

Aus dem Institut für Medizinische Biometrie und Informatik der Universität Heidelberg
(Geschäftsführender Direktor: Prof. Dr. sc. hum. Meinhard Kieser)

MECHANISMEN ZUR AUTOMATISIERTEN QUALITÄTSSICHERUNG UND VALIDIERUNG DER DATEN IN MEDIZINISCHEN REGISTERN

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)
an der
Medizinischen Fakultät Heidelberg
der
Ruprecht-Karls-Universität

vorgelegt von
LENNART KÖSTER
aus
Heidelberg

2018

Dekan: Herr Prof. Dr. med. Wolfgang Herzog

Doktormutter: Frau Prof. Dr. sc. hum. Petra Knaup-Gregori

Inhaltsverzeichnis

1	Einleitung	1
1.1	Gegenstand und Motivation	1
1.2	Problemstellung	2
1.3	Zielsetzung	2
1.4	Aufbau der Dissertation	3
2	Grundlagen	5
2.1	Cooperative European Paediatric Renal TransplAnt INitiative (CERTAIN) Registry	5
2.2	Cooperative EuROpean Paediatric TransplAnt INitiative LIver (CERTAIN-LI) Registry	6
2.3	Kinderrheumatologie Online - Kerndokumentation (KRhOKo)	7
2.4	CERTAIN-Registerplattform	8
2.4.1	Architektur	9
2.4.2	Datenmodell	10
2.4.3	Datenqualitätsmanagementprozess	12
2.5	Java Bean Validation	13
2.6	Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.	16
2.7	Leitlinie Datenqualität	16
2.7.1	Vorgehen bei der Anwendung	18
2.7.2	Qualitätsindikatoren	18
2.7.3	Qualitätsscore	45
2.7.4	Registerempfehlungen	46
2.8	Verwandte Arbeiten	46
3	Methoden und Werkzeuge	49
3.1	Entwicklung des Ansatzes	49
3.2	Analyse der automatischen Validierung des CERTAIN-Ansatzes	49
3.3	Entwicklung des allgemeinen Vorgehens	52
3.4	Implementierung von Modulen zur Qualitätsprüfung	53
3.5	Auswertung der Effektivität der neuen Validierungsmodule	53
4	Überprüfung des Ansatzes der automatischen Validierung	57
4.1	Überprüfung der Digitalisierung der Fragebogen	57
4.2	Überprüfung des Ansatzes der automatischen Validierung	58
4.3	Überprüfung der Vollständigkeit der Fragebogen	59

5	Vorgehen zur Verbesserung der Datenqualität in Registern	61
5.1	QI-Kategorien	61
5.2	Auswahl der relevanten QI	63
5.3	Priorisierung der Umsetzung der QI	64
5.4	Festlegung der Gewichte der QI	64
5.5	Allgemeines Vorgehen zur Anwendung des Ansatzes	66
6	Registerempfehlung CERTAIN/-LI und KRhOKo	67
6.1	Vergleich der Register	67
6.2	Technische Anforderungen	68
6.3	Analyse der TMF-Qualitätsindikatoren im Rahmen der CERTAIN-Register	69
6.3.1	Qualitätsindikatoren der Ebene Integrität	69
6.3.2	Qualitätsindikatoren der Ebene Organisation	109
6.3.3	Qualitätsindikatoren der Ebene Richtigkeit	124
6.4	Zusammenfassung der Analyse der QI	131
6.5	Gewichte der QI für die CERTAIN-Register	132
6.6	Vorlage für elektronische Register	132
7	Implementierung	134
7.1	Auswahl der umzusetzenden Validierungsregeln	134
7.2	Etablierung von Warnungen	135
7.3	Nutzung von Annotationen und Java Reflection	135
7.4	Verhinderung von Synonymen	136
7.5	Prüfung auf Übereinstimmung mit Vorwert für variable Datenwerte	137
7.6	Prüfung auf Übereinstimmung mit Vorwert für invariable Datenwerte	138
7.7	Reihenfolge der fortlaufenden Medikationsangaben	139
7.8	Häufigkeit der fortlaufenden Medikationsangaben	141
7.9	Prüfung auf fehlende Datenwerte bei optionalen Datenelementen	142
7.10	Validierung von Wertebereichen	143
7.11	Aktualität der gespeicherten Daten	146
7.12	Rekrutierungsrate	147
7.13	Test der Implementierung	148
7.14	Auswertungsmethoden	149
8	Datenqualität des CERTAIN-Registers	152
8.1	Datenqualität des Datenbestandes D1	152
8.2	Prüfung des Datenbestandes D1 gegen die Validierungsregeln V2	153
8.3	Datenqualität Datenbestand D2 nach Durchführung der Korrekturen geprüft gegen Validierungsregeln V2	157

8.4	Datenqualität ein Jahr nach Etablierung der Validierungsregeln V2	159
8.5	Validierungsergebnis der Validierungsregeln V3	160
8.6	Vergleich der Subgruppen der Visiten des Datenbestandes D3	161
8.7	Inkompatible Einheiten	164
8.8	Korrekturen von Ausreißern bei numerischen Datenelementen	165
8.9	Clustering unterhalb der Validierungsgrenzen von numerischen Datenelementen	166
9	Diskussion und Ausblick	170
9.1	Diskussion der Ergebnisse	170
9.2	Diskussion der Methoden	177
9.3	Ausblick	180
10	Zusammenfassung	183
	Glossar	184
	Literatur	186
	Eigene Veröffentlichungen	195
	Anhang	196
A	Zusammenfassung QI-Analyse der CERTAIN-Register	196
B	QI-Empfehlung für elektronische Register	201
	Lebenslauf	206
	Danksagung	207

Abbildungsverzeichnis

1	CERTAIN-Serverarchitektur (Köster et al. 2015)	9
2	Projektstruktur CERTAIN 2014	10
3	Projektstruktur CERTAIN-Registerplattform	11
4	Ausschnitt Klassendiagramm der drei Register	12
5	Datenqualitätsmanagementprozess der CERTAIN-Register	13
6	Ausschnitt Klassendiagramm ConstraintValidator	14
7	Adaptives Monitoring der Datenqualität nach (Nonnemacher et al. 2014)	17
8	Prozeduren zur Sicherung der Datenqualität in medizinischen Registern (Arts et al. 2002)	47
9	Entwicklung der allgemeinen und registerspezifischen QI-Empfehlungen	52
10	Auswertung der Effektivität der Validierung	55
11	Zusammenfassung Ansatz zur Ermittlung der QI Gewichte	66
12	Erweitertes allgemeines Vorgehen	133
13	Ausschnitt Klassendiagramm ContinuousValuesValidator	138
14	Ausschnitt Klassendiagramm InvariableValueValidator	139
15	Erlaubte Abfolge Medikationseinträge	140
16	Ausschnitt Klassendiagramm EntryOrderValidator	141
17	Ausschnitt Klassendiagramm SingleOrderedEntryPerDayValidator	142
18	Ausschnitt Klassendiagramm NotEmptyOptionalValidator und NotNullOptionalValidator	142
19	Ausschnitt Klassendiagramm RangeValidator und EntryRangeValidator	144
20	Aktualität der Daten für CERTAIN Einrichtung Heidelberg im November 2017	146
21	Globale Aktualität der Daten des CERTAIN-Registers November 2017	147
22	Rekrutierungsrate für CERTAIN Einrichtung Heidelberg im November 2017	148
23	Globale Rekrutierungsrate der ET-Einrichtungen des CERTAIN-Register im November 2017	148
24	Aggregierte Kalium-Werte des CERTAIN-Registers November 2017	150
25	Aggregierte Protein-Werte des Datenbestandes D1	156
26	Lokale Maxima der Dichtefunktion	167
27	Lokale Maxima der Dichtefunktion in Nähe der Validierungsgrenzen	168

Tabellenverzeichnis

1	Basis-Constraints der <i>Bean Validation</i> (Morling 2017)	15
2	Dokumente mit Fehlern durch den Digitalisierungsprozess	57
3	Fehleranzahl im Digitalisierungsprozess aufgeteilt nach Fehlertyp	58
4	Validierungsergebnis der digitalisierten Fragebogen	58
5	Vollständigkeit der Patientenbogen	59
6	QI-Kategorien des entwickelten Vorgehens	63
7	Bewertung TMF-1001 - Übereinstimmung mit Vorwert	71
8	Bewertung TMF-1002 - Konkordanz	73
9	Bewertung TMF-1003 - Widerspruchsfreiheit	75
10	Bewertung TMF-1035 - Endlosüberlebende	77
11	Bewertung TMF-1004 - Sichere Widersprüche/Fehler	79
12	Bewertung TMF-1005 - Mögliche Widersprüche/Warnungen	80
13	Bewertung TMF-1006 - Werteverteilung	82
14	Bewertung TMF-1007 - Bevorzugung bestimmter Endziffern	83
15	Bewertung TMF-1009 - Werteverteilung der durch Untersucher erfassten Parameter	84
16	Bewertung TMF-1010 - Werteverteilung der durch Geräte erfassten Parameter	84
17	Bewertung TMF-1011 - Werteverteilung von Befunden	85
18	Bewertung TMF-1052 - Werteverteilung von Parametern zwischen Zentren	86
19	Bewertung TMF-1008 - Untersuchungen am Wochenende	88
20	Bewertung TMF-1012 - Fehlende Module	90
21	Bewertung TMF-1013 - Fehlende Werte bei Datenelementen	91
22	Bewertung TMF-1014 - Fehlende Werte bei mandatorischen Datenelementen	93
23	Bewertung TMF-1015 - Fehlende Werte bei optionalen Datenelementen . .	94
24	Bewertung TMF-1016 - Datenelemente mit Wert unbekannt o.ä.	95
25	Bewertung TMF-1017 - Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten	97
26	Bewertung TMF-1018 - Ausreißer bei stetigen Datenelementen	99
27	Bewertung TMF-1019 - Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten	99
28	Bewertung TMF-1020 - Werte aus Standards	101
29	Bewertung TMF-1021 - Unerlaubte Werte bei qualitativen Datenelementen	102
30	Bewertung TMF-1022 - Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings	104
31	Bewertung TMF-1023 - Unerlaubte Werte zur Kodierung von fehlenden Mo- dulen	105

