

# **Abschlussbericht**

## **Leitlinie Datenqualität als Rahmen für empirische Forschungsvorhaben**

Dr. Carsten Oliver Schmidt, Institut für Community Medicine-SHIP/KEF, Greifswald

Prof. Dr. med. Jürgen Stausberg, Ludwig-Maximilians-Universität München

Dr. Ron Pritzkeleit, Institut für Krebsepidemiologie e.V., Universität Lübeck

Prof. Dr. Thomas Schrader, Institut für Pathologie, Charité- Universitätsmedizin Berlin,  
Fachhochschule Brandenburg, Fachbereich Informatik & Medien

Dr. Michael Nonnemacher, Institut für Medizinische Informatik, Biometrie und Epidemiologie  
(IMIBE), Universitätsklinikum Essen

09.03.2012



## Inhalt

1	Zusammenfassung.....	3
2	Indikatordefinition.....	4
2.1	Aufbau einer Indikatordefinition.....	4
2.2	Liste der Qualitätsindikatoren .....	5
3	Übersicht neuer bzw. modifizierter Indikatoren.....	9
3.1	Neue Indikatoren der Ebene Integrität .....	9
3.2	Neue Indikatoren der Ebene Organisation.....	29
3.3	Neue Indikatoren der Ebene Metadaten .....	40
4	Anhang .....	42
4.1	Teilbesprechung : Krebsregister.....	42
4.2	Teilbesprechung : SHIP .....	46
4.3	Teilbesprechung : OpEN.SC .....	49

# 1 Zusammenfassung

Aus der TMF wurde 2006 eine Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern vorgelegt, die u. a. 24 Kennzahlen zur Datenqualität beschreibt, diese in eine Struktur einordnet und ein Verfahren zur Verdichtung auf einen Score beinhaltet. Seit 2006 hat das Thema Datenqualität weiter an Bedeutung zugenommen, weshalb mit diesem Projekt zum Zwecke einer Überprüfung und möglichen Adaptation der Leitlinie Kennzahlen auf drei sehr verschiedene Vorhaben abgebildet wurden:

- den epidemiologischen Krebsregistern, vertreten durch die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID)
- der Study of Health in Pomerania (SHIP), vertreten durch das Institut für Community Medizin der Universitätsmedizin Greifswald,
- sowie dem Open European Nephrology Science Center (OpEN.SC), vertreten durch die Charité - Universitätsmedizin Berlin.

Die Ergebnisse belegen die Anwendbarkeit der meisten bestehenden Indikatoren der TMF Leitlinie. Zugleich bestand die Anforderung, einige Kennzahlen neu aufzunehmen sowie bestehende Kennzahlen zu modifizieren. Dies zeigte sich auf Basis der im Anhang aufgeführten Gespräche mit den Vertretern der drei Vorhaben (S. 42ff). Heterogene Anregungen ergaben sich aus der sehr verschiedenartigen Struktur der Projekte. Während SHIP als Kohortenstudie in der Qualitätssicherung vor allem auf die Datengenese fokussiert, arbeiten GEKID bzw. das Daten Repository OpEN.SC mit vorhandenen Daten und QS fokussiert somit stärker auf deren Bewertung. Einen kurzen Überblick zu feineren, identischen, breiteren oder ergänzenden Indikatoren im Vergleich zur vorliegenden TMF Leitlinie liefert die folgende Tabelle:

Kategorie	SHIP	OpEN.SC	GEKID
Feiner	3	6	3
Identisch	1	3	4
Breiter	2	1	0
Ergänzend	2	2	3
Neu	1	1	0
Summe	9	13	10

In den beiden folgenden Abschnitten des Berichtes folgt eine detaillierte Aufstellung zu Indikatordefinitionen (S. 4-8) und der neuen bzw. modifizierten Indikatoren (S. 9-49).

Insgesamt wurde die Grundstruktur der bestehenden TMF Leitlinie nicht verändert. Unsere Projektergebnisse bestätigen die Eignung der TMF-Leitlinie als Referenzrahmen für Kennzahlen zur Datenqualität.

## 2 Indikatordefinition

### 2.1 Aufbau einer Indikatordefinition

<b>Beschreibung</b>	z.B. <i>Anzahl fehlender Werte</i>
<b>Begriffsdefinitionen</b>	werden auch im Glossar aufgeführt
<b>Identifikationsnummer</b>	neutral, TMF-<vierstellige Zahl>
<b>Quelle</b>	Literaturquellen, Expertenmeinung
<b>Bezug</b>	Berechnung für einzelne Datenelemente oder für größere Informationseinheiten (z.B. Meldung, Register)  Berichtsebene: Datensatz, Beobachtungseinheit, Datenbestand
<b>Alternative Definition</b>	Hinweis auf alternative Definitionen
<b>Anmerkungen</b>	
<b>Zähler</b>	z.B. <i>Anzahl fehlender Werte des Datenelements</i>
<b>Nenner</b>	z.B. <i>Anzahl überprüfter Datenelemente, Anzahl aller Patienten</i>
<b>Subkategorien</b>	Subkategorien zur Adjustierung des Indikators, optional
<b>Berechnung</b>	Beschreibung der Datensammlung und –aggregation
<b>Interpretation</b>	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 &lt; 0,5 % für akzeptabel.</i>
<b>Einflussfaktoren</b>	Faktoren auf den Ebenen Struktur und Prozesse, die den Indikator beeinflussen können, sind für eine Schwachstellenanalyse hilfreich  z.B. <i>Einführung von Kategorien für not applicable/not done verringert die Anzahl echter fehlender Werte</i>

## 2.2 Liste der Qualitätsindikatoren

Qualitätsindikator	ID alt	ID neu	Datensatz/ Meldung	Beobachtungs- einheit	Daten- bestand
<b>Ebene Integrität<sup>1</sup></b>					
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
Wertevertellung	B.2.1.4	TMF-1006		x	x
Bevorzugung bestimmter Endziffern	B.2.1.9	TMF-1007			x
Wertevertellung der durch Untersucher erfassten Parameter [SHIP]		TMF-1009			x
Wertevertellung der durch Geräte erfassten Parameter [SHIP]		TMF-1010			x
Wertevertellung von Befunden [SHIP]		TMF-1011			x
Wertevertellung 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

Qualitätsindikator	ID alt	ID neu	Datensatz/ Meldung	Beobachtungs- einheit	Daten- bestand
Fehlende Werte bei optionalen Datenelementen	B.2.1.6	TMF-1015	x	x	x
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 <sup>1</sup>	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
<b>Ebene Organisation</b>					
Aktualität der gespeicherten Daten	B.2.2.1	TMF-1028	x	x	x
Dubletten im Datenbestand <sup>1</sup>	B.2.2.2	TMF-1029		x	x
Rekrutierungsrate	B.2.2.3	TMF-1030			x

Qualitätsindikator	ID alt	ID neu	Datensatz/ Meldung	Beobachtungs- einheit	Daten- bestand
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 <sup>1</sup>	B.2.2.6	TMF-1036			x
Homonyme <sup>1</sup>	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
<b>Ebene Richtigkeit</b>					
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

<b>Qualitätsindikator</b>	<b>ID alt</b>	<b>ID neu</b>	<b>Datensatz/ Meldung</b>	<b>Beobachtungs- einheit</b>	<b>Daten- bestand</b>
Übereinstimmung mit Verfahrensregeln <sup>1</sup>	B.2.3.5	TMF-1047			x
Repräsentativität der Registerdaten	B.2.3.7	TMF-1048			x
<b>Ebene Metadaten</b>					
Umfang der Metadaten bei Untersuchungen [Open.SC]		TMF-1050	x		x

<sup>1</sup>Bezeichnung geändert

## **3 Übersicht neuer bzw. modifizierter Indikatoren**

### **3.1 Neue Indikatoren der Ebene Integrität**

## **Bezeichnung in Übersicht**

Endlosüberlebende

## **Beschreibung**

Anteil Endlosüberlebender an allen Patienten/Probanden

## **Begriffsdefinitionen**

Endlosüberlebender: Patient oder Proband, für den zur Laufzeit eines Vorhabens niemals formal der Tod festgestellt wird (z.B. durch Todesbescheinigung oder Information vom Einwohnermeldeamt)

## **Identifikationsnummer**

TMF-1035

## **Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID); Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Dieser Indikator sagt etwas aus über die Vollständigkeit und Validität von Follow-ups. Wenn Patienten die zu erwartende Überlebenszeit für ihre Tumorerkrankung deutlich überschreiten, kann dies ein Hinweis auf fehlende Informationen zum Tod oder eine fehlerhafte Recherche sein. Dies gilt speziell für Tumore mit schlechter Prognose oder kurzer Überlebenszeit (z.B. Pankreaskarzinom).

## **Zähler**

Anzahl Endlosüberlebender

## **Nenner**

Anzahl Patienten/Probanden im Datenbestand insgesamt

## **Subkategorien**

-

## **Berechnung**

- 1) Identifizierung aller Endlosüberlebenden im Datenbestand
- 2) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Fehlende Informationen von den Meldestellen, unzureichendes Follow-up

## **Bezeichnung in Übersicht**

Sichere Widersprüche/Fehler

## **Beschreibung**

Durch Plausibilitätsprüfungen zwischen mehreren Merkmalen entdeckte Widersprüche und Fehler können nach ihrer Schwere (und der daraus resultierenden Reaktion) unterschieden werden. Dieser Indikator misst sichere Widersprüche und Fehler.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1004

## **Quelle**

TMF-Projekt V020-04: Open European Nephrology Science Center (OpEN.SC), Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID);

Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für einen einzelnen Datensatz, für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Sichere Widersprüche und Fehler müssen bei epidemiologischen Krebsregistern zwingend korrigiert werden. Unzureichende Korrekturen können in diesem Fall zur Zurückweisung einer Meldung führen (siehe Indikator TMF-1040: „Zurückgewiesene Meldungen“).

