMIBE, Universitätsklinikum Essen, Germany

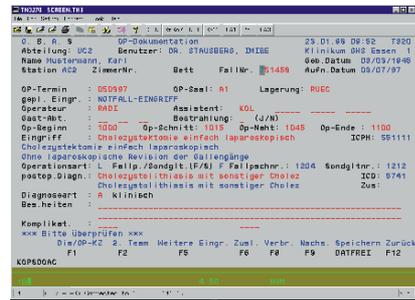
^eInstitut für Krebsepidemiologie e.V., Universität Lübeck, Germany

^fInstitut für Community Medicine, Greifswald, Germany

^gDepartment of Pathology Charite, Berlin, Germany

Kontinuierliche Verbesserung der Datenqualität

Überprüfung



Angepasste Empfehlungen

Verbesserung der Datenqualität



Durchführung: SDV, Analyse

Gliederung

- Leitlinie Version 1.0
- Vorgehen
- Veränderungen der Qualitätsindikatoren
- Zusammenfassung

Leitlinie aus 2007

- 24 Indikatoren zur Datenqualität
 - Kategorien Plausibilität (10 Indikatoren), Organisation (7), Richtigkeit (7)
- Definition eines Scores als Globalindikator
- Empfehlungen zur Anpassung von Maßnahmen des Qualitätsmanagements an die gemessene Datenqualität
 - Fallzahlkalkulation für einen Originaldatenabgleich
 - Art des Feedback zur Datenqualität
- Prozessmodell

M. Nonnemacher
D. Weiland | J. Stausberg



Datenqualität in der medizinischen Forschung

Leitlinie zum adaptiven Management
von Datenqualität in
Kohortenstudien und Registern

Beispiel für Datenqualitätsindikator 1

Beschreibung	Anzahl fehlender Werte bei mandatorischen Datenelementen
Begriffsdefinitionen	---
Typ	Ergebnisqualität
Quelle	Aronsky und Haug 2000, Gaus 2003, ... [weitere Quellen]
Bezug	Die Berechnung des Indikators ist nur für <u>mandatorische</u> Datenelemente sinnvoll.
Alternative Definition	---
Anmerkungen	<p>Der Indikator ist wichtig für die korrekte Schätzung von Inzidenzen und Prävalenzen. Der Anteil der Fälle mit vollständigen Daten ist für spätere multivariate Analysen von großer Bedeutung.</p> <p>Ein entsprechender Indikator ist auch für optionale Datenelemente vorhanden. Die separate Behandlung von mandatorischen und optionalen Datenelementen wird von Aronsky und Haug vorgeschlagen, da jeweils unterschiedliche Fehlerraten tolerabel sind (Aronsky und Haug 2000).</p>

Beispiel für Datenqualitätsindikator 2

Zähler	Anzahl fehlender Werte bei mandatorischen Datenelementen
Nenner	Anzahl überprüfter Werte bei mandatorischen Datenelementen
Subkategorien	---
Berechnung	1) Suche nach fehlenden Werten bei mandatorischen Datenelementen 2) Berechnung von Zähler, Nenner und Rate Alternative Berechnungsmöglichkeit: Anzahl fehlender Werte pro Patient/Person
Interpretation	Je höher die Rate, desto schlechter ist die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.
Einflussfaktoren	Einführung von separaten Kategorien für „not applicable“, „not done“

Indikatoren für Datenqualität 1

Qualitätsindikator	Schwellenwert für Auffälligkeit	Gewicht
Übereinstimmung mit Vorwert	> 5 %	1
Konkordanz	> 5 %	1
Widerspruchsfreiheit	> 5 %	3
Werteverteilung innerhalb und zwischen Zentren	> 10 %	1
Untersuchungen am Wochenende oder an gesetzlichen Feiertagen	> 2 %	1
Anzahl fehlender Werte bei optionalen Datenelementen	> 5 %	1
Ausreißer bei stetigen Datenelementen	> 10 %	2
Erlaubte Werte bei qualitativen Datenelementen	> 0 %	1
Bevorzugung bestimmter Endziffern	> 5 %	1
Nachweis bekannter Korrelationen	< 10 %	1