32	Bewertung TMF-1024 - Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen	105
33	Bewertung TMF-1025 - Datenelemente mit unspezifischen Werten	106
34	Bewertung TMF-1026 - Beobachtungseinheiten mit unbekanntem Primärtumor	108
35	Bewertung TMF-1027 - Nachweis bekannter Korrelationen	109
36	Bewertung TMF-1050 - Umfang der Metadaten bei Untersuchungen	109
37	Bewertung TMF-1028 - Aktualität der gespeicherten Daten	111
38	Bewertung TMF-1029 - Dubletten im Datenbestand	112
39	Bewertung TMF-1030 - Rekrutierungsrate	114
40	Bewertung TMF-1051 - DCO-Rate (Death Certificate Only)	114
41	Bewertung TMF-1031 - Verweigerungsrate von Untersuchungen	114
42	Bewertung TMF-1032 - Verweigerungsrate von Modulen	115
43	Bewertung TMF-1033 - Verweigerungsrate von einzelnen Datenelementen	115
44	Bewertung TMF-1034 - Vorzeitig aus dem Register ausscheidende Beobachtungseinheiten	117
45	Bewertung TMF-1036 - Synonyme	118
46	Bewertung TMF-1037 - Homonyme	119
47	Bewertung TMF-1038 - Einzelmeldungen pro Beobachtungseinheit	119
48	Bewertung TMF-1039 - Solitäre Meldungen von Pathologen	120
49	Bewertung TMF-1040 - Zurückgewiesene Meldungen	122
50	Bewertung TMF-1041 - Datenquellen pro Beobachtungseinheit	123
51	Bewertung TMF-1042 - Beobachtungseinheiten mit Follow-up	124
52	Bewertung TMF-1043 - Genauigkeit der Registerdaten	125
53	Bewertung TMF-1044 - Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente	126
54	Bewertung TMF-1045 - Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten	127
55	Bewertung TMF-1046 - Vollständigkeit der Registerdaten	128
56	Bewertung TMF-1047 - Übereinstimmung mit Verfahrensregeln	130
57	Bewertung TMF-1048 - Repräsentativität der Registerdaten	131
58	Validierungsergebnis Datenbestand D1 geprüft gegen Validierungsregeln V1	153
59	Validierungsergebnis Datenbestand D1 geprüft gegen Validierungsregeln V2	154
60	Validierungsergebnis Datenbestand D2 geprüft gegen Validierungsregeln V2	157
61	Validierungsergebnis Datenbestand D2 geprüft gegen Validierungsregeln V3	158
62	Validierungsergebnis Datenbestand D3 geprüft gegen Validierungsregeln V3	159
63	Validierungsergebnis Datenbestand D2 geprüft gegen in dieser Dissertation entwickelte Validierungsregeln	160

64	Validierungsergebnis Datenbestand D3 geprüft gegen in dieser Disseration entwickelte Validierungsregeln	160
65	Validierungsergebnis Datenbestand D3 nur Visiten berücksichtigt mit letzter Bearbeitung vor dem 17.11.2016	162
66	Validierungsergebnis Datenbestand D3 nur Visiten mit letzter Bearbeitung nach 17.11.2016	162
67	Validierungsergebnis Datenbestand D3 nur Visiten erstellt in Registerversion R2 oder R3	163
68	Validierungsergebnis Datenbestand D3 nur Visiten erstellt nach 01.07.2017 .	163
69	Inkompatible Einheiten Datenbestand D2	164
70	Inkompatible Einheiten Datenbestand D3	164

1 Einleitung

1.1 Gegenstand und Motivation

Seltene Erkrankungen sind eine Herausforderung für die klinische Forschung, da aufgrund der geringen Fallzahl und Erkrankungsraten oftmals nur wenige Patienten für Studien zur Verfügung stehen (Tönshoff et al. 2015). Um die Situation dennoch durch Forschung verbessern zu können, ist es notwendig, die Population der Erkrankten systematisch zu beobachten und ihre Daten ganzheitlich zu erfassen, um die geringe Anzahl der Fälle standardisiert zu dokumentieren und anhand dieser Dokumentation die Wirksamkeit von Therapien zu beurteilen, neue Therapieformen zu entwickeln und alte verbessern zu können (Bellgard et al. 2013; Gliklich et al. 2014). Eines der Instrumente, um Erkrankungen, Therapien oder Populationen zu dokumentieren, sind klinische Register, die den Verlauf einer Erkrankung oder Therapie abbilden (Lim et al. 2008). Ein Register ist eine systematische Sammlung von medizinischen und demographischen Daten zu einer Patientengruppe mit bestimmten Charakteristika (Gliklich et al. 2014; Veer et al. 2010). Die Aussagekraft der Daten eines Registers hängt stark von der Qualität der Daten ab und somit von Mechanismen, welche die Qualität absichern sollen (Bray und D. M. Parkin 2009; Couchoud et al. 2013). Bei multizentrischen Registern kann die Qualität der Daten aus unterschiedlichen Einrichtungen stark schwanken. Dieser Effekt verstärkt sich noch bei internationalen Registern aufgrund der geographischen Verteilung der Einrichtungen und etwaiger Sprachbarrieren, welche eine Kontrolle der Erhebungsmechanismen und Datenqualität vor Ort einschränken. Sekundäre Datenquellen, wie beispielsweise andere Register, stellen für das Datenqualitätsmanagement eine zusätzliche Herausforderung dar.

Um den Therapieverlauf bei pädiatrischen Nierentransplantationen besser verfolgen zu können, entstand 2010 unter der Führung der Gesellschaft für pädiatrische Nephrologie ein Register für pädiatrische Nierentransplantationen, das so genannte CERTAIN-Register (Cooperative European Paediatric Renal TransplAnt Initiative (Köster et al. 2015; Plotnicki et al. 2013; Tönshoff et al. 2015)). Auf der technischen Basis dieses Registers wurden 2014 zwei weitere Register ins Leben gerufen, ein pädiatrisches Lebertransplantations- (CERTAIN-LI) und ein pädiatrisches Rheumaregister (KRhOKo). Die drei Register teilen sich eine Codebasis und bieten neben demselben Rollen- und Benutzer- auch dasselbe Qualitätssicherungskonzept (CERTAIN-Registerplattform). Die Qualitätssicherung beruht hierbei auf einer automatischen Prüfung der eingegebenen Daten gegen vordefinierte Regeln, welche die eingegebenen Daten auf Plausibilität prüfen. Verletzungen der Regeln werden direkt an die Eingabekräfte gemeldet. Erst wenn keine Verletzungen der Regeln mehr vorliegen, werden die eingegebenen Daten in einem zweiten Schritt durch einen Datenqualitätsmanager geprüft und akzeptiert.

1.2 Problemstellung

Während der Auswertungen der Daten des Nierentransplantationsregisters kam es trotz der bestehenden Datenqualitätssicherungsmaßnahmen vereinzelt zu Ungereimtheiten, welche Forschungsvorhaben durch aufwändige Rückfragen verzögerten. Daher sollen in die CERTAIN-Registerplattform weitergehende, rechnerbasierte Mechanismen für komplexere Qualitätsprüfungen implementiert werden, die insbesondere auch eine größere Flexibilität erlauben. Da klinische Register noch nicht so stark regulatorisch geregelt sind wie klinische Studien, liegen bisher auch nur vereinzelte Forschungsergebnisse zu Qualitätsindikatoren (QI) für Register vor. Ein Beispiel ist die Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern der TMF (Nonnemacher et al. 2014). Diese schlägt 51 QI vor, die durch eine Recherche der Literatur zu Qualitätsindikatoren erfasst und für die Leitlinie ausgewählt wurden. Eine Vorabanalyse des Leitfadens zeigte, dass dieser einen adäquaten Grundstein für eine verbesserte Validierung in den CERTAIN-Registern darstellen kann und in einigen Bereichen jedoch noch Spielraum für weitere Verbesserungen zulässt. Außerdem wurde deutlich, dass für die Bereiche der Transplantations- und Rheumamedizin Anpassungen und eine gezielte Auswahl der Kriterien erforderlich sind. Ausgehend von den Ergebnissen der Vorabanalyse muss aus den 51 QI eine gezielte Auswahl der umzusetzenden QI getroffen werden. Auf Basis dieser Auswahl muss entschieden werden, welche Kriterien den größten Nutzen erwarten und welche sich technisch in der CERTAIN-Architektur umsetzen lassen. Die TMF-Leitlinie bietet keine Unterstützung, wie diese QI in einem Register technisch implementiert werden können. Die Auswahl der QI ist zudem auch für die Berechnung eines Gesamtscores relevant, der die Datenqualität eines Datenbestandes wiedergibt. Der Gesamtscore wird anhand der ausgewählten QI berechnet, denen für diese Berechnung ein Gewicht zugewiesen werden muss. Die Leitlinie enthält beispielhafte Gewichte für alle QI, jedoch kein allgemeines Vorgehen zur Ermittlung der Gewichte.

1.3 Zielsetzung

Allgemeines Ziel dieser Dissertation ist es, Methoden und Werkzeuge zu erarbeiten, welche die Anwendung der Qualitätsmechanismen für medizinische Register vereinfachen und effektiv die Datenqualität verbessern können. Dies soll am Beispiel der CERTAIN-Register implementiert und erprobt werden. Die konkreten Ziele der vorliegenden Arbeit sind:

1. Analyse, welches Potential weitergehende, automatisierte Methoden der Qualitätssicherung im KRhOKo-Register haben.
2. Entwicklung eines allgemeinen Vorgehens zur Planung von IT-gestützten Maßnahmen zur Verbesserung der Datenqualität in Registern.

3. Spezialisierung des zu Ziel 2 entwickelten Vorgehens und Implementierung in der CERTAIN-Registerplattform in Form von wiederverwendbaren Modulen zur Qualitätsprüfung in medizinischen Registern.

Teilaufgaben:

- (a) Anwendung des TMF-Leitfadens für die drei CERTAIN-Register, insbesondere Ermittlung der Gewichte der QI für die Berechnung des Gesamtscores und Abstimmung mit den Arbeitsgemeinschaften der Registerzentralen der drei Register.
 - (b) Implementierung von generischen, automatisierten Methoden und Werkzeugen als Java-Module für Webservices und JUnit-Tests für diese Module.
 - (c) Spezialisierung der Methoden und Implementierung in der CERTAIN-Registerplattform der drei CERTAIN-Register.
4. Evaluation der eingeführten Qualitätsprüfungsmethoden durch Analysen, in denen überprüft wird, ob durch die Anwendung der automatisierten Methoden und Werkzeuge die Datenqualität in den CERTAIN-Registern verbessert werden konnte, insbesondere durch den Vergleich der Datenqualität, vor der Etablierung der neuen Maßnahmen, mit der Datenqualität nachdem diese Maßnahmen über einen größeren Zeitraum aktiv waren.