## **Zähler**

Anzahl Plausibilitätsprüfungen, die einen sicheren Widerspruch oder Fehler zeigen

## **Nenner**

Anzahl aller durchgeführten Plausibilitätsprüfungen

## **Subkategorien**

-

## **Berechnung**

1) Definition des Begriffes „sicherer Widerspruch/Fehler“ (d.h. Widersprüche oder Fehler, die eine große Bedeutung für das Vorhaben besitzt)

2) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Motivation der Studienteilnehmer, Vollständigkeit und Korrektheit der primären Datenquellen

## **Bezeichnung in Übersicht**

Mögliche Widersprüche/Warnungen

## **Beschreibung**

Durch Plausibilitätsprüfungen zwischen mehreren Merkmalen entdeckte Widersprüche und Fehler können nach ihrer Schwere (und der daraus resultierenden Reaktion) unterschieden werden. Dieser Indikator misst mögliche Widersprüche und Fehler.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1005

## **Quelle**

TMF-Projekt V020-04: Open European Nephrology Science Center (OpEN.SC), Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID);

Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für einen einzelnen Datensatz, für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Für epidemiologische Krebsregister liegen mögliche Widersprüche und Fehler dann vor, wenn Daten unplausibel aber prinzipiell möglich sind (z.B. Brustkrebs beim Mann).

## **Zähler**

Anzahl Plausibilitätsprüfungen, die einen möglichen Widerspruch oder Fehler zeigen

## **Nenner**

Anzahl aller durchgeführten Plausibilitätsprüfungen

## **Subkategorien**

-

## **Berechnung**

1) Definition des Begriffes „möglicher Widerspruch/Fehler“ (d.h. Widerspruch oder Fehler, der eine untergeordnete Bedeutung für das Vorhaben besitzt)

2) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Die Rate kann auch einen Hinweis auf unzutreffende Plausibilitätsprüfungen geben.

## **Einflussfaktoren**

Motivation der Studienteilnehmer, Vollständigkeit und Korrektheit der primären Datenquellen, Umfang und Tiefe der Plausibilitätsprüfungen

## **Bezeichnung in Übersicht**

Wertevertelung der durch Untersucher erfassten Parameter

## **Beschreibung**

Bei technischen Untersuchungen betrachtet SHIP die Qualität getrennt für das Gerät, den verantwortlichen Techniker (Observer/Untersucher) sowie die befundende Person (Reader). Dieser Indikator befasst sich mit der Qualität der Daten in Bezug auf den Untersucher.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1009

## **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

## **Bezug**

Die Berechnung erfolgt für einen Datenbestand.

## **Alternative Definition**

-

## **Anmerkungen**

Der Indikator kann sowohl für einzelne Geräte als auch für einzelne technische Untersuchungen mit mehreren Geräten berechnet werden. Der Indikator ist nur dann sinnvoll, wenn als Zwischenschritt zwischen der Erzeugung von Messwerten und der Befundung eine manuelle Erfassung durch Untersucher stattfindet.

## **Zähler**

Anzahl Untersucher mit auffälliger Wertevertelung

## **Nenner**

Anzahl Untersucher

## **Subkategorien**

Der Indikator ist eine Subkategorie des Indikators TMF-1006 „Wertevertelung“.

## **Berechnung**

- 1) Auswahl der technischen Untersuchung
- 2) Festlegung des Kriteriums für Auffälligkeit, z. B. Abweichung von einem Verteilungsparameter um einen bestimmten Betrag
- 3) Berechnung des Verteilungsparameters
- 4) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Bei Einschluss mehrerer Geräte sind Verzerrungen durch feste Bindungen zwischen Untersucher und Geräten auszuschließen. Die Anzahl von Untersuchungen je Untersucher ist zu berücksichtigen.

## **Einflussfaktoren**

Standardisierung der Erfassung von Werten, Schulung der Untersucher

## **Bezeichnung in Übersicht**

Werteverteilung der durch Geräte erfassten Parameter

## **Beschreibung**

Bei technischen Untersuchungen betrachtet SHIP die Qualität getrennt für das Gerät, den verantwortlichen Techniker (Observer) sowie die befundende Person (Reader). Dieser Indikator befasst sich mit der Qualität der Daten in Bezug auf die Geräte.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1010

## **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

## **Bezug**

Die Berechnung erfolgt für einen Datenbestand.

## **Alternative Definition**

-

## **Anmerkungen**

Der Indikator wird für mehrere Geräte berechnet.

## **Zähler**

Anzahl Geräte mit auffälliger Werteverteilung

## **Nenner**

Anzahl Geräte

## **Subkategorien**

Der Indikator ist eine Subkategorie des Indikators TMF-1006 „Werteverteilung“.

## **Berechnung**

- 1) Auswahl der Geräte
- 2) Festlegung des Kriteriums für Auffälligkeit, z. B. Abweichung von einem Verteilungsparameter um einen bestimmten Betrag
- 3) Berechnung des Verteilungsparameters
- 4) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Die Anzahl der Untersuchungen pro Gerät ist zu berücksichtigen.

## **Einflussfaktoren**

Regelmäßige Wartung der Geräte, korrekte Einstellung und Handhabung der Geräte

## **Bezeichnung in Übersicht**

Werteverteilung von Befunden

## **Beschreibung**

Bei technischen Untersuchungen betrachtet SHIP die Qualität getrennt für das Gerät, den verantwortlichen Techniker (Observer) sowie die befundende Person (Reader). Dieser Indikator befasst sich mit der Qualität der Daten in Bezug auf die befundende Person.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1011

## **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

## **Bezug**

Die Berechnung erfolgt für einen Datenbestand.

## **Alternative Definition**

-

## **Anmerkungen**

Der Indikator kann sowohl für einzelne Geräte als auch für einzelne technische Untersuchungen mit mehreren Geräten berechnet werden.

## **Zähler**

Anzahl Befunder mit auffälliger Werteverteilung

## **Nenner**

Anzahl Befunder

## **Subkategorien**

Der Indikator ist eine Subkategorie des Indikators TMF-1006 „Werteverteilung“.

## **Berechnung**

- 1) Auswahl der technischen Untersuchung
- 2) Festlegung des Kriteriums für Auffälligkeit, z. B. Abweichung von einem Verteilungsparameter um einen bestimmten Betrag
- 3) Berechnung des Verteilungsparameters
- 4) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Die Anzahl der Befunde je Befunder ist zu berücksichtigen.

## **Einflussfaktoren**

Qualifikation der befundenden Personen, regelmäßige Zweitbefundung mit Feedback

## **Bezeichnung in Übersicht**

Werteverteilung von Parametern zwischen Zentren

## **Beschreibung**

Messwerte können sich systematisch zwischen Zentren unterscheiden. Mögliche Unterschiede werden mit diesem Punkt abgebildet.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1052

## **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

## **Bezug**

Die Berechnung erfolgt für einen Datenbestand.

## **Alternative Definition**

-

## **Anmerkungen**

Der Indikator wird für alle Zentren berechnet.

## **Zähler**

Anzahl Zentren mit auffälliger Werteverteilung

## **Nenner**

Anzahl Zentren

## **Subkategorien**

Der Indikator ist eine Subkategorie des Indikators TMF-1006 „Werteverteilung“.

## **Berechnung**

- 1) Auswahl der Zentren
- 2) Festlegung des Kriteriums für Auffälligkeit, z. B. Abweichung von einem Verteilungsparameter um einen bestimmten Betrag
- 3) Berechnung des Verteilungsparameters
- 4) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Schulungsvoraussetzungen, organisatorische Gegebenheiten, Gerätespezifika etc. in den Zentren.

## **Bezeichnung in Übersicht**

Fehlende Module

## **Beschreibung**

Der Datensatz eines Vorhabens kann in verschiedene Module unterteilt sein, z. B. nach Verfahren wie EKG oder nach sachlogischen Bezügen wie Soziodemographie. Module können auch über Gruppen von zusammenhängenden Variablen definiert sein. Dieser Indikator prüft die Vollständigkeit, in denen die Module vorliegen.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1012

## **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

## **Bezug**

Die Berechnung kann für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Dieser Indikator kann sowohl beschränkt auf ein Modul wie Soziodemographie als auch unter Einschluss verschiedener Module durchgeführt werden. Betrachtet werden nur komplette Module. Fehlende Module können z. B. durch fehlende Compliance von Studienteilnehmern in Bezug auf bestimmte Untersuchungen bedingt sein.

## **Zähler**

Anzahl fehlender Module

## **Nenner**

Anzahl überprüfter Module

## **Subkategorien**

-

## **Berechnung**

- 1) Suche nach fehlenden Modulen
- 2) Berechnung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.

## **Einflussfaktoren**

Motivation der Studienteilnehmer, Belastung der Studienteilnehmer durch Untersuchungen, Geräteausfällen, Studiendesign

## **Bezeichnung in Übersicht**

Fehlende Werte bei Datenelementen

## **Beschreibung**

Anteil fehlender Werte bei Datenelementen

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1013

## **Quelle**

TMF-Projekt V020-04: Ergänzung erfolgt aus Gründen der Systematik

## **Bezug**

Die Berechnung kann für einen einzelnen Datensatz, für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Dieser Indikator wurde eingeführt, um eine Berechnung der Anzahl fehlender Werte unter Einschluss von optionalen und mandatorischen Datenelementen zu ermöglichen.