Berechnung des Qualitätsscores

- Für jeden Qualitätsindikator
 - Schwellenwert für auffällig/unauffällig
 - Relatives Gewicht
- Berechnung des Qualitätsscores
 - Wert 1 für jeden unauffälligen, Wert 0 für jeden auffälligen Indikator
 - Multiplikation des Wertes mit dem relativen Gewicht
 - Addition der Werte für alle bestimmmbaren Indikatoren
 - Transformation auf eine Skala von 0 bis 100
 - Bestimmung der Datenqualitätskategorie
 - 0-20 sehr schlecht
 - 21-40 schlecht
 - 41-60 moderat
 - 61-80 gut
 - 81-100 sehr gut

Empfehlungen

- Source Data Verification (SDV)
 - Fallzahlschätzung für Anzahl der Patienten
 - Fallzahlschätzung für Anzahl der Datenelemente
 - Häufigkeit der Besuche in den Studienzentren
- Feedback über Datenqualität (je Zentrum)
 - Eigene Ergebnisse, oder
 - Eigene Ergebnisse + anonymisierte Vergleichs-daten der anderen Zentren, oder
 - Eigene Ergebnisse + offene Vergleichsdaten der anderen Zentren
- Zeitpunkt der nächsten Leitlinienanwendung

Framework of Procedures for the Assurance of Data Quality in Medical Registries

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 itemsdefine data & data characteristics in data dictionarydraft a data collection protocoldefine pitfalls in data collectioncompose data checkscreate user friendly case record formscreate quality assurance plan	At the onset of participating in the registry <ul style="list-style-type: none">assign a contact personcheck developed software for data entry and for extractioncheck reliability and completeness of extraction sourcesstandardise correction of data items
In case of new participating sites <ul style="list-style-type: none">perform site visittrain new participants	Continuously <ul style="list-style-type: none">train (new) data collectorsmotivate data collectorsmake data definitions availableplace date & initials on completed formskeep completed case record formsdata collection close to the source and as soon as possibleuse the registry data for local purposes
Continuously <ul style="list-style-type: none">motivate participantscommunicate with local sites	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
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	
Detection during data collection	
During import of data into the central database <ul style="list-style-type: none">perform automatic data checks	Continuously <ul style="list-style-type: none">visually inspect completed formsperform automatic data checkscheck completeness of registration
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 variabilityperform analyses on the data	
Actions for quality improvement	
After data import and data checks <ul style="list-style-type: none">provide local sites with data quality reportscontrol local correction of data errors	After receiving quality reports <ul style="list-style-type: none">check detected errorscorrect inaccurate data & fill in incomplete dataresolve causes of data errors
After data audit or variability test <ul style="list-style-type: none">give feedback of results and recommendationsresolve causes of data errors	After receiving feedback <ul style="list-style-type: none">implement recommended changescommunicate with personnel

Gliederung

- Leitlinie Version 1.0
- Vorgehen
- Veränderungen der Qualitätsindikatoren
- Zusammenfassung

Workshop: Messung der Qualität biomedizinischer Daten

Programm

09.00 Uhr Einführung

Prof. Dr. Jürgen Stausberg (LMU München)

09.15 Uhr Methoden

Leitlinie zum adaptiven Management von Datenqualität
In Kohortenstudien und Registern
Dr. Michael Nonnemacher (Universität Duisburg-Essen)

Optimierung der Datenqualität von Kohortenstudien
mittels Qualitätsindikatoren
Christian Jacke | Prof. Dr. Ute-Susann Albert (Brustzentrum
Regio, Universitätsklinikum Gießen und Marburg GmbH)

Messung von Datenqualität mit Kennzahlen in Open.SC.
Sabine Hanß | Prof. Dr. Thomas Schrader (Charité Berlin)

10.30 Uhr Kaffeepause

10.45 Uhr Datenqualität in der Forschung

Datenqualität in epidemiologischen Krebsregistern
Prof. Dr. Alexander Katalinic (Krebsregister Schleswig-
Holstein)

Qualitätssicherung in einer komplexen epidemiologischen
Studie am Beispiel der Study of Health in Pomerania (SHIP)
Dr. Carsten Oliver Schmidt (SHIP-Studie, Universität
Greifswald)

11.30 Uhr Pause und Imbiss

11.45 Uhr Datenqualität in der Versorgung

Qualität von Routinedaten in der stationären Versorgung
Dr. Jan Helling (DRG-Research-Group, Universitätsklinikum
Münster)

Überprüfung der Datenqualität in der externen stationären
Qualitätssicherung
Yvonne Ziert | PD Dr. Günther Heller (AQUA-Institut
für angewandte Qualitätsförderung und Forschung im
Gesundheitswesen GmbH)