1.4 Aufbau der Dissertation

Die Dissertation gliedert sich in neun Kapitel. Die Arbeit gibt zunächst in Kapitel 2 einen Einblick in die Architektur, den Aufbau der CERTAIN-Register und die Ziele der einzelnen Register. Ebenso wird ein Überblick über die Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF) und deren Aktivitäten gegeben. Genauer betrachtet wird insbesondere der TMF-Leitfaden zur Datenqualität für Register. Schließlich werden die für die Arbeit wichtigen Spezifikationen dargelegt und das Kapitel mit einem Überblick über bisherige Arbeiten zum Thema der Dissertation abgeschlossen.

Kapitel 3 stellt die verwendeten Methoden und Werkzeuge dieser Arbeit sowie die gewählten Ansätze für die Analysen zur Datenqualität der CERTAIN-Register und der QI des TMF-Leitfadens dar.

In Kapitel 4 wird das Ergebnis einer Auswertung zum Potential des Ansatzes der automatischen Validierung im Hinblick auf die Umstellung der papierbasierten Erhebung der Kinder-Kerndokumentation zur elektronischen Datenerhebung des KRhOKo-Registers dargelegt.

Das entwickelte Vorgehen zur Verbesserung der Datenqualität und Datenqualitätssicherungsmaßnahmen wird in Kapitel 5 vorgestellt.

Kapitel 6 enthält die Spezialisierung dieses Vorgehens für die drei CERTAIN-Register.

In Kapitel 7 erfolgt die Beschreibung der implementierten Module.

Kapitel 8 stellt das Ergebnis der Überprüfung der Datenqualität ein Jahr nach Etablierung der neuen Datenqualitätssicherungsmaßnahmen dar.

Die Arbeit schließt mit Diskussion und Ausblick in Kapitel 9.

2 Grundlagen

Ausgehend von der dargelegten Problemstellung und den umzusetzenden Zielen, die sich in den einzelnen Kapiteln der Dissertation widerspiegeln, sollen in diesem Kapitel die Grundlagen zum Verständnis der Arbeit vermittelt werden. Dies beinhaltet neben einer genaueren Darstellung der einzelnen QI der TMF-Leitlinie, sowie der TMF selbst, auch einen Überblick über relevante Arbeiten Dritter, die dieser Dissertation vorausgingen. Zudem werden die verwendeten Technologien erläutert sowie die technischen Rahmenbedingungen beschrieben, welche die zu entwickelnden technischen Lösungen einhalten sollen. Zunächst wird ein Überblick über die drei Register gegeben, deren Datenqualität durch diese Arbeit verbessert werden sollen.

2.1 Cooperative European Paediatric Renal TransplAnt INitiative (CERTAIN) Registry

Das Cooperative European Paediatric Renal TransplAnt INitiative (CERTAIN) Registry ist ein pädiatrisches Nierentransplantationsregister, das 2009 vom Arbeitskreis „pädiatrische Nierentransplantation“ der Gesellschaft für pädiatrische Nephrologie (GPN) ins Leben gerufen wurde (Tönshoff et al. 2015). CERTAIN wurde mit dem Ziel etabliert, alle pädiatrischen Nierentransplantationen der Länder des Aktivitätsbereichs der GPN zu erfassen. Das Register sollte mit den erhobenen Daten klinische Forschung und Qualitätssicherung vereinfachen und es ermöglichen, Diagnose- und Therapiekonzepte zu vereinheitlichen (Plotnicki et al. 2013). CERTAIN wurde von Beginn an als webbasiertes Register entwickelt, um den teilnehmenden Einrichtungen Zugriff auf ihre Daten zu geben und um die Eingabe von Daten unabhängig von lokalen Informationssystemen zu ermöglichen. Aufgrund der geringen Fallzahlen der Zielpopulation wurde CERTAIN als multizentrisches Register entwickelt und ging 2010 in den Regelbetrieb über.

Der Datensatz des CERTAIN-Registers besteht aus einem minimalen, verpflichtend auszufüllenden, und einem extendierten, optionalen Datensatz. Hinzu kommen Studienmodule zur Ermöglichung der Auswertung von aktuellen wissenschaftlichen Fragestellungen, die dem Datensatz bei Bedarf hinzugefügt werden können und welche speziell für diese Studien relevante Datenelemente enthalten. Diese Studien können sowohl pro- als auch retrospektiver Natur sein und durch den großen Datensatz und die Studienmodule sehr unterschiedliche Themen abdecken (Brunkhorst et al. 2015; Habbig et al. 2017; Höcker et al. 2018a,b, 2016a,b; Weitz et al. 2018). Ebenso ist es allen teilnehmenden Einrichtungen möglich, zu eigenen Fragestellungen Studienpläne einzureichen und diese Studien nach Akzeptanz durch das Studienkomitee durchzuführen. CERTAIN definiert für die Datenerhebung vorgegebene Zeitpunkte in Form eines Visitenplans, welcher automatisch vorausberechnet wird.

Der Fokus des Datensatzes liegt auf klinischen Parametern wie Laborwerten und Medikationsgaben.

CERTAIN sollte zudem den Aufwand zur Dateneingabe reduzieren und die Vollständigkeit der Daten durch Datenaustausch mit Organisationen wie Eurotransplant (ET) oder der Collaborative Transplant Study (CTS) erhöhen. Hierfür wurden bidirektionale Schnittstellen entwickelt, welche die jeweils gewünschten Daten automatisiert übertragen (Köster et al. 2015). ET nutzt die von CERTAIN übertragenen Daten zur Auswertung des Transplantatsüberlebens, um die von ET durchgeführte Organallokation zu verbessern.

Ursprünglich lag der Betrachtungsrahmen des Registers auf der deutschsprachigen GPN-Region, durch die gestiegene wissenschaftliche Bedeutung des CERTAIN-Registers nehmen nunmehr nicht nur Einrichtungen aus der deutschsprachigen GPN-Region teil, sondern beispielsweise auch aus Polen, Tschechien und der Türkei. Seit dem Start des Registers wurden bis zum Januar 2018 2008 Patienten mit insgesamt 18382 Visiten durch 58 teilnehmende Einrichtungen und deren 183 Benutzer in das Register eingegeben.

2.2 Cooperative European Paediatric Transplant Initiative Liver (CERTAIN-LI) Registry

Das Cooperative European Paediatric Transplant Initiative Liver (CERTAIN-LI) Registry ist ein pädiatrisches Lebertransplantationsregister, dessen Entwicklung 2014 auf Basis des CERTAIN Registers begann (CERTAIN-LI 2018; Teufel et al. 2015). CERTAIN-LI verfolgt analog zu CERTAIN die Beantwortung wissenschaftlicher Fragestellungen zu pädiatrischen Lebertransplantationen und soll zudem zur Qualitätssicherung in diesem Bereich dienen. Der Fokus des Registers lag zunächst auf deutschen Einrichtungen, eine Erweiterung des Betrachtungsrahmens wurde durch die Arbeitsgemeinschaft „Pädiatrische Lebertransplantation“ des Auftraggebers der Gesellschaft für Pädiatrische Gastroenterologie und Ernährung (GPGE) längerfristig angestrebt. Diese Erweiterung wurde durch die Einbindung des GIGOLO-Projektes¹ in CERTAIN-LI erreicht. Infolgedessen nehmen nun auch ausländische Einrichtungen aus aller Welt an CERTAIN-LI teil. CERTAIN-LI ging Anfang 2015 in den Regelbetrieb über.

CERTAIN-LI folgt in seiner Definition des Datensatzes dem CERTAIN-Register, wurde in Leber-transplantationsspezifischen Punkten jedoch angepasst. Die Aufteilung des Datensatzes, in einen minimalen, extendierten und studienbezogenen Datensatz, wurde beibehalten. Gleichmaßen konnten viele Teile des Datensatzes mit wenigen Anpassungen übernommen werden, die klinische Ausrichtung blieb bestehen. Wie bei CERTAIN ist es den teilnehmenden Einrichtungen ebenso möglich, Studienpläne einzureichen und die Beantwortung eigener Fragestellungen zu verfolgen. Die vorausberechneten Zeitpunkte zur

¹<http://www.graftinjurygroup.com/>

Datenerhebung im ersten Jahr sind erhalten geblieben, die Anzahl der Visiten wurde jedoch um eine Visite reduziert. Für Visiten nach dem ersten Jahr nach der Transplantation wurde die Anzahl von zwei Visiten pro Jahr auf eine Visite pro Jahr aufgrund des mit der Datenerhebung verbundenen Aufwands reduziert.

Im Januar 2018 waren 190 Patienten mit insgesamt 574 Visiten durch 17 teilnehmende Einrichtungen und deren 37 Benutzer in CERTAIN-LI dokumentiert.

2.3 Kinderrheumatologie Online - Kerndokumentation (KRhOKo)

Die Kerndokumentation rheumakranker Kinder und Jugendlicher (Kinder-KD) ist ein 1997 unter Führung des Deutschen Rheuma-Forschungszentrum Berlin gestartetes Projekt, um den Stand der Versorgung bei rheumaerkrankten Kindern und Jugendlichen zu untersuchen sowie epidemiologische Fragestellungen zu beantworten (Minden et al. 2002). 1998 nahmen zunächst 18 pädiatrische Rheumaeinrichtungen an der Kinder-KD teil, die 2488 Patienten zur Auswertung beitrugen. 2014 begannen die Arbeiten zur Entwicklung einer webbasierten Version der Kinder-Kerndokumentation, der Kinderrheumatologie Online - Kerndokumentation (KRhOKo), die Anfang 2017 in den Regelbetrieb übergang und eine Alternative zur papierbasierten Datenerhebung bietet.