## **Zähler**

Anzahl fehlender Werte

## **Nenner**

Anzahl überprüfter Werte

## **Subkategorien**

TMF-1014 „Fehlende Werte bei mandatorischen Datenelementen“

TMF-1015 „Fehlende Werte bei optionalen Datenelementen“

## **Berechnung**

- 1) Suche nach fehlenden Werten
- 2) Berechnung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.

## **Einflussfaktoren**

Einführung von separaten Kategorien für „not applicable“, „not done“. Nutzung eines EDC-Systems mit Erzwingung von Einträgen bei mandatorischen Datenelementen.

### **Bezeichnung in Übersicht**

Datenelemente mit Wert unbekannt o. ä.

### **Beschreibung**

Anteil von Datenelementen mit Werten unbekannt o. ä.

### **Begriffsdefinitionen**

-

### **Identifikationsnummer**

TMF-1016

### **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

### **Bezug**

Die Berechnung kann für einen einzelnen Datensatz, für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

### **Alternative Definition**

-

### **Anmerkungen**

Dieser Eintrag ist nur für mandatorische Datenelemente mit einer vorgegebenen Werteliste sinnvoll, bei denen ein Eintrag unbekannt o. ä. verfügbar ist. Dieser Indikator ist verwandt mit dem Indikator TMF-1025 „Datenelemente mit unspezifischen Werten“

### **Zähler**

Anzahl Werte unbekannt o. ä.

### **Nenner**

Anzahl überprüfter Werte

### **Subkategorien**

-

### **Berechnung**

- 1) Festlegung der relevanten Bezeichnungen („unbekannt“ o. ä.)
- 2) Suche nach den entsprechenden Einträgen
- 3) Berechnung von Zähler, Nenner und Rate

### **Interpretation**

Je höher die Rate, desto schlechter die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.

### **Einflussfaktoren**

Motivation des Erhebungspersonals, Art der Erhebung

### **Bezeichnung in Übersicht**

Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten

### **Beschreibung**

Anteil von Datenelementen mit bestehenden Einträgen bei allen Beobachtungseinheiten

### **Begriffsdefinitionen**

-

### **Identifikationsnummer**

TMF-1017

### **Quelle**

TMF-Projekt V020-04: Open European Nephrology Science Center (OpEN.SC)

### **Bezug**

Die Berechnung erfolgt für einen Datenbestand.

### **Alternative Definition**

-

### **Anmerkungen**

Die Berechnung ist sowohl für optionale als auch für mandatorische Datenelemente sinnvoll. Festzulegen ist der Umgang mit Einträgen wie „unbekannt“ umgegangen wird.

### **Zähler**

Anzahl von Datenelementen mit Einträgen bei allen Beobachtungseinheiten

### **Nenner**

Anzahl von überprüften Datenelementen

### **Subkategorien**

-

### **Berechnung**

- 1) Festlegung der einzuschließenden Datenelemente
- 2) Bestimmung für jedes Datenelement, ob Werte von allen Beobachtungseinheiten vorliegen
- 3) Berechnung von Zähler, Nenner und Rate

### **Interpretation**

Je niedriger die Rate, desto schlechter die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.

### **Einflussfaktoren**

Die Einflussfaktoren sind vielfältig, da dieser Indikator Aspekte anderer Kennzahlen bündelt. Es ist daher sinnvoll, diesen Indikator nicht getrennt sondern zusammen mit anderen aus dem Bereich fehlender Einträge zu betrachten.

### **Bezeichnung in Übersicht**

Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten

### **Beschreibung**

Anteil von Werten, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten

### **Begriffsdefinitionen**

-

### **Identifikationsnummer**

TMF-1019

### **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

### **Bezug**

Die Berechnung kann für einen einzelnen Datensatz, für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

### **Alternative Definition**

-

### **Anmerkungen**

Die Berechnung ist nur dann sinnvoll, wenn die Werte, die Messbarkeitsgrenzen unter- oder überschreiten, identifizierbar sind, z. B. durch eine bestimmte Kodierung. Die Auswertung kann verfahrensspezifisch erfolgen. Es kommen nur technische Verfahren in Betracht. Zudem müssen Messbarkeitsgrenzen in der Praxis relevant sein.

### **Zähler**

Anzahl von Werten, die die Messbarkeitsgrenzen unter- oder überschreiten

### **Nenner**

Anzahl von überprüften Werten

### **Subkategorien**

-

### **Berechnung**

- 1) Auswahl des Verfahrens
- 2) Suche nach Werten, die die Messbarkeitsgrenzen unter- oder überschreiten
- 3) Berechnung von Zähler, Nenner und Rate

### **Interpretation**

Je höher die Rate, desto schlechter die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab. Hohe Werte zeigen eine unangemessene Wahl des Verfahrens für das untersuchte Kollektiv an.

### **Einflussfaktoren**

Verfahrensgerechte Durchführung der Untersuchungen, korrekte Indikationsstellung

## **Bezeichnung in Übersicht**

Werte aus Standards

## **Beschreibung**

Anteil von Werten, die mit Bezeichnungen aus kontrollierten Vokabularen übereinstimmen

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1020

## **Quelle**

TMF-Projekt V020-04: Open European Nephrology Science Center (OpEN.SC)

## **Bezug**

Die Berechnung erfolgt für einen Datenbestand.

## **Alternative Definition**

-

## **Anmerkungen**

Die Berechnung ist nur sinnvoll für Freitextfelder oder Datenelemente, bei denen die Liste der erlaubten Werte bei der Erfassung ergänzt werden kann. Es ist sinnvoll, dass zum Textvergleich eingesetzte Verfahren einer expliziten Qualitätsprüfung zu unterziehen.

## **Zähler**

Anzahl von Werten mit Bezeichnungen aus kontrollierten Vokabularen.

## **Nenner**

Anzahl überprüfter Werte

## **Subkategorien**

-

## **Berechnung**

- 1) Festlegung der zu berücksichtigenden kontrollierten Vokabulare
- 2) Durchführung einer Normalisierung der Einträge bzw. Festlegung eines Verfahrens zum Textvergleich
- 3) Suche nach Werten, die mit den Bezeichnungen aus kontrollierten Vokabularen übereinstimmen
- 4) Berechnung von Zähler, Nenner und Rate

## **Interpretation**

Je niedriger die Rate, desto schlechter die Datenqualität. Da bei der Definition der Datenelemente auf die explizite Verwendung eines kontrollierten Vokabulars verzichtet wurde, ist eine Rate von 100 % nicht zu erwarten. Die für diese Entscheidung relevanten Gründe müssen bei der Interpretation der Ergebnisse berücksichtigt werden.

## **Einflussfaktoren**

Bedeutung und Bekanntheitsgrad von kontrollierten Vokabularen bei den betrachteten Datenelementen

## **Bezeichnung in Übersicht**

Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings

## **Beschreibung**

Anzahl unerlaubter Werte bei qualitativen Datenelementen zur Kodierung von Missings

## **Begriffsdefinitionen**

Qualitatives Datenelement: Ein Datenelement, das nur eine endliche Zahl qualitativer Werte annehmen kann, d.h. Werte die sich nicht quantifizieren lassen (z.B. nach ICD-10 kodierte Diagnose).

## **Identifikationsnummer**

TMF-1022

## **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

## **Bezug**

Die Berechnung des Indikators ist nur für qualitative Datenelemente sinnvoll. Die Berechnung kann für einen einzelnen Datensatz, für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Dieser Indikator macht nur Sinn, wenn das Auftreten von Missings technisch überhaupt möglich und eine Kodierung vorgesehen ist.

## **Zähler**

Anzahl der unerlaubten Werte

## **Nenner**

Anzahl überprüfter Werte

## **Subkategorien**

-

## **Berechnung**

- 1) Festlegung der erlaubten Werte für Missings pro Datenelement
- 2) Suche nach unerlaubten Werten
- 3) Berechnung von Zähler, Nenner und Rate

## **Interpretation**

Jeder unerlaubte Wert ist ein Datenfehler. Je höher die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Automatische Belegung von Missings mit definierten Codes

**Bezeichnung in Übersicht**

Unerlaubte Werte zur Kodierung von fehlenden Modulen

**Beschreibung**

Anzahl unerlaubter Werte zur Kodierung von fehlenden Modulen

**Begriffsdefinitionen**

-

**Identifikationsnummer**

TMF-1023

**Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

**Bezug**

Die Berechnung kann für einen einzelnen Datensatz, für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

**Alternative Definition**

-

**Anmerkungen**

Der Indikator ist verwandt mit Indikator TMF-1012 „Fehlende Module“.

**Zähler**

Anzahl der unerlaubten Werte

**Nenner**

Anzahl überprüfter Werte

**Subkategorien**

-

**Berechnung**

- 1) Festlegung der erlaubten Werte für fehlende Module
- 2) Suche nach unerlaubten Werten
- 3) Berechnung von Zähler, Nenner und Rate

**Interpretation**

Jeder unerlaubte Wert ist ein Datenfehler. Je höher die Rate, desto schlechter ist die Datenqualität.

**Einflussfaktoren**

Automatische Belegung von fehlenden Modulen mit definierten Codes

### **Bezeichnung in Übersicht**

Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen

### **Beschreibung**

Anzahl unerlaubter Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen

### **Begriffsdefinitionen**

-

### **Identifikationsnummer**

TMF-1024

### **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

### **Bezug**

Die Berechnung kann für einen einzelnen Datensatz, für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

### **Alternative Definition**

-

### **Anmerkungen**

Dieser Indikator ist verwandt mit Indikator TMF-1019 „Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten“

### **Zähler**

Anzahl der unerlaubten Werte

### **Nenner**

Anzahl überprüfter Werte

### **Subkategorien**

-

### **Berechnung**

- 1) Festlegung der erlaubten Werte zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen
- 2) Suche nach unerlaubten Werten
- 3) Berechnung von Zähler, Nenner und Rate

### **Interpretation**

Jeder unerlaubte Wert ist ein Datenfehler. Je höher die Rate, desto schlechter ist die Datenqualität.

