

Projekt zur

Entwicklung, Umsetzung und Evaluation von Leitlinien
zum adaptiven Management von Datenqualität in
Kohortenstudien und Registern

gefördert durch die

Telematikplattform für
Medizinische Forschungsnetze e. V.



Indikatoren von Datenqualität

Michael Nonnemacher

Institut für Medizinische Informatik, Biometrie und Epidemiologie
am Universitätsklinikum Essen



Ein **Qualitätsindikator** ist ein quantitatives Maß, das die Qualität einer Einheit durch Zahlen bzw. Zahlenverhältnisse indirekt abbildet. Er ist kein direktes Maß der Qualität, sondern mehr ein Werkzeug zur Leistungsbewertung und dient hauptsächlich der gezielten Lenkung qualitätssichernder Aktivitäten auf auffällige Bereiche.

GMDS-Arbeitsgruppe "Qualitätsmanagement in der Medizin".
Begriffe und Konzepte des Qualitätsmanagements. Informatik,
Biometrie und Epidemiologie in Medizin und Biologie 2003;34:1-61.



Aufbau der Indikatordefinition 1



Angelehnt an Empfehlungen der Joint Commission on Accreditation of Healthcare Organizations für evidenz-basierte klinische Qualitätsindikatoren im Gesundheitswesen

Joint Commission on Accreditation of Healthcare Organizations. Primer on Indicator Development and Application. Measuring Quality in Health Care. 1990. Illinois.

Beschreibung	z.B. <i>Anzahl fehlender Werte</i>
Begriffsdefinitionen	werden auch im Glossar aufgeführt
Typ	Strukturqualität/Prozessqualität/Ergebnisqualität
Quelle	Literaturquellen, Expertenmeinung
Bezug	Berechnung für einzelne Datenelemente oder für größere Informationseinheiten (z.B. Meldung, Register)
Alternative Definition	Hinweis auf alternative Definitionen
Anmerkungen	
Zähler	z.B. <i>Anzahl fehlender Werte des Datenelements</i>
Nenner	z.B. <i>Anzahl überprüfter Datenelemente, Anzahl aller Patienten</i>
Subkategorien	Subkategorien zur Adjustierung des Indikators, optional
Berechnung	Beschreibung der Datensammlung und –aggregation
Interpretation	Interpretation des Indikatorwertes, auch Vorschläge aus der Literatur z.B. <i>Je höher die Rate, desto schlechter die Datenqualität. Meier/Schulze 2005 halten eine Rate < 0,5 % für akzeptabel.</i>
Einflussfaktoren	Faktoren auf den Ebenen Struktur und Prozesse, die den Indikator beeinflussen können z.B. <i>Einführung von Kategorien für not applicable/not done verringert die Anzahl echter fehlender Werte</i>

Identifikation von Datenqualitätsindikatoren aus der Literatur und Zuordnung zu den Bereichen

- Plausibilität (Strukturen)
- Organisation (Prozesse)
- Richtigkeit (Ergebnisse)

- Übereinstimmung des Ergebnisses einer Messung oder Untersuchung mit dem Vorwert
 - Rate: Anzahl der Werte mit absoluter Differenz zum Vorwert $>$ Schwellenwert / Anzahl überprüfter Werte
 - Schwellenwert individuell für jedes Datenelement
- Konkordanz (Übereinstimmung des Wertes eines Datenelements mit einer Referenzquelle)
 - Referenzquelle unabhängig von Datenerhebung
 - Rate: Anzahl der Werte, die nicht mit Referenzquelle übereinstimmen / Anzahl überprüfter Werte

- Widerspruchsfreiheit
 - Festlegung von Plausibilitätsregeln
 - Rate: Anzahl der Werte, die mindestens eine Plausibilitätsregel verletzen / Anzahl überprüfter Werte
- Werteverteilungen innerhalb und zwischen Zentren
 - Berechnung statistischer Kenngrößen (Min, Max, Mittelwert etc.) zur Beschreibung der Werteverteilung
 - Rate: Anzahl auffälliger Werteverteilungen / Anzahl überprüfter Datenelemente

- Untersuchungen am Wochenende oder an gesetzlichen Feiertagen
 - Rate: Anzahl der Datumsangaben, die auf ein Wochenende oder einen gesetzlichen Feiertag fallen / Anzahl überprüfter Werte
- Fehlende Werte bei optionalen Datenelementen
 - „harmloser“ als bei mandatorischen Datenelementen
 - Rate: Anzahl fehlender Werte / Anzahl überprüfter Werte