KRhOKo wurde auf Basis des CERTAIN-Registers entwickelt, wesentliche Teile der Architektur und der existierenden Funktionalitäten wurden übernommen. Da sich die Themengebiete und Erhebungszeitpunkte zu den zwei Transplantationsregistern stark unterscheiden, musste ein neuer Datensatz definiert werden, der in die bestehende CERTAIN-Architektur übernommen werden konnte. Dieser Datensatz orientierte sich sehr stark an der bisherigen papierbasierten Datenerhebung, um die Vergleichbarkeit der beiden Erhebungsstränge zu gewährleisten. Grundlegende Unterschiede zwischen KRhOKo und den beiden Transplantationsregistern bestehen in den frei wählbaren Erhebungszeitpunkten des KRhOKo-Registers sowie in dem Verzicht auf eine Unterteilung des Datensatzes in einen minimalen und extendierten Datensatz. Aufgrund der angestrebten Niederschwelligkeit der Datenerhebung existieren im KRhOKo-Datensatz zudem nur wenige Pflichtangaben. Einen wesentlichen Unterschied bildet des Weiteren die Patientenbeteiligung des KRhOKo-Registers, die es Patienten ermöglicht, ihren Patientenbogen, über einen per E-Mail versandten Link, außerhalb der Einrichtung selbstständig auszufüllen. Neben dem klinisch orientierten Arztfragebogen existieren für die Patienten jeweils zwei Fragebogen. Für Kinder unter 13 Jahre der Elternbogen und für Kinder von 13 und mehr Jahren der Bogen für Jugendliche.

Im Januar 2018 waren 1562 Patienten mit 1146 Visiten durch 15 teilnehmende Einrichtungen und deren 84 Benutzer in KRhOKo dokumentiert. Nicht alle Einrichtungen dokumentieren die Visiten der Patienten über KRhOKo, sondern dokumentieren weiter auf Papier. Diese Einrichtungen legen die Patienten zeitgleich jedoch auch im Register an, um eine

fortlaufende Dokumentation zu gewährleisten und eine Übersicht über ihre Patienten zu haben.

2.4 CERTAIN-Registerplattform

Da die beiden neueren Register auf dem CERTAIN-Register aufbauen, wird im Folgenden, im Hinblick auf die technischen Gemeinsamkeiten der drei Register, von der CERTAIN-Registerplattform gesprochen. Die Entwicklung des CERTAIN-Registers begann 2009 als eigenständige Entwicklung auf Basis von GWT², vormals Google Webtoolkit. GWT ist eine Anwendungsbibliothek, welche die Entwicklung von komplexen Webanwendungen auf Basis der Programmiersprache Java ermöglicht. Die eigentliche Entwicklung der Anwendung erfolgt dabei auf Basis von Java-Klassen, die durch den GWT-Compiler in HTML und Javascript übersetzt werden, welches wiederum durch den Web-Browser der Benutzer dargestellt werden kann. Die Web-Anwendung der drei Register ist als Thin-Client konzipiert, um ohne Installation einer lokalen Anwendung auszukommen, da die Installation von Programmen Dritter in vielen Kliniken oft nicht möglich ist.

Auf Serverseite wird die Anwendung in einem Tomcat-Applikationsserver eingesetzt. Als Programmiersprache der Registeranwendung kommt hauptsächlich Java zum Einsatz, in geringem Umfang wird die Programmiersprache Scala für Funktionalitäten zur Analyse des Datenbestandes eingesetzt.

Die Verbindung zur PostgreSQL-Datenbank wird durch den ORM-Mapper Hibernate auf Basis von JDBC realisiert. Kommunikation zwischen den Servern des Registers sowie über die Schnittstellen zu externen Partnern erfolgt über REST-Schnittstellen auf Basis von JAX-RS.

Das Kompilieren der Anwendung erfolgt durch Apache Maven, zur Versionskontrolle wird git eingesetzt.

Alle drei CERTAIN-Register bauen auf einer strikteren Version des TMF-Datenschutzkonzept A (Pommerening et al. 2009) auf, um die nationalen und lokalen Datenschutzgesetze aller potentiell teilnehmenden Einrichtungen einzuhalten. Dieses Datenschutzkonzept basiert auf der Trennung von medizinischen und identifizierenden Daten und der Speicherung dieser Daten in verschiedenen Datenbanken, deren Informationen nur auf dem lokalen System zugangsberechtigter Personen wieder zusammengeführt werden dürfen. Die Zusammenführung erfolgt über ein temporäres Pseudonym, die PID der Patienten wird nie in Verbindung mit den identifizierenden Daten (IDAT) an die Einrichtungen weitergegeben. Der IDAT-Service beinhaltet die Datenbank, welche die IDAT-Daten speichert. Die medizinischen Daten (MDAT) der Patienten werden zudem unter einem weiteren Pseudonym in einer Forschungsdatenbank gespeichert, dem MDAT-Service. Der MDAT-Service stellt die

²<http://www.gwtproject.org/>

MDAT über den Registry-Service den Benutzern zur Verfügung. Die Gesamtarchitektur ist in Abbildung 1 dargestellt.

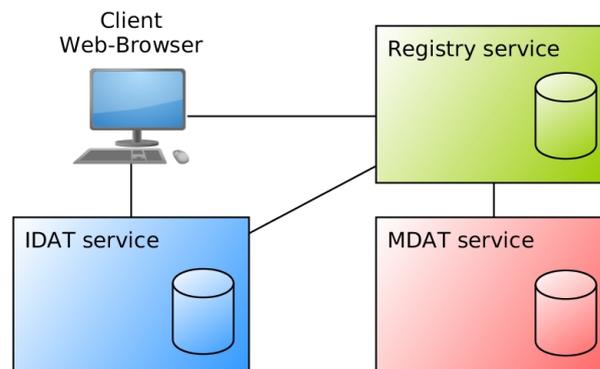


Abbildung 1: CERTAIN-Serverarchitektur (Köster et al. 2015)

Die CERTAIN-Registerplattform stellt grundlegende Funktionalitäten in Form von generischen Modulen bereit, in die ein Datenmodell integriert werden kann. So teilen sich alle Register, die auf der Plattform aufbauen, dieselbe Benutzer- und Einrichtungsverwaltung sowie das Rollenkonzept und notwendige Teile zur Darstellung und Navigation der Weboberfläche im Browser der Benutzer. Die wesentlichen Formulare der Weboberfläche sind generisch entwickelt und können in den einzelnen Registern erweitert werden, um registerspezifische Sonderfälle abzudecken. Zu den weiteren Gemeinsamkeiten gehören zudem die Dokumentenverwaltung, Darstellungen diverser Übersichtslisten und Statistiken zum Datenbestand, überdies E-Mailversand an Benutzer und Maßnahmen zur Sicherung der Datenqualität. Ebenso teilen sich alle drei Register dasselbe Berechtigungskonzept, das für registerspezifische Anforderungen erweitert werden kann.

2.4.1 Architektur

Die Architektur des Registry-Service der CERTAIN-Registerplattform geht auf die ursprüngliche Architektur des Registry-Service des CERTAIN-Registers zurück. Zunächst bestand diese Architektur aus drei Teilkomponenten, *Register*, *AnalysisModule* und *Common*, siehe Abbildung 2. Die Teilkomponente *Common* beinhaltete die Domänen-Objekte in Form von Java-POJOs. Im *AnalysisModule* befanden sich die in der Programmiersprache Scala geschriebenen Analysemethoden für Auswertungen zu medizinischen Fragestellungen und zur Generierung von Statistiken. Die Komponente *Register* beinhaltete die Informationen der Weboberfläche sowie die Programmteile zur Durchführung der allgemeinen Abläufe, wie beispielsweise das Anlegen und Speichern von Visiten oder die Benutzer- und Einrichtungsverwaltung und das zugehörige Berechtigungskonzept.



Abbildung 2: Projektstruktur CERTAIN 2014

Zur optimalen Wiederverwendung der CERTAIN-Codebasis für die neuen Register wurden die existierenden Teilkomponenten aufgeteilt und projektspezifische Komponenten der Gesamtarchitektur hinzugefügt. Für die jeweiligen Register wurden hierfür neue Komponenten etabliert, für CERTAIN *NTx*, für CERTAIN-LI *LTx* und für KRhOKo *KRh*, Abbildung 3. Aufgrund der Gemeinsamkeiten zwischen CERTAIN und CERTAIN-LI wurde zudem eine *Tx* Komponente etabliert indem die Gemeinsamkeiten der Transplantationsdomäne abgebildet wurden. Die grundlegende Aufteilung in *Register*, *AnalysisModule* und *Common* wurde dabei beibehalten. In den *Core*-Komponenten befinden sich hierbei alle Programmteile, welche in allen drei Registern gleich sind, beispielsweise die Benutzer- und Einrichtungsverwaltung oder grundlegende Datentypen.

2.4.2 Datenmodell

Unter dem Gesichtspunkt der Wiederverwendbarkeit und vereinfachten Wartung der drei Register wurden die Datenmodelle der Register in den zentralen Punkten abgestimmt. So folgen alle Register der zentralen Beziehung zwischen Einrichtungen, deren Patienten und ihren Visiten, Abbildung 4. Die Verknüpfungen dieser Klassen bilden den Dreh- und Angelpunkt des Registers, da nur durch die gemeinsame Definition dieser Objekte die Wiederverwendung der generischen Programmteile der *Core*-Komponenten sichergestellt werden kann. Visiten bilden das übergeordnete Objekt der Datenerhebung, sie bilden den Container für die anderen Objekte des jeweiligen Datensatzes und enthalten neben den projektspezifischen Objekten auch Informationen zur Verwaltung des Registers.

Neben den projektspezifischen Datentypen gibt es auch Datentypen und Objekte, die gleichermaßen in allen drei Registern genutzt werden. Zu diesen gehören unter anderem *Entry*, *ClinicalValue* und *OverviewItem*.