### **Einflussfaktoren**

Automatische Belegung von Unter- oder Überschreitungen von Messbarkeitsgrenzen mit definierten Codes

## **Bezeichnung in Übersicht**

Datenelemente mit unspezifischen Werten

## **Beschreibung**

Anzahl von Datenelementen mit unspezifischen Werten

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1025

## **Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID);

Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Dieser Indikator beschreibt die Informationsdichte.

## **Zähler**

Anzahl Datenelemente mit unspezifischer Angabe

## **Nenner**

Anzahl überprüfter Datenelemente

## **Subkategorien**

Bei epidemiologischen Krebsregistern wird dieser Indikator berechnet für

- Uterus-Malignome
- Histologie
- Lokalisation
- Diagnosen (ICD-10-Kodes der Form „xxx.9“)

## **Berechnung**

- 1) Festlegung der zu untersuchenden Datenelemente
- 2) Definition der Kategorie „unspezifisch“ für die zu untersuchenden Datenelemente
- 3) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Auf der Ebene des einzelnen Datensatzes kann der Indikator u.U. ignoriert werden, da bei der Zusammenführung von Datensätzen für eine Beobachtungseinheit unspezifische Angaben in einem Datensatz durch spezifischere Angaben in

einem anderen Datensatz überschrieben werden können. Bei Uterus-Malignomen sollte die Rate unter 5 % liegen.

### **Einflussfaktoren**

Genauigkeit der Diagnostik, Vollständigkeit und Korrektheit der primären Datenquellen

**Bezeichnung in Übersicht**

Beobachtungseinheiten mit unbekanntem Primärtumor

**Beschreibung**

Anteil der Beobachtungseinheiten mit unbekanntem Primärtumor

**Begriffsdefinitionen**

-

**Identifikationsnummer**

TMF-1026

**Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID);

Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

**Bezug**

Die Berechnung kann für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

**Alternative Definition**

-

**Anmerkungen**

Dieser Indikator hat eine enge Beziehung zum Indikator TMF-1025 „Datenelemente mit unspezifischen Werten“ und kann ggf. als weitere Präzisierung aufgefasst werden.

**Zähler**

Anzahl der Beobachtungseinheiten mit unbekanntem Primärtumor

**Nenner**

Anzahl aller überprüften Beobachtungseinheiten

**Subkategorien**

-

**Berechnung**

- 1) Identifizierung aller Beobachtungseinheiten mit unbekanntem Primärtumor
- 2) Bestimmung von Zähler, Nenner und Rate

**Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Die Rate sollte unter 5 % liegen.

**Einflussfaktoren**

Genauigkeit der Diagnostik, Vollständigkeit und Korrektheit der primären Datenquellen

## **3.2 Neue Indikatoren der Ebene Organisation**

## **Bezeichnung in Übersicht**

DCO-Rate (Death Certificate Only)

## **Beschreibung**

Anteil der Krebserkrankungen, für die nur eine Todesbescheinigung mit einem Hinweis auf die Krebserkrankung vorliegt.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1051

## **Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID); Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Dieser Indikator beschreibt die Vollzähligkeit der Krebsregistrierung. Er ist der wichtigste Indikator für epidemiologische Krebsregister und ist dort zur Beurteilung eines Datenbestandes der entscheidende Indikator. Nach seinem Wert entscheiden Forschungsprojekte (z.B. Survivalanalysen auf regionaler Ebene) und internationale Datensammlungen (Globocan, Cancer Incidence in five continents), ob die Daten eines Krebsregisters verwendet werden oder nicht. Bei der Berechnung und Beurteilung anderer Indikatoren ist die DCO-Rate als Einflussgröße zu berücksichtigen.

## **Zähler**

Anzahl der Krebserkrankungen, für die nur eine Todesbescheinigung mit einem Hinweis auf die Krebserkrankung vorliegt.

## **Nenner**

Anzahl aller erfassten Krebserkrankungen

## **Subkategorien**

Die einzelnen Tumorentitäten

## **Berechnung**

- 1) Identifizierung der Krebserkrankungen, für die nur eine Todesbescheinigung mit einem Hinweis auf die Krebserkrankung vorliegt
- 2) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Die DCO-Rate sollte insgesamt unter 5 % liegen. Für die einzelnen Tumorentitäten können die Referenzwerte davon abweichen. Bei jungen Krebsregistern wird die DCO-Rate vergleichsweise hoch sein, da auch noch Jahre nach Beginn der

Registrierung Personen an oder mit Krebs sterben, die vor Beginn der Registrierung an Krebs erkrankt sind.

**Einflussfaktoren**

Bereitschaft der Meldestellen zur Meldung, fehlerhafte Zusammenführung von Meldungen

**Bezeichnung in Übersicht**

Verweigerungsrate von Untersuchungen

**Beschreibung**

Anteil von verweigerter Untersuchungen

**Begriffsdefinitionen**

-

**Identifikationsnummer**

TMF-1031

**Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

**Bezug**

Die Berechnung erfolgt für einen Datenbestand.

**Alternative Definition**

-

**Anmerkungen**

Der Indikator ist verwandt mit Indikator TMF-1012 „Fehlende Module“. Voraussetzung zur Ermittlung des Indikators ist die Möglichkeit, eine Verweigerung von Untersuchungen explizit zu dokumentieren. Dieser Indikator kann sowohl für eine spezifische Untersuchung, z. B. MRT, als auch für verschiedene Untersuchungen zusammen bestimmt werden.

**Zähler**

Anzahl verweigerter Untersuchungen

**Nenner**

Anzahl überprüfter Untersuchungen

**Subkategorien**

-

**Berechnung**

- 1) Identifikation des Datenelements, in dem die Verweigerung erfasst wird, für jeden Untersuchungstyp
- 2) Bestimmung des Wertes, der eine Verweigerung anzeigt
- 3) Suche nach dem Auftreten dieses Wertes
- 4) Berechnung von Zähler, Nenner und Rate

**Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.

**Einflussfaktoren**

Studiendesign, Inzentives für Studienteilnehmer, Belastung der Studienteilnehmer

### **Bezeichnung in Übersicht**

Verweigerungsrate von Modulen

### **Beschreibung**

Anteil von verweigerter Modulen

### **Begriffsdefinitionen**

-

### **Identifikationsnummer**

TMF-1032

### **Quelle**

TMF-Projekt V020-04: Open European Nephrology Science Center (OpEN.SC), Study in Health in Pomerania (SHIP)

### **Bezug**

Die Berechnung erfolgt für einen Datenbestand.

### **Alternative Definition**

-

### **Anmerkungen**

Der Indikator ist eng verwandt mit den Indikatoren TMF-1012 „Fehlende Module“ und TMF-1031 „Verweigerungsrate von Untersuchungen“. Voraussetzung zur Ermittlung des Indikators ist die Möglichkeit, eine Verweigerung von Modulen explizit zu dokumentieren.

### **Zähler**

Anzahl verweigerter Module

### **Nenner**

Anzahl überprüfter Module

### **Subkategorien**

-

### **Berechnung**

- 1) Identifikation des Datenelements, in dem die Verweigerung erfasst wird
- 2) Bestimmung des Wertes, der eine Verweigerung anzeigt
- 3) Suche nach dem Auftreten dieses Wertes
- 4) Berechnung von Zähler, Nenner und Rate

### **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.

### **Einflussfaktoren**

Studiendesign, Inzentives für Studienteilnehmer

## **Bezeichnung in Übersicht**

Verweigerungsrate bei einzelnen Datenelementen

## **Beschreibung**

Anteil von verweigerten Datenelementen

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1033

## **Quelle**

TMF-Projekt V020-04: Study in Health in Pomerania (SHIP)

## **Bezug**

Die Berechnung erfolgt für einen Datenbestand.

## **Alternative Definition**

-

## **Anmerkungen**

Der Indikator ist komplementär zum Indikator TMF-1013 „Fehlende Werte bei Datenelementen“ und dessen Subkategorien TMF-1014 und TMF-1015. Voraussetzung zur Ermittlung des Indikators ist die Möglichkeit, eine Verweigerung der Beantwortung von Datenelementen explizit zu dokumentieren. Es wird sich bei den Datenelementen regelhaft um Elemente einer Befragung handeln, kaum um einen isolierten Parameter aus einer technischen Untersuchung.

## **Zähler**

Anzahl von Werten mit Verweigerung

## **Nenner**

Anzahl überprüfter Werte

## **Subkategorien**

-

## **Berechnung**

- 1) Identifikation von Datenelementen, bei denen eine Verweigerung erfasst wird
- 2) Bestimmung des Wertes, der eine Verweigerung anzeigt
- 3) Suche nach dem Auftreten dieses Wertes
- 4) Berechnung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität. Welche Rate noch akzeptabel ist, hängt von der vorgesehenen Datennutzung ab.

## **Einflussfaktoren**

Sensibilität des Datenelements, Art der Befragung, Schulung der Interviewer

## **Bezeichnung in Übersicht**

Solitäre Meldungen

## **Beschreibung**

Zu einem Tumorfall können mehrere Meldungen vorliegen. Der Indikator prüft, wie häufig zu einem Tumorfall nur eine einzige Meldung vorliegt.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1038

## **Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID);

Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Zu einem Tumorfall können mehrere Meldungen (z.B. vom niedergelassenen Arzt, Klinikarzt und Pathologen) vorliegen. Wenn nur eine einzelne Meldung zu einem Fall vorliegt, kann dies ein Hinweis auf schlechte Datenqualität sein, hervorgerufen durch fehlende oder unvollständige Meldungen. Dies gilt speziell für solitäre Meldungen von Pathologen, die hier mit berücksichtigt werden, aber auch durch einen separaten Qualitätsindikator (TMF-1039) zusätzlich untersucht werden können.