12.30 Uhr Abschlussdiskussion

Moderation: Prof. Dr. Jürgen Stausberg

13.00 Uhr Ende des Workshops

Organisatorisches

Veranstaltungstermin:

Montag, 4. April 2011
9.00 bis 13.00 Uhr

Ort:

Veranstaltungsräume der TMF
Georgenstraße 22 | 10117 Berlin

Anreise mit der Bahn:

Vom Hauptbahnhof mit der S-Bahn eine Station
bis zum S-Bahnhof Friedrichstraße, der Eingang
zu den Räumen in der Georgenstr. 22 liegt gleich
gegenüber dem Bahnhof

Wissenschaftliche Leitung:

Prof. Dr. Jürgen Stausberg
(Hep-Net | LMU München)

Organisation:

Geschäftsstelle TMF e.V.
Juliane Gehrke
Tel.: 030 - 31 01 19 78
E-Mail: info@tmf-ev.de

Anmeldung und Teilnahmegebühr:

Die Teilnahme an dem Workshop ist kostenfrei.

Die Anmeldung erfolgt online unter:

www.tmf-ev.de/anmelden

Bottom-up-Ansatz

- Eingeladene Best Practice-Beispiele
 - Kohorte: Study of Health in Pomerania (SHIP), Greifswald
 - Datenrepository: Open European Nephrology Science Center (OpEN.SC), Berlin
 - Krebsregister: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID), vertreten durch das Institut für Krebs Epidemiologie, Lübeck
- Abbildung der Kennzahlen der Best Practice-Beispiele aus den entsprechenden SOPs auf die Indikatoren der Leitlinie
- Identifikation von Lücken: Aufnahme von Kategorien, Ergänzung von Kennzahlen, Verfeinerung von Kennzahlen, etc.

Anpassungen in der Darstellung

- Empfehlungen zu Schwellenwerten
 - Empfehlungen zu Gewichten (Beispiele)
 - Implementierungshilfen
 - Software
-
- + Eindeutiges Identifikationskennzeichen
 - + Anwendungsempfehlungen

Konzeptionelle Anpassungen

- Umbenennung der Kategorie Plausibilität in Integrität
- Aufnahme eines Indikators zu Metadaten
- Einführung von Modulen mit Gewichten für bestimmte Vorhabenstypen
 - Register, Kohorten, Datenrepositories
- Berechnung der Kennzahlen für verschiedene Beobachtungseinheiten: Datenbestand, Person, Meldung
- Aufnahme Domain-spezifischer Indikatoren wie die DCO-Rate bei Krebsregistern

Gliederung

- Leitlinie Version 1.0
- Vorgehen
- Veränderungen der Qualitätsindikatoren
- Zusammenfassung

Erweiterung des Indikatorensets

Kategorie	Version 2007	Revision
Integrität	10	29
Organisation	7	15
Richtigkeit	7	6
Metadaten		1
Summe	24	51

Beispiele

- Integrität
 - neu: Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten
- Organisation
 - neu: Verweigerungsrate von Untersuchungen
- neu: Metadaten
 - neu: Umfang der Metadaten bei Untersuchungen

Überarbeitung: Ebene Integrität 1

Qualitätsindikator	ID alt	ID neu	Datensatz/ Meldung	Beobachtungs- einheit	Daten- bestand
Ebene Integrität¹					
Übereinstimmung mit Vorwert	B.2.1.1	TMF-1001		x	x
Konkordanz	B.2.1.2	TMF-1002			x
Widerspruchsfreiheit	B.2.1.3	TMF-1003	x	x	x
Endlosüberlebende [Krebsregister]		TMF-1035			x
Sichere Widersprüche/Fehler [OpEN.SC, Krebsregister]		TMF-1004	x	x	x
Mögliche Widersprüche/Warnungen [OpEN.SC, Krebsregister]		TMF-1005	x	x	x
Werteverteilung	B.2.1.4	TMF-1006		x	x
Bevorzugung bestimmter Endziffern	B.2.1.9	TMF-1007			x
Werteverteilung der durch Untersucher erfassten Parameter [SHIP]		TMF-1009			x
Werteverteilung der durch Geräte erfassten Parameter [SHIP]		TMF-1010			x
Werteverteilung von Befunden [SHIP]		TMF-1011			x
Werteverteilung von Parametern zwischen Zentren		TMF-1052			x
Untersuchungen am Wochenende	B.2.1.5	TMF-1008	x	x	x
Fehlende Einträge					
Fehlende Module [SHIP]		TMF-1012		x	x
Fehlende Werte bei Datenelementen [Stausberg]		TMF-1013	x	x	x
Fehlende Werte bei mandatorischen Datenelementen	B.2.3.6	TMF-1014	x	x	x
Fehlende Werte bei optionalen Datenelementen	B.2.1.6	TMF-1015	x	x	x