Das Interface *Entry* dient als Basis für alle Klassen der fortlaufenden Dokumentation des Datensatzes und wird beispielsweise zur detaillierten Dokumentation von Laborwerten oder Medikationsangaben des extendierten Datensatzes der beiden Transplantationsregister verwendet. Objekte dieses Typs besitzen mindestens ein Datum und einen Namen, so dass sich ein kontinuierlicher Verlauf des jeweiligen Medikamentes oder Laborwertes über mehrere Dokumentationszeitpunkte nachvollziehen lässt. *ClinicalValue* besteht aus einem Datenwert-Einheit-Paar und bildet den Grunddatentyp aller Angaben zu Laborwerten und Medikationsangaben der drei Register. Das *OverviewItem* ergänzt den *ClinicalValue* um

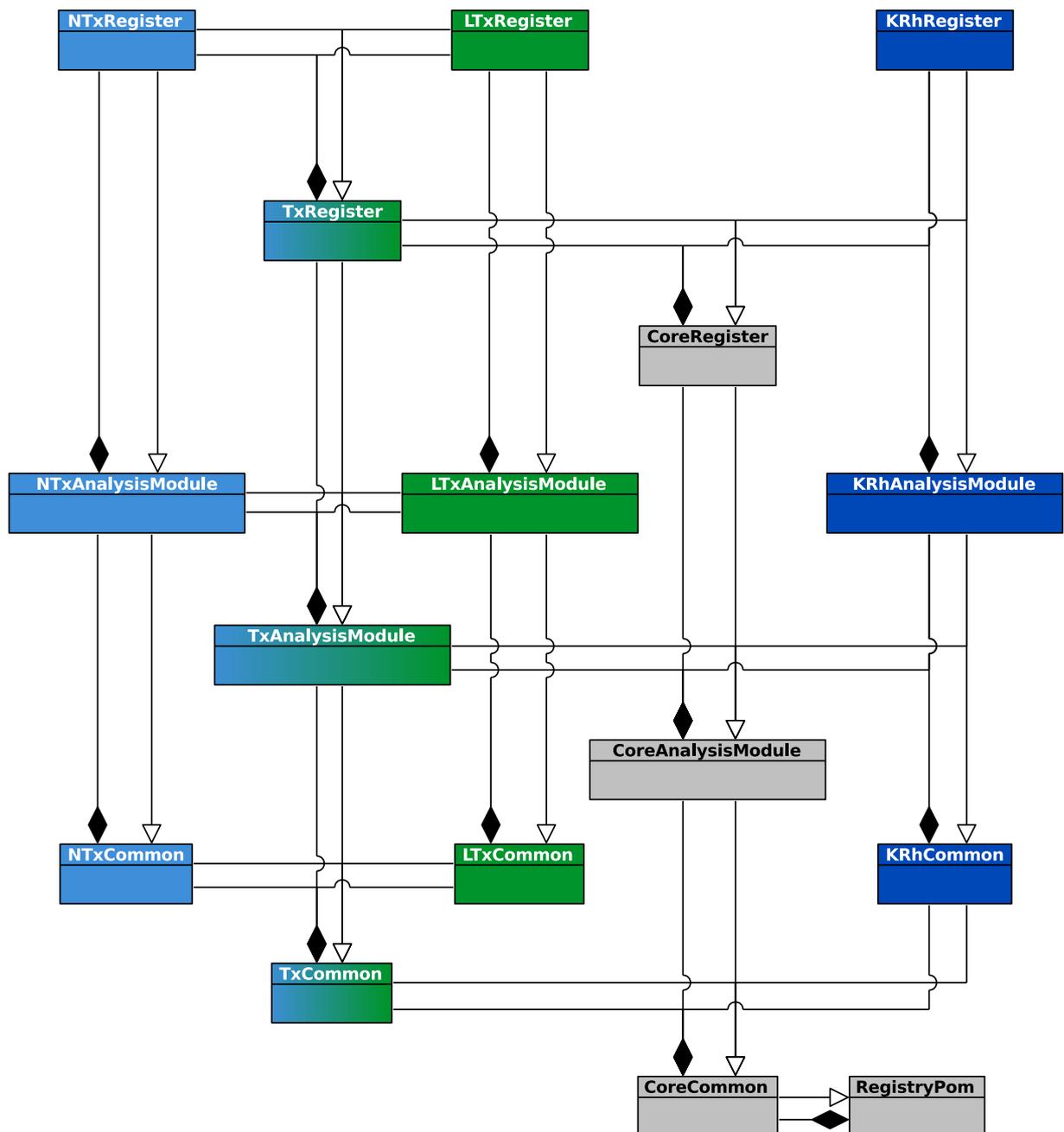


Abbildung 3: Projektstruktur CERTAIN-Registerplattform

eine zusätzliche Statusangabe, beispielsweise „unbekannt“, um das Fehlen des Datenwertes dokumentieren zu können.

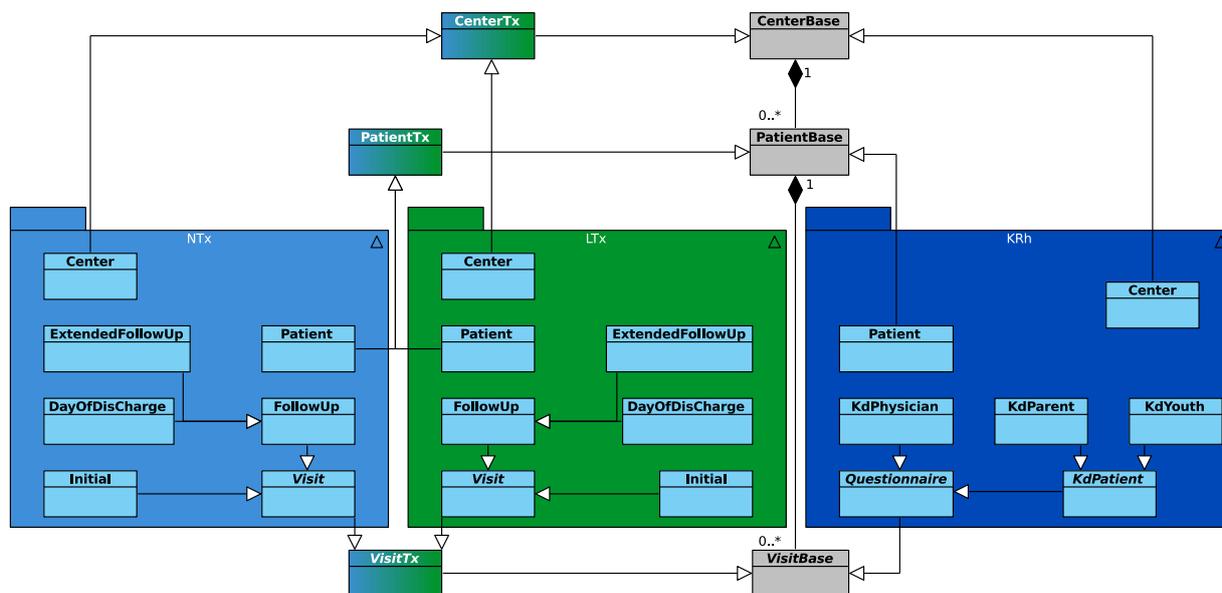


Abbildung 4: Ausschnitt Klassendiagramm der drei Register

2.4.3 Datenqualitätsmanagementprozess

Die drei Register setzen auf eine dreistufige Qualitätskontrolle auf Basis von automatischer Validierung und Überprüfung der eingegebenen Datenwerte durch einen Datenqualitätsmanager (DQM). Die automatische Validierung basiert auf festgelegten Plausibilitätsüberprüfungen sowie der Erzwingung von Angaben für Pflichtfelder durch technische Maßnahmen zur Eingabezeit. Das Ergebnis der automatischen Validierung wird der Eingabekraft bei jedem Speichern der Visite angezeigt. Die Eingabe wird zudem durch Markierung der als fehlerhaft festgestellten Datenelemente in den Eingabemasken unterstützt.

Durch die Einrichtung angelegte Visiten sind zunächst im Zustand „In Bearbeitung“. In diesem Zustand kann die Einrichtung die Visite soweit vervollständigen wie sie möchte, mindestens aber bis die automatische Validierung keine Fehler mehr feststellt. Visiten, die in Bearbeitung sind, können dabei nur durch die zugehörige Einrichtung eingesehen werden. Hat die Einrichtung die Eingabe der Visite abgeschlossen, alle Pflichtfelder ausgefüllt und Fehler korrigiert, kann die Eingabekraft die Visite in den Zustand „Freigegeben“ versetzen. In diesem Zustand wird die Visite für den DQM sichtbar, der die Visite einer Überprüfung im Hinblick auf die Plausibilität der Daten auf Basis seiner medizinischen Erfahrungen unterzieht. Sollte der DQM nicht plausible Daten feststellen, kann er die Visite mit einem Kommentar versehen und in den Zustand „Zurückgewiesen“ versetzen. Die Visite wird hierdurch zurück an die Einrichtung verwiesen, welche die Visite auf Basis des Kommentars des DQM überarbeitet und erneut, gegebenenfalls mit einem weiteren Kommentar zur Erklärung der Beanstandung, in den Zustand „Freigegeben“ versetzen kann. Sollte der DQM

keine Beanstandungen haben, versetzt er die Visite in den Zustand „Akzeptiert“, wodurch die Visite für Auswertungen zu wissenschaftlichen Fragestellungen verfügbar gemacht wird. Dieser Ablauf wird in Abbildung 5 dargestellt.

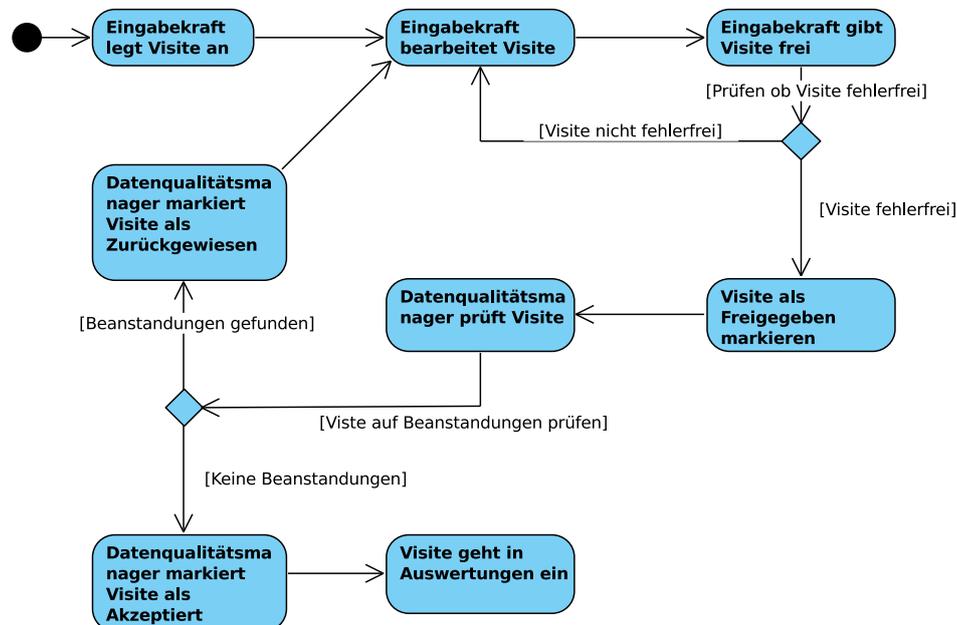


Abbildung 5: Datenqualitätsmanagementprozess der CERTAIN-Register

Die automatische Validierung der CERTAIN-Registerplattform stützt sich auf das Framework und die Funktionalitäten der Spezifikation *Bean Validation* der Programmiersprache Java.