## **Zähler**

Anzahl Beobachtungseinheiten mit nur einer Meldung

## **Nenner**

Anzahl überprüfter Beobachtungseinheiten

## **Subkategorien**

-

## **Berechnung**

1) Bestimmung der Anzahl Meldungen für jede Beobachtungseinheit

2) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Vollständigkeit der Meldungen, Motivation der Meldestellen zur Meldung

## **Bezeichnung in Übersicht**

Solitäre Meldungen von Pathologen

## **Beschreibung**

Zu einem Tumorfall können mehrere Meldungen vorliegen. Der Indikator prüft, wie häufig zu einem Tumorfall nur eine solitäre Meldung eines Pathologen vorliegt.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1039

## **Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID);

Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Zu einem Tumorfall können mehrere Meldungen (z.B. vom niedergelassenen Arzt, Klinikarzt und Pathologen) vorliegen. Wenn nur eine einzelne Pathologen-Meldung zu einem Fall vorliegt, bedeutet dies häufig eine schlechte Datenqualität, da in einer Pathologen-Meldung wichtige Informationen (z.B. Datum der Erstdiagnose), die dem Pathologen nicht bekannt sind, fehlen. Ein separater Qualitätsindikator (TMF-1038) überprüft solitäre Meldungen unabhängig vom Typ der Meldestelle.

## **Zähler**

Anzahl Beobachtungseinheiten mit solitärer Pathologen-Meldung

## **Nenner**

Anzahl überprüfter Beobachtungseinheiten

## **Subkategorien**

-

## **Berechnung**

1) Bestimmung der Anzahl Beobachtungseinheiten mit einer solitären Pathologen-Meldung

2) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Vollständigkeit der Meldungen, Motivation der Meldestellen zur Meldung

## **Bezeichnung in Übersicht**

Zurückgewiesene Meldungen

## **Beschreibung**

Meldungen, die aufgrund fehlerhafter Angaben nicht verarbeitet werden können, werden zurückgewiesen. Der Indikator misst den Anteil solcher Meldungen

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1040

## **Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID); Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Meldungen, die aufgrund fehlerhafter Angaben nicht verarbeitet werden können, werden zurückgewiesen. Die meldende Stelle hat dann die Möglichkeit, die Meldung mit korrigierten Daten erneut vorzunehmen.

## **Zähler**

Anzahl zurückgewiesener Meldungen

## **Nenner**

Anzahl überprüfter Meldungen

## **Subkategorien**

-

## **Berechnung**

- 1) Identifizierung zurückgewiesener Meldungen
- 2) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Vollständigkeit und Korrektheit der primären Datenquellen, Motivation der Melder

## **Bezeichnung in Übersicht**

Solitäre Datenquelle

## **Beschreibung**

Zu einem Tumorfall können Meldungen aus einer oder mehreren Datenquellen vorliegen. Der Indikator prüft, wie häufig zu einem Tumorfall nur eine einzige Datenquelle existiert.

## **Begriffsdefinitionen**

-

## **Identifikationsnummer**

TMF-1041

## **Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID);

Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

## **Bezug**

Die Berechnung kann für eine einzelne Beobachtungseinheit und für einen Datenbestand durchgeführt werden.

## **Alternative Definition**

-

## **Anmerkungen**

Zu einem Tumorfall können Meldungen aus mehreren Datenquellen (z.B. vom niedergelassenen Arzt, Klinikarzt und Pathologen) vorliegen. Wenn nur eine einzelne Datenquelle zu einem Fall zur Verfügung steht, fehlen möglicherweise wichtige Daten, da sie in dieser Datenquelle nicht erfasst wurden. Die Todesbescheinigung darf nicht mit in die Berechnung eingeschlossen werden.

## **Zähler**

Anzahl Beobachtungseinheiten mit nur einer Datenquelle

## **Nenner**

Anzahl überprüfter Beobachtungseinheiten

## **Subkategorien**

-

## **Berechnung**

1) Bestimmung der Anzahl Datenquellen für jede Beobachtungseinheit

2) Bestimmung von Zähler, Nenner und Rate

## **Interpretation**

Je höher die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Motivation der Meldestellen zur Meldung, Zeitpunkt der Berechnung

**Bezeichnung in Übersicht**

Beobachtungseinheiten mit Follow-up

**Beschreibung**

Anteil Beobachtungseinheiten mit Follow-up

**Begriffsdefinitionen**

-

**Identifikationsnummer**

TMF-1042

**Quelle**

TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID);

Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008

**Bezug**

Die Berechnung kann für einen Datenbestand durchgeführt werden.

**Alternative Definition**

-

**Anmerkungen**

Dieser Indikator sagt etwas aus über die Vollzähligkeit von Follow-ups. Follow-ups sind wichtig für Verlaufsuntersuchungen (z.B. Bestimmung der Überlebenszeit nach Erstdiagnose eines Tumors). Voraussetzung ist die Kenntnis über die durchzuführenden Follow-ups.

**Zähler**

Anzahl Beobachtungseinheiten mit Follow-up

**Nenner**

Anzahl überprüfter Beobachtungseinheiten

**Subkategorien**

-

**Berechnung**

- 1) Identifizierung der Beobachtungseinheiten, für die mindestens ein Follow-up vorhanden ist
- 2) Bestimmung von Zähler, Nenner und Rate

**Interpretation**

Je geringer die Rate, desto schlechter ist die Datenqualität.

**Einflussfaktoren**

Bereitschaft der Meldestellen zur Meldung, Mobilität der Patienten

### 3.3 Neue Indikatoren der Ebene Metadaten

## **Bezeichnung in Übersicht**

Umfang der Metadaten bei Untersuchungen

## **Beschreibung**

Anteil der von Untersuchungen übermittelten Metadaten

## **Begriffsdefinitionen**

Metadaten: Daten, die andere Daten definieren und beschreiben (aus ISO/IEC 11179-3 Edition 3); hier verstanden als Parameter einer technischen Untersuchung.

## **Identifikationsnummer**

TMF-1050

## **Quelle**

TMF-Projekt V020-04: Open European Nephrology Science Center (OpEN.SC)

## **Bezug**

Die Berechnung kann für einen einzelnen Datensatz und für einen Datenbestand durchgeführt werden. Sie beschränkt sich auf einen Untersuchungstyp, z. B. ein bildgebendes Verfahren.

## **Alternative Definition**

-

## **Anmerkungen**

Dieser Indikator ist nur dann sinnvoll, wenn die Möglichkeit zur Auslassung von Parametern bei der Übermittlung von Ergebnissen technischer Untersuchungen besteht.

## **Zähler**

Anzahl vorhandener Parameter

## **Nenner**

Anzahl der zu erwartenden Parameter

## **Subkategorien**

-

## **Berechnung**

- 1) Festlegung der möglichen Parameter interessierender Untersuchung
- 2) Bestimmung der vorhandenen Parameter
- 3) Berechnung von Zähler, Nenner und Rate

## **Interpretation**

Je niedriger die Rate, desto schlechter ist die Datenqualität.

## **Einflussfaktoren**

Umfang von Schnittstellen, Einstellung von Geräten

## 4 Anhang

### 4.1 Teilbesprechung : Krebsregister

Arbeitspaket 1.1

Besprechung am 03.11.2011 in Lübeck,

Teilnehmer: Prof. Dr. A. Katalinic, Dr. R. Pritzkeleit, Dr. M. Nonnemacher

---

Zur Verfügung stehende Unterlagen zu Krebsregistern:

1. Hentschel S, Katalinic A (Herausgeber). Das Krebsregister-Manual der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. 2008
  2. Pritzkeleit R, Rawal R, Holzmann M, Gerdemann U, Katalinic A. Krebs in Schleswig-Holstein. Kurzbericht zu Inzidenz und Mortalität im Jahr 2005 und Datenqualität im Krebsregister Schleswig-Holstein 2003-2005. Lübeck 2008
  3. Tyczynski JE, Demaret E, Maxwell Parkin D. Standards and guidelines for cancer registration in Europe: then ENCR recommendations. IARC Technical Publication No. 40. Lyon 2003
  4. Parkin DM, Chen VW, Ferlay J, Galceran J, Storm HH, Whelan SL. Comparability and quality control in cancer registration. IARC Technical Report No. 19. Lyon 1994
  5. Bullard J, Coleman MP, Robinson D, Lutz J-M, Bell J, Peto J. Completeness of cancer registration: a new method for routine use. British Journal of Cancer 2000;82(5):1111–1116
  6. Brenner H, Stegmaier C, Ziegler H. Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. Journal of Epidemiology and Community Health 1995;49:426-430
  7. Schouten LJ, Straatman H, Kiemeneij LALM, Gimbrere CHF, Verbeek ALM. The Capture-Recapture Method for Estimation of Cancer Registry Completeness: A Useful Tool? International Journal of Epidemiology 1994;23(6):1111-1116
  8. Haberland J, Schön D, Bertz J, Görsch B. Vollzähligkeitsschätzungen von Krebsregisterdaten in Deutschland. Bundesgesundheitsbl - Gesundheitsforsch - Gesundheitsschutz 2003 · 46:770–774
- 

Zusammenfassung der Diskussion:

- Die Definition der Qualitätsindikatoren ist unter den epidemiologischen Krebsregistern in Deutschland weitgehend harmonisiert und orientiert sich an Empfehlungen der IARC (International Agency for Research on Cancer).
- Das Robert-Koch-Institut ist eine zentrale Sammelstelle für Krebsdaten in Deutschland. Qualitätsindikatoren können helfen, die Qualität dieser Daten zu beurteilen, speziell im Hinblick auf die Datenauswertung.
- Es ergeben sich somit drei Zielgruppen für den Einsatz von Qualitätsindikatoren: Epidemiologische Krebsregister, klinische Krebsregister und Forscher, die mit den Krebsdaten arbeiten möchten.