- Ausreißer bei stetigen Datenelementen
 - Erwarteter Bereich: z.B. bestimmt durch 1,5-fachen Interquartilsabstand (Box-Whisker-Plots)
 - Rate: Anzahl der Ausreißer / Anzahl überprüfter Werte

- Erlaubte Werte bei qualitativen Datenelementen
 - Rate: Anzahl der unerlaubten Werte / Anzahl überprüfter Werte



- Bevorzugung bestimmter Endziffern bei numerischen Datenelementen
 - Annahme: gleichverteilte Endziffern
 - Rate: Anzahl Datenelemente mit auffälliger Verteilung der Endziffern / Anzahl überprüfter Datenelemente
- Nachweis bereits bekannter Korrelationen zwischen verschiedenen Datenelementen
 - Rate: Anzahl erwarteter aber nicht nachweisbarer Korrelationen / Anzahl überprüfter Korrelationen

- Aktualität der gespeicherten Daten
 - Festlegung individueller Schwellenwerte für die akzeptable Zeitdifferenz/Verfallzeit der Daten
 - Rate: Anzahl der Werte mit aktueller Zeitdifferenz > Schwellenwert / Anzahl überprüfter Werte
- Anzahl Dubletten im Datenbestand
 - Festlegung eines Sets von Datenelementen, deren Kombination die Erkennung von Dubletten erlaubt
 - Rate: Anzahl gefundener Dubletten / Anzahl überprüfter Sets von Datenelementen



- Rekrutierungsrate im Register und den Zentren
 - Rate: Anzahl rekrutierter Patienten im gewählten Zeitintervall / Anzahl rekrutierbarer Patienten im gewählten Zeitintervall
- Anzahl vorzeitig aus dem Register ausscheidender Patienten
 - Rate: Anzahl Patienten, die innerhalb des gewählten Zeitintervalls vorzeitig aus dem Register ausscheiden / Anzahl Patienten im Register im gewählten Zeitintervall



- Fachliche Qualifikation des Datenerfassungspersonals pro Zentrum
 - Rate: Anzahl der für das Register tätigen Personen mit der erforderlichen Qualifikation / Anzahl der für das Register insgesamt tätigen Personen
- Anzahl Synonyme im Register
 - Synonym: verschiedene Patientenidentifikatoren werden demselben Patienten zugewiesen
 - Rate: Anzahl Synonyme / Anzahl überprüfter Patienten



- Anzahl Homonyme im Register
 - Homonym: ein Patientenidentifikator wird mehreren Patienten zugewiesen
 - Rate: $\text{Anzahl Homonyme} / \text{Anzahl überprüfter Patienten}$

- Genauigkeit der Registerdaten
 - Rate: Anzahl der im Register vorhandenen und korrekten Werte / Anzahl überprüfter Werte
- Vollständigkeit der Registerdaten
 - Rate: Anzahl der im Register vorhandenen Werte / Anzahl der im Register vorhandenen Werte + Anzahl der im Register fehlenden Werte

- Übereinstimmung der Registerdaten mit den Originaldaten
 - Ergebnis der Source Data Verification
 - Einsatz bei Fallzahlplanung Source Data Verification
1. Bezogen auf Datenelemente
 - Rate: Anzahl der Werte, die nicht mit den Originaldaten übereinstimmen / Anzahl überprüfter Werte
 2. Bezogen auf Beobachtungseinheiten
 - Rate: Anzahl der Werte, die nicht mit den Originaldaten übereinstimmen / Anzahl überprüfter Beobachtungseinheiten

- Übereinstimmung mit dem Studienplan
 - Festlegung wesentlicher Elemente des Studienplans, bei denen Abweichungen gemessen werden können
 - Rate: Anzahl Abweichungen vom Studienplan / Anzahl überprüfter Patienten
- Fehlende Werte bei mandatorischen Datenelementen
 - Rate: Anzahl fehlender Werte / Anzahl überprüfter Werte

- Repräsentativität der Registerdaten
 - Repräsentativität: Die im Register erfasste Population ist eine Stichprobe der zu untersuchenden Gesamtpopulation, die in wesentlichen Einflussgrößen mit der Gesamtpopulation weitgehend übereinstimmt.
 - Rate: Anzahl der Einflussgrößen mit erwarteter Verteilung / Anzahl überprüfter Einflussgrößen

- Festlegung eines Schwellenwertes für jeden Qualitätsindikator zur Trennung in „auffällig“ und „unauffällig“
- Beispiele:
 - Ausreißer bei stetigen Datenelementen: 10 %
 - Genauigkeit der Registerdaten: 95 %
 - Vollständigkeit der Registerdaten: 95 %
 - Übereinstimmung der Registerdaten mit den Originaldaten: 5 %

Vielen Dank für Ihre Aufmerksamkeit!