2.5 Java Bean Validation

Die Programmiersprache Java unterstützt die Validierung von *JavaBeans* durch *Java Specification Request (JSR) 303 Bean Validation* seit November 2009 (Bernard et al. 2009). *JavaBeans* sind Klassen die mehrere Objekte zusammenfassen. Ziel der Spezifikation ist es, ein Metamodel und eine API zur Validierung von *JavaBeans* bereitzustellen (Bernard und Peterson 2009). Der Ansatzpunkt zur Validierung wird nicht vorgegeben und kann sowohl in der Web- als auch in der Persistenzschicht der Anwendung erfolgen. JSR 303 wurde im Mai 2013 durch JSR-349 *Bean Validation* Version 1.1 abgelöst, die wiederum durch JSR-380 *Bean Validation* Version 2.0 im August 2017 obsolet wurde (Bernard et al. 2013; Morling et al. 2017). Größte Neuerung in Version 1.1 war die Etablierung der Möglichkeit, neben Klassen und Feldern auch die Parameter und Rückgabewerte von Methoden zu validieren (Bernard 2013). Neuerungen in Version 2.0 beinhalteten Nutzung von JDK8

Features sowie neue Überprüfungen zur Validierung, beispielsweise auf die Wohlgeformtheit von E-Mail-Adressen (Morling 2017).

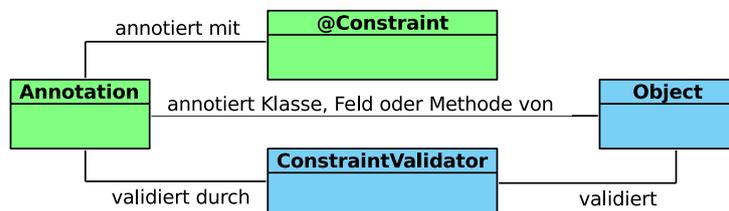


Abbildung 6: Ausschnitt Klassendiagramm ConstraintValidator

Die Validierung erfolgt dabei durch einen sogenannten *ConstraintValidator* in Verbindung mit einer *Constraint Annotation*, welche auf Klassen, Felder, Methoden und andere Annotationen angewendet werden kann, siehe Abbildung 6. Während die Annotation die Rahmenbedingungen der Validierungsregel bereithält, in Form von Parametern, die erfüllt sein müssen, führt der *ConstraintValidator* die eigentliche Validierung und Prüfung der annotierten Klasse, des Feldes oder der Methode gegen die Parameter der Annotation durch. Die *ConstraintValidator* werden durch einen *Validator* ausgeführt, der eine an ihn übergebene *JavaBean* nach *ConstraintValidator* durchsucht und das Ergebnis der einzelnen *ConstraintValidator* in Form eines Sets von *ConstraintViolations* gebündelt an die aufrufende Klasse zurückgibt. *ConstraintViolations* enthalten unter anderem neben Informationen zum *ConstraintValidator*, der den Verstoß gegen die Validierungsregel festgestellt hat, auch den Pfad zum Feld der Klasse, das den fehlerhaften Datenwert beinhaltet, sowie eine frei wählbare Nachricht.

Durch Vererbung der Klasse *ConstraintValidator* an eigene Klassen ist es Programmierern möglich, selbst Validierungsregeln auf Basis der Kombination aus *ConstraintValidator* und zugehöriger Annotation zu entwickeln und diese durch den *Validator* prüfen zu lassen.

Bean Validation bietet neben der Möglichkeit, eigene Constraints zu definieren auch eine Reihe von Basis-Constraints, die Teil der Spezifikation und in Tabelle 1 dargestellt sind.

Die Referenzimplementierung der JSR-380 bildet der Hibernate Validator Version 6.x, der in den CERTAIN-Registern seit November 2017 im Rahmen der automatischen Validierung zum Einsatz kommt. Frühere Versionen des Hibernate Validator waren von Beginn der Entwicklung des CERTAIN-Registers an Teil der Registerplattform. Der Hibernate Validator bietet zusätzliche Funktionen zur Validierung an, die nicht Teil der JSR sind. So existiert die Möglichkeit der gemeinsamen Validierung von mehreren Feldern innerhalb eines Objektes auf Basis der Annotation *ScriptAssert*. *ScriptAsserts* erlauben die statische Überprüfung von Sachverhalten durch die Formulierung der Validierungsregel in einer Skriptsprache. Diese Formulierung wird während der Validierung des Objektes verarbeitet und durch einen *ConstraintValidator* überprüft. Während sich *ScriptAsserts* zur Überprü-

<i>Annotation</i>	<i>Beschreibung</i>
@Null	Das annotierte Element muss <i>null</i> sein.
@NotNull	Das annotierte Element darf nicht <i>null</i> sein.
@AssertTrue	Das annotierte Element muss wahr sein.
@AssertFalse	Das annotierte Element muss unwahr sein.
@Min	Das annotierte Element muss eine Zahl sein deren Wert gleich dem oder höher als das spezifizierte Minimum ist.
@Max	Das annotierte Element muss eine Zahl sein deren Wert geringer als oder gleich dem spezifizierten Maximum ist.
@DecimalMin	Das annotierte Element muss eine Zahl sein deren Wert gleich dem oder höher als das spezifizierte Minimum ist.
@DecimalMax	Das annotierte Element muss eine Zahl sein deren Wert geringer als oder gleich dem spezifizierten Maximum ist.
@Size	Die Größe des annotierten Elementes muss zwischen den spezifizierten Grenzen liegen.
@Digits	Das annotierte Elementes muss die spezifizierte Anzahl an Ziffern und Nachkommastellen einhalten.
@Past	Das annotierte Element muss ein in der Vergangenheit liegendes Datum sein.
@Future	Das annotierte Element muss ein in der Zukunft liegendes Datum sein.
@Pattern	Das annotierte Element muss ein String sein, welcher der spezifizierten <i>regular expression</i> entspricht.
@Negative	Das annotierte Element muss eine negative Zahl sein.
@NegativeOrZero	Das annotierte Element muss eine negative Zahl oder 0 sein.
@Positive	Das annotierte Element muss eine positive Zahl sein.
@PositiveOrZero	Das annotierte Element muss eine positive Zahl oder 0 sein.
@PastOrPresent	Das annotierte Element muss ein in der Vergangenheit liegendes oder gegenwärtiges Datum sein.
@FutureOrPresent	Das annotierte Element muss ein in der Gegenwart oder Zukunft liegendes Datum sein.
@NotEmpty	Das annotierte Element darf nicht leer sein.
@NotBlank	Das annotierte Element darf nicht <i>null</i> sein und muss mindestens ein Zeichen enthalten.
@Email	Das annotierte Element muss das Format einer E-Mail-Adresse besitzen.

Tabelle 1: Basis-Constraints der *Bean Validation* (Morling 2017)

fung von Beziehungen zwischen mehreren Feldern innerhalb eines Objektes anbieten, steigt die Komplexität der Formulierung der Validierungsregel mit Zunahme der zu berücksichtigenden Felder sehr stark an. *ScriptAsserts* eignen sich daher nur mäßig zur Validierung von komplexen Sachverhalten, die beispielsweise mehrere Objekte umfassen.

2.6 Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.

Die Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF) bietet eine Plattform für standort- und projektübergreifende Zusammenarbeit in der medizinischen Verbundforschung und versteht sich als Dachorganisation dieses Bereichs (TMF 2018). Die TMF beschäftigt sich hierbei beispielsweise mit technologischen aber auch rechtlich-ethischen Problemstellungen der medizinischen Forschung. Neben Schulungs- und Beratungsangeboten stellt die TMF zudem IT-Anwendungen, generische Konzepte und Checklisten sowie Leitfäden zur Verfügung. Als ihre Ziele versteht es die TMF unter anderem, die Qualität, Organisation, Zusammenarbeit, Qualitätssicherung und das Qualitätsmanagement medizinischer Forschung zu verbessern. Dies soll durch die Bereitstellung von Musterlösungen zu rechtlichen und ethischen Problemstellungen unterstützt werden. Ebenso sollen vernetzte Strukturen durch die Bereitstellung und Entwicklung von IT-Infrastrukturen geschaffen werden.

Neben den Leitfäden zum Datenschutz in medizinischen Forschungsprojekten und zur informierten Einwilligung bietet die TMF auch die TMF-Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern (Leitlinie Datenqualität) an.

2.7 Leitlinie Datenqualität

Die TMF-Leitlinie Datenqualität enthält Empfehlungen zum Management der Datenqualität von Kohortenstudien, Registern sowie Data Repositories (Nonnemacher et al. 2014). Zielgruppe des Leitfadens sind hierbei Benutzer, die für Qualitätssicherung und Datenmanagement in den vorher genannten Datensammlungen verantwortlich sind. Die TMF-Leitlinie wurde erstmals am 31.03.2006 durch die TMF veröffentlicht und enthielt zu diesem Zeitpunkt 24 QI zur Datenqualität (Nonnemacher et al. 2007). 2011 wurde durch zwei Veröffentlichungen von Jacke et al. (Jacke et al. 2010) und Prokein et al. (Prokein et al. 2011) zur Anwendung der TMF-Leitlinie und durch Rückmeldung durch TMF-Mitglieder auf einem Workshop ersichtlich, dass eine Überarbeitung der TMF-Leitlinie erforderlich war (Nonnemacher et al. 2014, S. 1). Im Zuge der Betrachtung der Anforderungen zur Messung von Datenqualität aus dem Alltag von Registern und Kohortenstudien wurde der konzeptionelle Ansatz der ersten Version der TMF-Leitlinie bestätigt. Zugleich wurden jedoch auch die vorhandenen Indikatoren um neue Indikatoren erweitert. Diese Einbindung von

weiteren Indikatoren bildet die wesentliche Neuerung der TMF-Leitlinie in Version 2.0. Die neuen Indikatoren wurden dabei im Rahmen einer umfassenden Literaturrecherche von den Autoren der TMF-Leitlinie erfasst. Von zunächst 1916 gefundenen Literaturstellen wurden, nach Ausschluss auf Basis festgelegter Kriterien, schließlich 39 Literaturstellen in der TMF-Leitlinie dargestellt (ebd., S. 174). Die TMF-Leitlinie wurde zudem um beispielhafte Anwendungsempfehlungen zu Registern, Kohorten und Data Repositories erweitert, die als notwendig empfunden wurden, da die große Menge der Indikatoren eine Auswahl der anzuwendenden Indikatoren notwendig machte.