- Die in der TMF-Leitlinie niedergelegte Vorgehensweise ist teilweise für Krebsregister irrelevant. Irrelevant sind die Bildung des Qualitätsscores und die Planung der Source Data Verification.
- Die im Basisdatensatz definierten Datenelemente und die registerspezifischen Datenelemente sind als mandatorische Datenelemente im Sinne der TMF-Leitlinie aufzufassen.
- Referenzwerte für Qualitätsindikatoren sind nicht immer sinnvoll (z.B. bei Indikatoren zur Beschreibung der Prozessqualität).
- Es muss unterschieden werden zwischen einzelnen Qualitätsprüfungen, die auffällige Daten/Meldungen herausfiltern, und ihrer Umsetzung in Qualitätsindikatoren (als Rate auffälliger Daten/Meldungen bezogen auf die Gesamtzahl der untersuchten Daten/Meldungen).
- Im Krebsregister gibt es verschiedene unterschiedliche Datenbestände, deren Qualität jeweils mit Hilfe von Qualitätsindikatoren bestimmt wird:
  1. Erste Ebene ist der **gemeldete Datensatz**. Hier werden Vollständigkeit, Gültigkeit und in eingeschränktem Maß auch schon die Plausibilität der Angaben geprüft. Daten von schlechter Qualität werden abhängig vom Meldeweg u. U. nicht in das Krebsregister aufgenommen. Dies muss bei der Definition der Datenbasis berücksichtigt werden. Die Anforderungen an die Qualität der Daten hängen von der Meldestelle ab (z.B. unterschiedliche Referenzwerte für Kliniken und Pathologen). Weiterhin ist zu beachten, dass bei unplausiblen Angaben eine Datenkorrektur durch die Meldestelle erfolgen kann. Die auf dieser Ebene genutzten Qualitätsindikatoren müssen auf aggregierter Ebene gebildet werden (z.B. Anteil der Meldungen vom Pathologen ohne Angabe des Wohnorts an allen Pathologiemeldungen).
  2. Die zweite Ebene ist der **Auswertungsdatensatz (Best-of-Datensatz)**, der sich ggf. aus mehreren Einzeldatensätzen (d.h. verschiedene Meldungen zum gleichen Patienten und Tumor) zusammensetzt. Hier ist neben Vollständigkeit und Gültigkeit die Plausibilität entscheidend. Plausibilitätsprüfungen können zwei Ausgänge haben: Entweder sind die Daten fehlerhaft oder unwahrscheinlich. Es kann vorkommen, dass ein Datensatz unplausibel ist, aber trotzdem gültig (z.B. Ovariakarzinom bei einem Mann nach einer Geschlechtsumwandlung). Solche Datensätze müssen manuell gekennzeichnet werden. Auch hier müssen die Qualitätsindikatoren aggregiert erstellt werden.
  3. Die dritte Ebene ist der **Gesamtdatensatz**. Hier gibt es globale Qualitätsindikatoren (z.B. Vollzähligkeit, Anteil DCO-Fälle) mit konsentierten Referenzwerten. Die Indikatoren lassen sich auf einzelne Tumorgruppen herunter brechen und haben dann durchaus auch unterschiedliche Referenzwerte (z.B. würde man bei LungenCa höhere DCO- Raten akzeptieren als bei Mammakarzinom).

Eine vollständige Liste aller Qualitätsindikatoren mit ihrer Zuordnung zu den einzelnen Datenbeständen liegt derzeit noch nicht vor.

- Für die Abbildung der Qualitätsindikatoren von epidemiologischen Krebsregistern wäre eine Erweiterung des bisherigen Ansatzes der TMF-Leitlinie sinnvoll. Die dort aufgeführten Indikatordefinitionen sind Templates (z.B. „Anzahl fehlender Werte bei mandatorischen Datenelementen“), die für eine Anwendung instantiiert werden müssen (z.B. für eine bestimmte Tumorentität). Eine mögliche Erweiterung des Ansatzes könnte ähnlich wie bei den QSR-Indikatoren erfolgen; dort wird eine 2-dimensionale Matrix aus medizinischen Gegebenheiten (z.B. Herzinfarkt) und Indikatortemplates (z.B. Sterblichkeit im Krankenhaus) gebildet und eine Matrixzelle repräsentiert einen einzelnen Qualitätsindikator (z.B. „Sterblichkeit im Krankenhaus bei Aufnahme mit Herzinfarkt“). In unserem Falle benötigten wir mindestens folgende Dimensionen für eine solche Matrix:
  - Templates der Qualitätsindikatoren
  - Art des Datenbestandes
  - Typ der Meldestelle
  - Tumorentität
- Die Beschreibung des Qualitätsindikators „Vollständigkeit der Registerdaten“ (siehe TMF-Leitlinie: Abschnitt B.2.3.4) vermengt Aussagen zur **Vollzähligkeit** (d.h. Rate aller tatsächlich

erfassten PatientInnen bezogen auf die Anzahl aller theoretisch zu erfassenden PatientInnen) und zur **Vollständigkeit** der Daten. Deshalb sollte dieser Qualitätsindikator in zwei separate Indikatoren aufgespalten werden.

- In der folgenden Tabelle werden wichtige Qualitätsindikatoren bzw. Gruppen von Qualitätsindikatoren in Bezug zur TMF-Leitlinie betrachtet:

Qualitätsindikator / Qualitätsindikatorgruppe	Korrespondierender Qualitätsindikator in der TMF-Leitlinie (Abschnittsnummer, Name)	Anmerkungen	Einordnung
Anzahl fehlender Werte	B.2.3.6: „Anzahl fehlender Werte bei mandatorischen Datenelementen“		identisch
Vollzähligkeit	B.2.3.4: „Vollständigkeit der Registerdaten“	Anwendung auf Best-of-Datensatz; Indikatordefinition nur teilweise abgedeckt, da in der TMF-Leitlinie <b>Vollzähligkeit</b> und <b>Vollständigkeit</b> vermengt sind (s.o.)	feiner
Anzahl Meldungen pro Fall	---	Ebene Plausibilität (Strukturqualität)	ergänzend
Anteil der Personen mit bzw. ohne Follow-up (Todesmeldung, Abgleich mit Einwohnermeldeamt)	B.2.2.4: „Anzahl vorzeitig aus dem Register ausscheidender Patienten/Personen“		identisch
Anteil histologisch verifizierter Befunde bezogen auf alle Befunde (HV, MV)	B.2.3.5: „Übereinstimmung mit dem Studienplan“???	Indikator für die Vollzähligkeit für einzelne Tumorentitäten und das gesamte Register; Bezug zur Datenqualität und zur Qualität des diagnostischen Prozesses	feiner
Anteil der Fälle mit unbekanntem Primärtumor bezogen auf alle Fälle (PSU)	---	Ebene Richtigkeit (Ergebnisqualität)	ergänzend
Anteil der DCO-Fälle bezogen auf alle Fälle	B.2.3.4: „Vollständigkeit der Registerdaten“		identisch
Plausibilitätschecks	B.2.1.3: „Widerspruchsfreiheit“	Vergleich von 2 oder mehr Merkmalen; Anwendung auf einzelne	identisch

Qualitätsindikator / Qualitätsindikatorgruppe	Korrespondierender Qualitätsindikator in der TMF-Leitlinie (Abschnittsnummer, Name)	Anmerkungen	Einord- nung
		Meldung und Best-of-Datensatz; <u>zwei unterschiedliche Raten je Prüfung: Anteil Fehler</u> (d.h. Korrektur zwingend erforderlich) und <b>Anteil Warnungen</b> (d.h. unplausibel aber prinzipiell möglich)	
Gültigkeit	B.2.1.8: „Erlaubte Werte bei qualitativen Datenelementen“	Anwendung auf einzelne Meldung und Best-of-Datensatz; TMF-Leitlinie berücksichtigt nur <b>formale</b> Gültigkeit (z.B. existierender ICD-10-Kode)	feiner
Anteil Datenelemente mit spezifischer bzw. unspezifischer Angabe bezogen auf alle Datenelemente („Informationsdichte“)  <u>Ausprägungen:</u> <ul style="list-style-type: none"> <li>● Anteil nicht näher bezeichneter Uterus-Malignome bezogen auf alle Uterus-Malignome (Uterus NOS)</li> <li>● Histologie</li> <li>● Lokalisation</li> <li>● Diagnosen (ICD-10: xxx.9)</li> </ul>	---	Ebene Richtigkeit (Ergebnisqualität)	ergänzend

## 4.2 Teilbesprechung : SHIP

Arbeitspaket 1 –

Abbildung von Kennzahlen der Datenqualität aus der Study in Health in Pomerania (SHIP)

Treffen am 31.10.2011 in der Geschäftsstelle der TMF

Teilnehmer: Schmidt, Stausberg, Drepper (zeitweise)

Folgende Unterlagen von SHIP stehen zur Verfügung:

1. Checkliste SHIP Qualitätssicherungsberichte für Qualitätsverantwortliche und Bereichsleiter  
Version 1.6 – 10.2011
2. DM-SHIP. Standard Operating Procedure (SOP). Datenmanagement in SHIP-2 / SHIP-TREND – Teil  
II Data Validation + Documentation. V1.6-E vom 24.5.2011.
3. Qualitätssicherungsbericht zur Somatometrie für den Zeitraum vom 01.03.2011 - 31.05.2011
4. Schmidt CO, Westphal S. Qualitätsindikatoren von SHIP. Ein Überblick, orientiert an der  
Kategorisierung der TMF-Leitlinie Datenqualität in der medizinischen Forschung. 10.2011.