Überarbeitung: Ebene Integrität 2

Datenelemente mit Wert unbekannt o. ä. [SHIP]		TMF-1016	x	x	x
Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten [OpEN.SC]		TMF-1017			x
Ausreißer bei stetigen Datenelementen	B.2.1.7	TMF-1018	x	x	x
Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten [SHIP]		TMF-1019	x	x	x
Werte aus Standards [OpEN.SC]		TMF-1020			x
Unerlaubte Werte [Oberbegriff, SHIP]					
Unerlaubte Werte bei qualitativen Datenelementen ¹	B.2.1.8	TMF-1021	x	x	x
Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings [SHIP]		TMF-1022	x	x	x
Unerlaubte Werte zur Kodierung von fehlenden Modulen [SHIP]		TMF-1023	x	x	x
Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen [SHIP]		TMF-1024	x	x	x
Datenelemente mit unspezifischen Werten [Krebsregister]		TMF-1025		x	x
Beobachtungseinheiten mit unbekanntem Primärtumor [Krebsregister]		TMF-1026		x	x
Nachweis bekannter Korrelationen	B.2.1.10	TMF-1027			x

Überarbeitung: Ebene Organisation

Ebene Organisation					
Aktualität der gespeicherten Daten	B.2.2.1	TMF-1028	x	x	x
Dubletten im Datenbestand ¹	B.2.2.2	TMF-1029		x	x
Rekrutierungsrate	B.2.2.3	TMF-1030			x
DCO-Rate (Death Certificate Only) [Krebsregister]		TMF-1051		x	x
Verweigerungsraten [Oberbegriff, SHIP]					
Verweigerungsrate von Untersuchungen [SHIP]		TMF-1031			x
Verweigerungsrate von Modulen [SHIP, OpEN.SC]		TMF-1032			x
Verweigerungsrate bei einzelnen Datenelementen [SHIP]		TMF-1033			x
Vorzeitig aus dem Register ausscheidende Beobachtungseinheiten	B.2.2.4	TMF-1034			x
Fachliche Qualifikation des Datenerfassungspersonals [SHIP]	B.2.2.5				
Synonyme ¹	B.2.2.6	TMF-1036			x
Homonyme ¹	B.2.2.7	TMF-1037			x
Einzelmeldungen pro Beobachtungseinheit [Krebsregister]		TMF-1038		x	x
Solitäre Meldungen von Pathologen [Krebsregister]		TMF-1039			x
Zurückgewiesene Meldungen [Krebsregister]		TMF-1040			x
Datenquellen pro Beobachtungseinheit [Krebsregister]		TMF-1041		x	x
Beobachtungseinheiten mit Follow-up [Krebsregister]		TMF-1042			x

Überarbeitung: Ebene Richtigkeit

Ebene Richtigkeit					
Genauigkeit der Registerdaten	B.2.3.1	TMF-1043	x	x	x
Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente	B.2.3.2	TMF-1044			x
Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten	B.2.3.3	TMF-1045			x
Vollständigkeit der Registerdaten	B.2.3.4	TMF-1046	x	x	x
Übereinstimmung mit Verfahrensregeln ¹	B.2.3.5	TMF-1047			x
Repräsentativität der Registerdaten	B.2.3.7	TMF-1048			x

Überarbeitung: Ebene Metadaten

Ebene Metadaten					
Umfang der Metadaten bei Untersuchungen [Open.SC]		TMF-1050	x		x