Die Autoren der TMF-Leitlinie bauen die TMF-Leitlinie dabei auf der Arbeit von Arts et al. (Arts et al. 2002) auf, die sich mit einem generischen Modell zur Datenqualität in Registern befasst. Die Autoren identifizierten hierbei sowohl den Originaldatenabgleich zwischen Datenspeicherort und Originaldaten (Source Data Verification (SDV)), als auch das Feedback als besonders relevante Maßnahmen. Die Autoren betonen, dass eine vollständige SDV zwar wünschenswert, aber aufgrund von finanziellen Restriktionen kaum umsetzbar ist. Der Fokus der TMF-Leitlinie richtet sich daher auf das adaptive Monitoring. Dieses beschreibt Maßnahmen zur dynamischen Anpassung der Aktivitäten zur Prüfung von Datenqualität anhand der aktuell gemessenen Datenqualität. Die Prüfung der Datenqualität erfolgt dabei zum einen durch Indikatoren, die aus dem aktuellen Datenbestand erhoben werden können und zum anderen aus dem Grad der Übereinstimmung der gemeldeten Daten mit den Originaldaten, deren Vollständigkeit sowie ihre Vollzähligkeit. Das adaptive Monitoring verfolgt dabei das Ziel der Qualitätslenkung, durch die eine kontinuierliche Verbesserung der Datenqualität der teilnehmenden Leistungserbringer erreicht werden soll, siehe Abbildung 7.

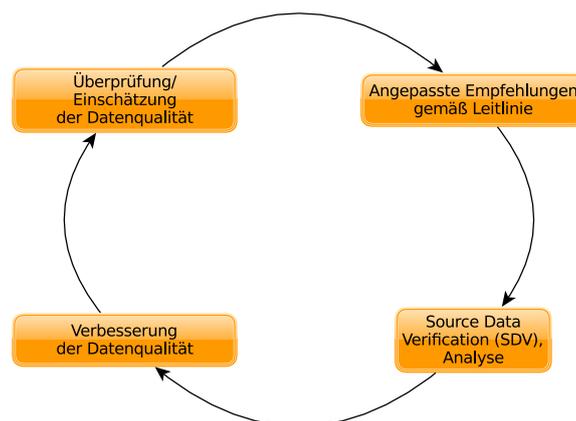


Abbildung 7: Adaptives Monitoring der Datenqualität nach (Nonnemacher et al. 2014)

Eine komplette Übereinstimmung der gemeldeten Daten mit den Originaldaten wird dabei nicht angestrebt. Der Ansatz des adaptiven Monitorings wurde gemäß TMF-Leitlinie

im Februar 2005 auf einem Workshop durch die Teilnehmer mit ihren unterschiedlichen Blickwinkeln (Register, Studien) bestätigt. Feedback für die beteiligten Einrichtungen oder Patienten biete sich, gemäß Leitlinie, zudem als Alternative zur SDV an, wenn das Studiensystem der einzige permanente Speicherort der erhobenen Daten ist, beispielsweise bei Daten, die von Patienten berichtet werden (z.B. Patient Reported Outcome) oder automatisierter Übermittlung von Sensordaten.

2.7.1 Vorgehen bei der Anwendung

Die Anwendung der TMF-Leitlinie auf Basis der SDV erfolgt in mehreren Schritten. Zunächst sollen die Daten aus dem Datenbestand extrahiert werden, die seit der letzten Analyse eingegangen sind. Folgend werden die Metadaten des Vorhabens zusammengestellt und die Parameter ermittelt, welche für die Berechnung der QI herangezogen werden. Die Berechnung soll für jedes Zentrum einzeln und bevorzugt rechnerunterstützt erfolgen.

Auf Basis dieser Vorarbeiten beginnt die Anwendung der TMF-Leitlinie mit der Berechnung der Werte der QI, welche im nächsten Schritt auf Basis von Schwellenwerten als un-/auffällig klassifiziert werden. Aus den Werten der QI wird ein Qualitätsscore berechnet, der auf einer durch den Anwender zu bestimmenden Gewichtung der einzelnen QI basiert. Die TMF-Leitlinie bietet eine beispielhafte Gewichtung der QI zur Berechnung des Qualitätsscores an, betont aber die Notwendigkeit der Anpassung der Gewichte für jedes Vorhaben, besonders im Hinblick auf die jeweils angewandten QI (Nonnemacher et al. 2014, S. 93). Gleichermaßen muss der Anwender die Schwellenwerte der QI zur Klassifizierung in die Kategorien un-/auffällig selbst wählen. Die TMF-Leitlinie betont im Hinblick auf die Schwellenwerte, dass in der Literatur keine Evidenz-basierten Empfehlungen gefunden werden konnten und deshalb auf die Angabe von Schwellenwerten verzichtet wird (ebd., S. 89).

Dem vorhergehenden Schritt folgt die Berechnung des Qualitätsscores unter Einbezug der vorherigen Klassifizierung und die Planung der künftig zu untersuchenden Anzahl an Beobachtungseinheiten des Datenbestandes auf Basis des ermittelten Qualitätsscores. Diese Fallzahlplanung schließt die Anzahl der zu untersuchenden Merkmale und der Besuche in der Einrichtung ein. Das Vorgehen schließt mit der Erstellung der Berichte für die Einrichtungen und der Festlegung des Termins für die nächste Analyse auf der Basis der festgestellten Datenqualität der jeweiligen Einrichtung.

2.7.2 Qualitätsindikatoren

Die Qualitätsindikatoren der TMF-Leitlinie zum adaptiven Datenmanagement von Datenqualität in Kohortenstudien und Registern sind in insgesamt drei Ebenen unterteilt. Die Ebene Integrität stellt mit 30 QI den Hauptanteil. Fünfzehn QI entfallen auf die Ebene Organisation, während die Ebene Richtigkeit mit sechs Indikatoren den Abschluss bildet.

Die Ebenen orientieren sich an den von Donabedian et al. (Donabedian 1988) eingeführten Kategorien Strukturqualität (in der Leitlinie: Ebene Integrität), Prozessqualität (in der Leitlinie: Ebene Organisation) und Ergebnisqualität (in der Leitlinie: Ebene Richtigkeit). Die Berechnung eines QI ergibt einen Wert, der in der Regel einen Anteil an fehlerhaften Datenwerten darstellt. Dieser Wert ist je nach QI unterschiedlich hinsichtlich seiner Bewertung der Datenqualität zu interpretieren. Die TMF-Leitlinie gibt für keinen QI Grenzen an, innerhalb derer die Datenqualität ausreichend ist. Dies zu beurteilen wird dem Anwender überlassen, damit die jeweils geplante Datennutzung berücksichtigt werden kann. In Version 1.0 der TMF-Leitlinie werden in den Implementierungshilfen zur TMF-Leitlinie Schwellenwerte für 24 QI empfohlen. Diese wurden nicht mehr in die Version 2.0 der TMF-Leitlinie übernommen.

Die Berechnung der QI soll je dokumentierender Einrichtung erfolgen. Welche Daten in welchem Umfang in die Berechnung einzubeziehen sind, hängt vom jeweiligen QI ab und kann von einzelnen Datenelementen über einzelne Datensätze zu einzelnen Beobachtungseinheiten bis zum gesamten Datenbestand der Einrichtung reichen. Im Folgenden werden die QI der TMF-Leitlinie vorgestellt, da sie die Ausgangsbasis für die weiteren Betrachtungen dieser Dissertationsschrift darstellen.

Qualitätsindikatoren der Ebene Integrität

Die Qualitätsindikatoren der Ebene Integrität treffen überwiegend Aussagen zur Strukturqualität der Daten des Registers. Die Ausnahmen bilden die QI „Fehlende Werte bei mandatorischen Datenelementen“ (Ergebnisqualität) und „TMF-1050 - Umfang der Metadaten bei Untersuchungen“ (Prozessqualität). Die Berechnung der QI basiert stets auf den im Register gespeicherten Datenwerten. Die einzige Ausnahme bildet hier der QI mit Aussagen zur Prozessqualität, in den die Parameter des zu untersuchenden Prozesses einfließen.

TMF-1001 - Übereinstimmung mit Vorwert

Die Berechnung des QI „Übereinstimmung mit Vorwert“ ermittelt den Anteil von Datenwerten, die sich in nicht plausibler Weise von ihrem Vorwert eines früheren Erhebungszeitpunkts unterscheiden und somit einen Datenfehler darstellen können, an allen überprüften Datenwerten. Wichtig ist hierbei die Unterscheidung zwischen Datenelementen zur Repräsentation von unveränderlichen Merkmalen, wie dem Geburtsdatum, und Datenelementen für Merkmale, welche im Rahmen von Messungenauigkeiten variieren können, beispielsweise die Körpergröße. Zusätzlich gibt es Datenelemente für veränderliche Merkmale, bei denen Abweichungen zu erwarten sind und auftreten dürfen, zum Beispiel Laborparameter. Für die Berechnung des QI bei unveränderlichen Merkmalen sind für jedes Datenelement Abweichungen zu definieren, die noch als akzeptabel hinzunehmen sind. Beispielführt die TMF-Leitlinie an, dass für das Geburtsdatum keine Abweichungen zulässig seien, je-

doch Abweichungen in der Größenordnung von wenigen Zentimetern für die Körpergröße akzeptabel sein sollten (Nonnemacher et al. 2014, S. 15). Bei veränderlichen Merkmalen kann es zu Abweichungen zum Vorwert kommen, die nicht plausibel sind. Daher ist es hier erforderlich, noch-plausible Abweichungsgrenzen zum Vorwert festzulegen.