Letztgenannter Bericht wird als Projektunterlage aufgenommen.

Die unterschiedlichen Zugänge in SHIP und in der Leitlinie lassen sich - unter der Voraussetzung, dass bislang spezifische Verweise auf Register allgemeiner gestaltet werden - weitgehend , auf das auf Seite 3 der Leitlinie abgebildete generische Rahmenmodell zur Datenqualität von Aarts et al. (2002a) abbilden. Die Leitlinie greift aus diesem Rahmenmodell Site Visits für einen Originaldatenabgleich, die Durchführung von Analysen zur Datenqualität mit Indikatoren, die Erstellung von Feedback-Berichten zur Datenqualität sowie ein Feedback zu den Site Visits heraus.

Die Leitlinie befasst sich mit der Bewertung der Datenqualität ganzer Studien oder einzelner Erhebungszentren. Liegen Erhebung, Erfassung und Datenmanagement in einer Hand sind andere Zugänge sinnvoll.

### Teil 1

Die Qualitätskennzahlen sind in der Unterlage 2 in Abschnitt 4.7 aufgeführt. Es erfolgt eine Bewertung der Abbildbarkeit gemäß Projektantrag mit den Kategorien gröber, identisch, feiner, ergänzend, nicht abzubilden (Definitionen s. Projektantrag). Aus der Durchsicht ergeben sich folgende Ergebnisse.

4.7.1. - irrelevant

4.7.2. - Vollständigkeit der Merkmale auf Modulebene (=Unit): Dies lässt sich mit B.2.1.6 sowie B.2.3.6 zu einer Superklasse zusammenführen. B.2.1.6 sowie B.2.3.6 werden dann ebenso wie die Kennzahlen von SHIP zu Subklassen. [Bewertung: gröber]

Füllung der Felder von fehlenden Modulen mit qualifizierten Missings: Hierbei handelt es sich um eine Subklasse von B.2.1.8. [Bewertung: feiner]

4.7.3. - Übereinstimmung des Erhebungsinstrumentes mit den Metadaten: Wird derzeit nicht abgebildet, da von der Leitlinie Metadaten und Erfassungsinstrument gleichgesetzt wurden. Vor dem Hintergrund des MDR-Projektes kann eine Aufnahme in die Leitlinie sinnvoll sein. [Bewertung: nicht abzubilden]

4.7.4. - Die Beachtung von Abhängigkeiten wird von der Leitlinie als Widerspruchsfreiheit bezeichnet und mit B.2.1.3 abgebildet. Bezüglich Missings wird auf die Erläuterung zu 4.7.2 hingewiesen. [Bewertung: identisch]

4.7.5. Füllung fehlende Felder mit qualifizierten Missings: Hierbei handelt es sich um eine Subklasse von B.2.1.8. (s. auch Ausführungen zu 4.7.2) [Bewertung: feiner]

4.7.6. - nicht relevant

4.7.7 - Anteil der Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten: von der Leitlinie nicht abgebildet; sollte neu aufgenommen werden. [Bewertung: ergänzend]

In den interessierenden Fällen sollen die entsprechenden Fehlwerte eingesetzt werden (ähnlich Missings): Hierbei handelt es sich um eine Subklasse von B.2.1.8. Die Nähe zu B.2.1.7 ist zu beachten. [Bewertung: feiner]

4.7.8 - Die Prüfung von Frequencies findet sich in der Leitlinie in B 2.1.7 und B.2.1.8. [Bewertung: gröber]

4.7.9 - nicht relevant

4.7.10 - nicht relevant

Aus der Diskussion der Rekrutierungsrate wird deutlich, dass die Leitlinie um eine „Verweigerungsrate“ von Modulen, Untersuchungen und Items ergänzt werden sollte. Dies könnte über einen Eintrag „abgelehnt“ in den entsprechenden Feldern berechnet werden. Eine ähnliche Betrachtung könnte auch für nicht mögliche Angaben („weiß nicht“) erfolgen. [Bewertung: ergänzend]

## **Teil 2**

An Hand der Somatometrie werden weitere Aspekte zu Kennzahlen aus dem Qualitätssicherungsbericht erarbeitet.

Angaben zum Einschluss lassen sich in B.2.2.3 und B.2.2.4 abbilden. Zusätzlich ist eine differenziertere Verweigerungsrate sinnvoll (s. oben).

Die Vollständigkeit der Daten ist über B.2.1.6 und B.2.3.6 abgebildet.

Extremwerte finden sich in der Leitlinie bei B.2.1.7.

Bei technischen Untersuchungen betrachtet SHIP die Qualität getrennt für das Gerät, den verantwortlichen Techniker (Observer) sowie die befundende Person (Reader). Dies lässt sich formal zum Teil unter B.2.1.4 abbilden. Eine weitere Differenzierung ist allerdings sinnvoll.

Zeittrends lassen sich unter B.2.1.4 einordnen. B.2.1.4 ist aber zu abstrakt, um allen möglichen Verteilungsprüfungen gerecht zu werden. B.2.1.4 sollte neu gefasst werden.

Schulung: B.2.2.5 „Fachliche Qualifikation“ befasst sich mit einer von den restlichen Indikatoren abweichenden Ebene. Ein Verzicht auf diesen Indikator sollte diskutiert werden.

### **Teil 3**

Die Unterlage 4 wird diskutiert.

Aus den Leitlinien sollten die Ebenen von Qualität (z. B. Struktur) nach Donabedian herausgenommen werden, da diese befremdlich wirken.

Konkordanz (B.2.1.2) und Genauigkeit (B.2.3.1) sind klarer abzugrenzen.

B2.1.4 sollte differenzierter gestaltet werden, um verschiedene, in einer Kohortenstudie relevante Aspekte besser abbilden zu können.

B.2.1.5 ist im Niveau zu spezifisch. Dieser Indikator sollte in eine Subklasse oder sogar in eine Instanz von B.2.1.3 überführt werden.

B.2.1.9 ist ebenfalls zu spezifisch. Die Bevorzugung bestimmter Endziffern kann als ungewöhnliche Werteverteilung aufgefasst werden. Dann wäre dies eine Subklasse oder Instanz von B.2.1.4.

B2.2.3 Von der Rekrutierungsrate sollte die Subresponse zu einzelnen Untersuchungsmodulen abgehoben werden.

B 2.2.5 Die fachliche Qualifikation des Datenerfassungspersonals fügt sich nicht in Leitlinie ein.

B.2.3.4 „Vollständigkeit“ ist von B.2.1.6 und B.2.3.6 abzugrenzen.

B.2.3.5 „Übereinstimmung mit dem Studienplan“ sollte präzisiert werden.

### **Teil 4 - allgemeine Anmerkungen**

Der oben bereits angedeutete unterschiedliche Abstraktionsgrad der Indikatoren sollte angeglichen werden.

Es ist sinnvoll, Empfehlungen für die Angemessenheit von Indikatoren in Bezug auf Datenbestände bestimmter Studientypen (Kohortenstudien, Registerstudien, etc.) zu formulieren.

### 4.3 Teilbesprechung : OpEN.SC

Arbeitspaket 1

Abbildung von Kennzahlen der Datenqualität aus OpEN.SC

Treffen am 18.11.2011 in der Charite

Teilnehmer: Hanß, N.N., Schrader, Stausberg

Folgende Unterlagen von OpEN.SC stehen zur Verfügung:

- a) Als Tischvorlage dieses Dokument unterhalb der Linie
- b) Schrader T, Zhou Y, Hanß S. Open European Nephrology Science Center – Datenqualitätskriterien

Das Ergebnis der Diskussion ist im [Überarbeitungsmodus](#) eingefügt.

## Completeness:

- PopulationCompleteness [In Deutschland üblicherweise als Vollzähligkeit bezeichnet. Bildet die Leitlinie über die Rekrutierungsrate ab \(B.2.2.3\). Mapping: identisch](#)

Member	data	comment	Umsetzung	Anmerkungen	Role
NumberOfAllClinicalCasesOfAResource	Alle Properties	Number of all clinical cases related to a domain calculated for a resource.	Transferproperty File - Definition	In unserem Fall: Wir bekommen alle Fälle, die im Quellsystem sind	
NumberOfOpenscCasesOfAResource	Alle Properties	Number of OpEN.SC Cases concerning one resource	Transferproperty File - Definition	Bei Datenintegration	
RatioOpenscCases2AllClinicalCasesOfAResource	Alle Properties	Ratio of OpEN.SC cases to all clinical cases to express the completeness of nephrology cases.	Transferproperty File - Definition	Bei Datenimport	

- SchemaCompleteness - [Bezieht sich auf jede Beobachtungseinheit, d. h. es liegt ein Ergebnis pro Patient vor. Damit bestimmt dieser Bereich die Vollständigkeit pro Beobachtungseinheit. Die Leitlinie deckt dies über B.2.1.6 und B.2.3.6 ab. Bei B.2.3.6 wird alternativ bereits die Berechnung der Vollständigkeit pro Person/Beobachtungseinheit benannt. Damit stellt dieser Komplex eine Verfeinerung dar. Mapping: feiner.](#)

Member	data	comment	Umsetzung	Anmerkungen	Role
NumberOfClinicalAttributes	Alle Properties	Count the number of attributes concerning a specific condition	Datenbankcheck	Anzahl der Attribute, die ein Patient entsprechend der quellenbezogenen Property-Datei mitbringt	
NumberOfJointlyUsedProperties	Alle Properties	Count of all jointly used properties of the resources	Datenbankcheck	Anzahl der Properties, zu denen von ALLEN <a href="#">Beobachtungseinheiten</a> Informationen vorliegen	<a href="#">Mapping „ergänzend“: wird auf Datenbestand bezogen berechnet, die Raten befinden sich in a) und b); passt in die Ebene der Organisation. Für die Menge der hier</a>