Gliederung

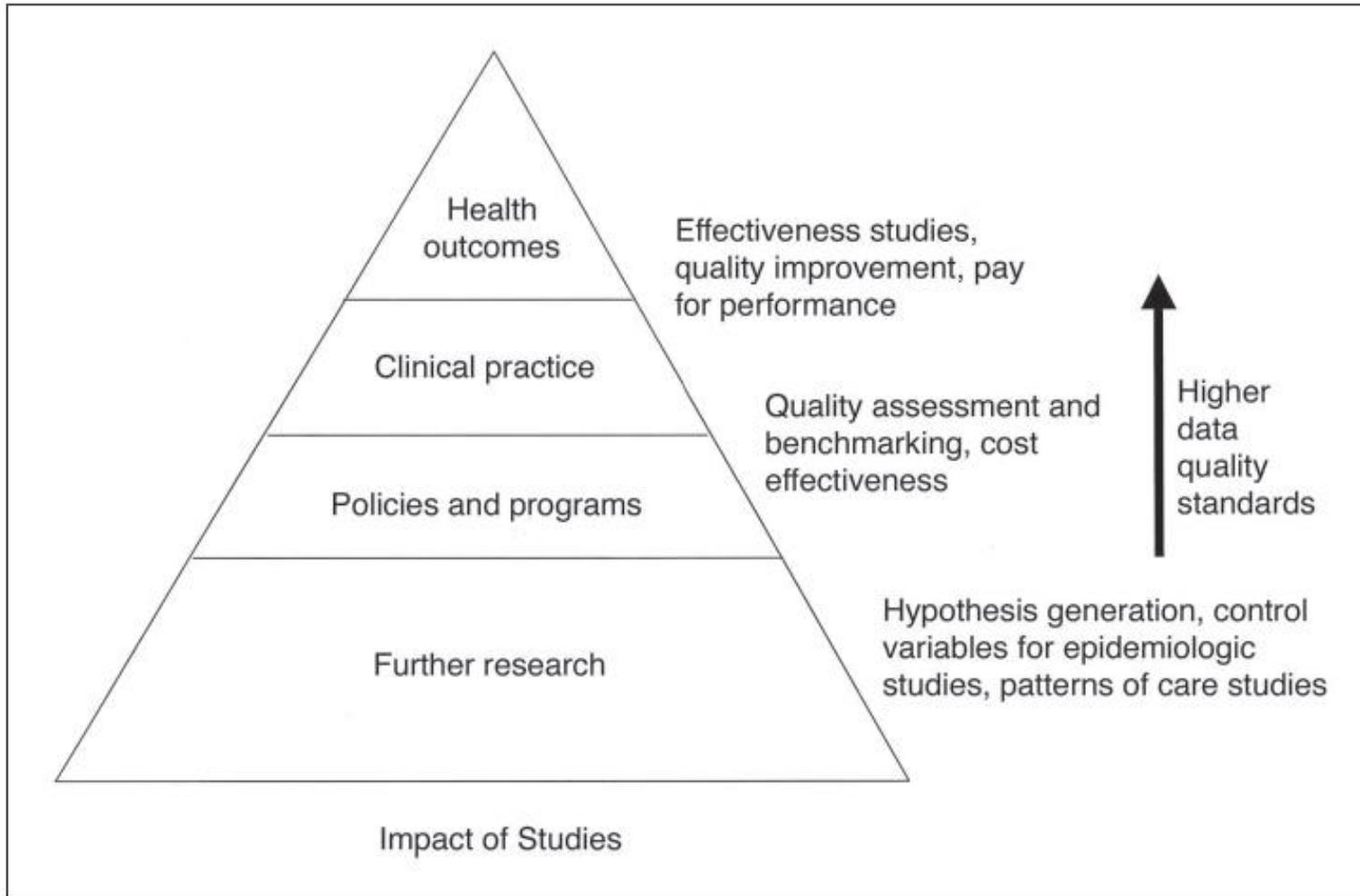
- Leitlinie Version 1.0
- Vorgehen
- Veränderungen der Qualitätsindikatoren
- Zusammenfassung

Leitlinie Version 2.0

- Umfassendes Repository von Indikatoren
 - Berücksichtigung spezifischer Indikatoren der Krebsregister
- Spezifische Anwendungsempfehlungen
 - U. Bauer: Register
 - C. Schmidt: Kohorten
 - T. Schrader: Data Repositories
- Zuwachs von 120 auf 160 Seiten
- Aktuelle Literaturübersicht
 - Breite Literaturübersicht bis 2005
 - Themenspezifische Literaturübersicht 2005 bis 2013

➤ Drucklegung in Kürze

Ausblick



Vielen Dank für Ihre Aufmerksamkeit!