					<a href="#">gewerteten Variablen ist die Vollständigkeit 100 %.</a>
NumberOfOpencProperties	Alle Properties	Count the number of different properties in the OpEN.SC Schema	Ist bekannt, ändert sich bei jeder neuen Quelle	Anzahl der Properties in OpEN.SC Datenbank	
RatioClinicalAttributes2OpencProperties	Alle Properties	Express the ratio between all attributes in the resource database and the used OpEN.SC properties.			<a href="#">Dies wäre dann pro Patient die Zahl der Variablen mit Information im Verhältnis zur Zahl aller verfügbaren Variablen.</a>
RatioJointlyProperties2AllPropertiesForAllResources	Alle Properties	ratio calculated for all resources			<a href="#">a) komplett</a>
RatioJointlyProperties2AllPropertiesForAResources	Alle Properties	ratio calculated for one resource			<a href="#">b) auf Teildatenbestand bezogen</a>

- [MetadataCompleteness: Dies ist nur dann zu berechnen, wenn sich die Anzahl der Variablen bei Messungen/Befunden der gleichen Untersuchung unterscheiden kann \(z. B. bei einem XML-Format\). Würde man dies auf die Werte übertragen, entspricht der Komplex der Vollständigkeit bezogen auf Untersuchungen und lässt sich als Verfeinerung von B.2.1.6 und B.2.3.6 auffassen.](#)
- [Allerdings ist die Nutzung als Qualitätskennzahl fraglich. So könnte der Verzicht auf die Übertragung von Variablen ohne Inhalt durchaus als gute Qualität im Sinne eines effizienten Datenaustausches bezeichnet werden. Die dem zu Grund liegende Unvollständigkeit wäre dann über einen anderen Indikator abgedeckt.](#)
- [Die hier vorgenommene Betrachtung der Qualität von Metadaten wäre neu und wird von der Leitlinie derzeit nicht abgedeckt. Zu diskutieren ist allerdings noch, ob es sich tatsächlich um Datenqualität handelt oder um ein davon zu unterscheidendes Konstrukt.](#)
- [Verwandtschaft besteht zu Indikator B.2.3.5 „Übereinstimmung mit dem Studienplan“.](#)
- [Mapping: nicht abzubilden.](#)

<b>Member</b>	<b>Data</b>	<b>comment</b>	<b>Umsetzung</b>	<b>Anmerkungen</b>	<b>Role</b>
NumberOfUsedMetadata RelatedToObservation	Metadata		Datenbankcheck	Anzahl der Metadaten bezogen auf eine Untersuchung oder Bild - bei Untersuchung - Datum, Einheit, Code, Spezifikation der Untersuchungsmethode - beim Bild - Anzahl der verwendeten Metadaten	
NumberOfJointlyUsedMe tadataInObservations	Metadata		Datenbankcheck	Anzahl der gemeinsam genutzen Metadaten in allen Observations	
RatioOfUsedMetadata2N umberObservations	Metadata		Datenbankcheck	Durchschnitt der Anzahl verwendeter Metadaten - Information darüber, wie gut oder schlecht überhaupt Metainformationen vorliegen	

Member	Data	comment	Umsetzung	Anmerkungen	Role
NumberOfAllConsultedCases			Telekonsultationsprozess		Studien <a href="#">Glaubwürdigkeit der Daten durch Absicherung von Diagnosen. Nicht Vollständigkeit.</a> <a href="#">Bei Betrachtung des Telekonsultationsprozesses als Modul ließe sich dies als Verfeinerung der Rekrutierungsrate (B.2.2.3) abbilden.</a>
NumberOfAllTerms		number of all medical terms		Differenzierbare medizinische Begriffe wie z.B. Diagnosen	
NumberOfConsens			Telekonsultationsprozess		<a href="#">Ein Dissenz lässt sich als Widerspruch in den Daten auffassen. Dann entspricht dies B.2.1.3.</a>
NumberOfMajorDissens		The difference between two opinions has a great impact.	Telekonsultationsprozess		<a href="#">Mapping: feiner als B.2.1.3</a>
NumberOfMinorDissens		First opinion and second opinion have only a difference of minor importance.	Telekonsultationsprozess		<a href="#">Mapping: feiner als B.2.1.3</a>
NumberOfStandardMatchedTerms2AStandard	ICD10, OPS, Labordaten, Rote Liste, Datum?	Number of terms which are matched to a standard		Siehe Statistik von Bea, Dietmar (Length(ICD10Code)<3 wird verworfen,	<a href="#">Überschneidung mit B.2.1.8, wobei in der Leitlinie von Vorgaben hinsichtlich der Einträge</a>

				fDInDom, Statistik?), P_ImplausibleDataLog	<a href="#">ausgegangen wird.</a> <a href="#">Hierüber wird bei</a> <a href="#">Open.SC nichts</a> <a href="#">ausgesagt.</a> <a href="#">Mapping: gröber</a>
RatioOfStandardTerms2AI Terms	Alle Properties		Einfacher Datenbankcheck (Dietmar?)	Terms oder properties? Test, wie viel überhaupt kodiert vorliegt – von wievielen Feldern könnte ein Code vorliegen, von wie vielen liegt er tatsächlich vor (egal richtig oder falsch)	<a href="#">B.2.1.8</a> <a href="#">Mapping: identisch</a>

**Interpretability:** Wird nicht gesondert behandelt.

<b>Member</b>	<b>Data</b>	<b>comment</b>	<b>Umsetzung</b>	<b>Anmerkungen</b>	<b>Role</b>
NumberOfAllTerms		number of all medical terms		Differenzierbare medizinische Begriffe wie z.B. Diagnosen	
NumberOfStandardMatchedTerms2AStandard		Number of terms which are matched to a standard			
NumberOfStandardMatchedTerms2Standards		Number of terms which are matched to more than one standard			
RatioOfStandardTerms2AllTerms					

## Timerelation:

- RegularityOfObservations [Es handelt sich hier um datumsbezogene Angaben.](#)

Member	Data	comment	Umsetzung	Anmerkungen	Role
AverageDistanceOfTimeBetweenObservations	Alle datumsbezogenen Properties				Musteranalyse <a href="#">Verfeinerung von B.2.1.4</a>
CountOfObservationsWithSameTestSet	Alle datumsbezogenen Properties	A factor wich describes the regulary application of the same testset for laboratory evaluations			Musteranalyse <a href="#">Hat mit Datum eigentlich nichts zu tun. Alle folgenden Zahlen beziehen sich auf eine Beobachtungseinheit. Entspricht der Verteilung von Variablen, die mehrfach bei einem Patienten auftreten können. Damit ließe sich dies als Verfeinerung von B.2.1.4 interpretieren. Alternativ kann an eine Trennung zwischen Kennzahlen bezogen auf ein Kollektiv und Kennzahlen bezogen auf einzelne Beobachtungseinheiten gedacht werden.</a>
FrequencyDistributionOfTimeBetweenObservations	Alle datumsbezogenen Properties	Distribution of time distances between lab evaluations as descriptive property of regularity of lab evaluations		#Daten regelmäßig ausgefüllt/gesamte # Daten, die regelmäßig ausgefüllt werden sollen	Musteranalyse <a href="#">s.o.</a>
FrequencyDistributionOf	Alle datumsbezogenen	Frequency distribution of			Musteranalyse

TimeBetweenObservationsRelatedToEvents	Properties	time between lab evaluations related to events An event is a new hospitalization related to any reason.			<a href="#">S.O.</a>
MaximumDistanceOfTimeBetweenObservations	Alle datumsbezogenen Properties				Musteranalyse <a href="#">S.O.</a>
MedianDistanceOfTimeBetweenObservations	Alle datumsbezogenen Properties				Musteranalyse <a href="#">S.O.</a>
MinimumDistanceOfTimeBetweenObservations	Alle datumsbezogenen Properties				Musteranalyse <a href="#">S.O.</a>
ModalDistanceOfTimeBetweenObservations	Alle datumsbezogenen Properties				Musteranalyse <a href="#">S.O.</a>

- Timeliness [Hier geht es um die Abstände bzw. Intervalle, in denen Daten oder Beobachtungseinheiten geändert oder neu aufgenommen werden. Dieser Komplex hat Überlappung mit der Aktualität \(B.2.2.1\) und der Übereinstimmung mit dem Studienplan \(B.2.3.5\). Dies wird bei Open.SC allerdings nicht als Qualitätsmerkmal der Erhebung sondern vielmehr des Datenimports verstanden, da Open.SC vorhandene Daten speichert. Eine aktive Erhebung findet bei Open.SC ja nicht statt.](#)
- [Open.SC lässt sich als Data Repository auffassen. Dadurch können Unterschiede zwischen dem Ansatz von Open.SC und den Überlegungen der Leitlinie erklärt werden. Kennzahlen, die auf die Erhebung bezogen sind, spielen für Open.Sc eine untergeordnete Rolle, da Open.SC keine eigenen Daten erhebt. Hingegen sind für die Bewertung der Qualität des Data Repository andere Aspekte relevant, die z. B. den Kontext der Datenübernahme von Quellsystemen abbilden.](#)

Member	Data	comment	Umsetzung	Anmerkungen	Role/Use Case
AvgCasesUpdated			Datenimport		
AvgFrequencyUpdated			Datenimport		
CasesUpdatedPerUpdate			Datenimport		
FrequencyOfUpdate			Datenimport		