In die Berechnung des QI geht der Betrag der Differenz des überprüften Datenwertes zum Vorwert ein; sollte dieser über der festgelegten, akzeptablen Abweichung liegen, geht dieser Datenwert im Zähler der Berechnungsformel für den QI ein. Im Nenner findet sich die Anzahl der insgesamt im Rahmen der Berechnung des QI überprüften Datenwerte. Höhere Werte des QI sind entsprechend ein Zeichen vieler nicht plausibler Datenwerte und somit von schlechter Datenqualität. In der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 wurde als Schwellenwert für Auffälligkeiten $> 5\%$ empfohlen (Nonnemacher et al. 2007, S. 61). Rahmen der Berechnung können einzelne Beobachtungseinheiten oder ein Datenbestand sein. Die Anwendung ist nur auf mehrfach erhobene Datenelemente sinnvoll. Beeinflusst wird der Wert des QI durch standardisierte Untersuchungsverfahren und Analysemethoden. Unveränderliche Werte nur einmal zu erfassen, ist ein weiterer Einflussfaktor. Laut Aussagen in der TMF-Leitlinie wurde der QI auf Basis der Ergebnisse mehrerer Arbeiten übernommen (Gassman et al. 1995; Naumann und Rolker 2000; O’Kelly 2004).

TMF-1002 - Konkordanz

Die TMF-Leitlinie definiert Konkordanz als Übereinstimmung eines Datenwertes mit einer Referenzquelle (z. B. Patientenakte, CRF) (Nonnemacher et al. 2014, S. 17). Ergebnis der Berechnung des QI ist der Anteil der Datenwerte eines Datenelements in einem Datenbestand, welche nicht mit der Referenzquelle übereinstimmen, an der Gesamtzahl der überprüften Datenwerte. Für die Berechnung muss zunächst die Wahl der Referenzquelle erfolgen, anschließend werden die Datenwerte verglichen und Zähler und Nenner bestimmt. Ein hoher Wert des QI steht für viele Abweichungen zwischen der Referenzquelle und den gespeicherten Daten und somit für schlechte Datenqualität. In der Implementierungshilfe der TMF-Leitlinie 1.0 wurde als Schwellenwert für Auffälligkeiten $> 5\%$ empfohlen (Nonnemacher et al. 2007, S. 61).

Für die Beurteilung ist zu beachten, dass nicht die Richtigkeit der Datenwerte selbst überprüft wird, sondern nur ihre Dokumentation in zwei Quellen. Aussagen über den Wahrheitsgehalt der Daten sind daher, gemäß TMF-Leitlinie nach Aussage einer der Quellen des QI, nicht möglich (Brennan und Stead 2000). Die Referenzquelle soll als „wahr“ angenommen und entsprechend sorgfältig gewählt werden. Weiter gibt die TMF-Leitlinie für die Wahl der Referenzquelle zu bedenken, dass, bei einer Übereinstimmung von gewählter Referenzquelle und der Quelle der Datenerhebung, dieser QI dem QI „Übereinstimmung der Registerdaten mit den Originaldaten“ entspricht. Die TMF-Leitlinie bietet zudem eine alternative Definition des QI an; die Vollständigkeit der Datenerhebung könne auch als Konkordanz verstanden werden, wenn betrachtet wird, ob alle Daten einer Referenzquel-

le übernommen wurden (Aronsky und Haug 2000). Die TMF-Leitlinie nennt eine weitere Literaturquelle auf deren Basis der QI in die TMF-Leitlinie aufgenommen wurde (Cook et al. 2002).

Als Einflussfaktoren auf den QI werden Qualität der Datenerhebung, zur Verfügung stehende Eingabezeit und -personal, sowie Qualifikation des Eingabepersonals genannt.

TMF-1003 - Widerspruchsfreiheit

Widerspruchsfreiheit von Datenwerten wird für die Berechnung des QI durch Plausibilitätsprüfungen überprüft. Plausibilitätsprüfungen werden durch logische Ausdrücke abgebildet und können mehrere Datenelemente umfassen. Als Beispiel für Plausibilitätsprüfungen nennt die TMF-Leitlinie die Prüfung auf weibliches Geschlecht bei schwangeren Personen. Die Berechnung des QI erfolgt durch die Bestimmung des Anteils der Datenwerte, die mindestens eine Plausibilitätsprüfung verletzen an der Gesamtanzahl der überprüften Datenwerte. Die Berechnung kann für die Datenelemente einzelner Datensätze, Beobachtungseinheiten oder eines Datenbestands durchgeführt werden. Die Berechnung ist nur für Datenelemente sinnvoll und erfolgt durch Anwendung der Plausibilitätsprüfungen auf die Datenwerte des ausgewählten Datenelementes, um Widersprüche zu identifizieren. Ein hoher berechneter Wert des QI steht für viele Widersprüche und somit schlechte Datenqualität. Als Schwellenwert wurden $> 5\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61). Der Wert des QI lässt sich durch automatisierte Plausibilitätsprüfungen während der Dateneingabe beeinflussen. Gemäß TMF-Leitlinie sind widerspruchsfreie Daten wichtig für die Aussagekraft von Auswertungen (Nonnemacher et al. 2014, S. 19). Entsprechend ist der QI gemeinhin bekannt und wurde der TMF-Leitlinie zufolge auf Basis der Ergebnisse mehrerer Quellen in die TMF-Leitlinie übernommen (Bobrowski et al. 1999; Gaus 2003; Jung und R. Winter 2000; Naumann und Rolker 2000).

TMF-1035 - Endlosüberlebende

Die TMF-Leitlinie definiert Endlosüberlebende als Patienten oder Probanden, für die der Tod während der Laufzeit des Forschungsvorhabens niemals formal festgestellt wird (Nonnemacher et al. 2014, S. 21). Der Wert des QI entspricht dem Anteil der identifizierten Endlosüberlebenden an der Anzahl der überprüften Patienten eines Datenbestands. Ein hoher Wert kann für bestimmte Erkrankungen für schlechte Datenqualität stehen, wobei der Wert durch unzureichendes Follow-up und fehlende Informationen von Meldestellen negativ beeinflusst wird. Anhand des Werts des QI lassen sich daher Aussagen über Vollständigkeit und Validität der Nachsorge des Patienten treffen, sowie zur Recherche zu diesem Themenkomplex. Besonders bei Tumoren mit schlechter Prognose und kurzer Überlebenszeit seien deutlich längere Überlebenszeiten ein Hinweis auf fehlende Informationen oder mangelnde Recherche (ebd., S. 21). Gemäß der Aussagen der TMF-Leitlinie hat der QI

seinen Ursprung im Bereich der Krebserkrankungen (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a). Der QI ist eine Subkategorie des QI „Widerspruchsfreiheit“.

TMF-1004 - Sichere Widersprüche/Fehler

Eine weitere Unterkategorie von „Widerspruchsfreiheit“ ist der QI „Sichere Widersprüche/Fehler“. Dieser QI entstammt gemäß TMF-Leitlinie in Teilen den selben Quellen wie der vorherige QI (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a,b). Im Rahmen der Berechnung der QI der TMF-Leitlinie werden, von Plausibilitätsprüfungen erfasste, Widersprüche und Fehler in unterschiedliche Schweregrade unterteilt. Der berechnete Wert dieses Indikators beschreibt den Anteil von sicheren Widersprüchen und Fehlern bei der Durchführung von Plausibilitätsprüfungen an der Gesamtzahl der durchgeführten Plausibilitätsprüfungen. Entsprechend ist ein hoher Wert des QI aufgrund der festgestellten Fehler mit schlechter Datenqualität assoziiert.

Für die Berechnung ist es erforderlich, die Plausibilitätsprüfungen zu identifizieren, deren festgestellte Widersprüche oder Fehler eine große Relevanz für das Vorhaben und somit den geforderten Schweregrad besitzen. Die Berechnung kann sowohl für einzelne Datensätze, Beobachtungseinheiten als auch den Datenbestand ausgeführt werden. Der Wert des QI soll sich durch vollständige, korrekte Datenerhebung und die Motivation der Studienteilnehmer verbessern lassen (Nonnemacher et al. 2014, S. 22). Die TMF-Leitlinie merkt an, dass sichere Widersprüche und Fehler bei epidemiologischen Krebsregistern aufgrund ihrer Schwere zwingend zu korrigieren seien. Unzureichend korrigierte Meldungen könnten demnach zurückgewiesen werden und somit in TMF-1040 „Zurückgewiesene Meldungen“ eingehen.

TMF-1005 - Mögliche Widersprüche/Warnungen

Dieser QI ist mit dem vorherigen QI verwandt, bildet ebenfalls eine Unterkategorie von „Widerspruchsfreiheit“ und basiert gemäß TMF-Leitlinie auf den selben Quellen wie der vorherige QI (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a,b). Vom vorherigen QI unterscheidet sich dieser QI in der Schwere der festgestellten Widersprüche. In die Berechnung des Wertes dieses QI gehen nur Plausibilitätsprüfungen ein, deren gemeldete Fehler auf Daten beruhen, die als nicht plausibel aber prinzipiell möglich erscheinen. Als Beispiel für solche Plausibilitätsprüfungen führt die TMF-Leitlinie Brustkrebs beim Mann an (Nonnemacher et al. 2014, S. 23). Der berechnete Wert des QI entspricht dem Anteil der in möglichen Fehlern oder Widersprüchen resultierenden Plausibilitätsprüfungen an der Gesamtheit der durchgeführten Plausibilitätsprüfungen. Ein hoher berechneter Wert ist Zeichen schlechter Datenqualität. Zu bedenken sei gemäß Leitlinie, dass ein hoher Wert auch die Folge von unzutreffenden Plausibilitätsprüfungen sein könne. Die Einflussfaktoren entsprechen denen des vorherigen QI TMF-1004, zusätzlich werden noch der Umfang und die Tiefe der Plausibilitätsprüfungen genannt. Die Berechnung kann gleich TMF-1004

für einzelne Datensätze, Beobachtungseinheiten oder den Datenbestand der Einrichtung durchgeführt werden.

TMF-1006 - Werteverteilung

Für numerische Datenelemente bietet die TMF-Leitlinie den QI „Werteverteilung“ zur Berechnung des Anteils der auffälligen Werteverteilungen an der Gesamtzahl der überprüften Werteverteilungen an. Die Berechnung des Wertes des QI erfolgt zunächst durch die Berechnung der Kenngrößen Minimum, Maximum, der Quartile, Mittelwert, Median sowie der Standardabweichung für jedes einbezogene Datenelement. Aus dem früheren Datenbestand der Einrichtung oder der anderen Einrichtungen werden Referenzwerte festgelegt, die mit den berechneten Kenngrößen verglichen werden. Sollten die Kenngrößen von den Referenzwerten abweichen, gilt die Werteverteilung gemäß TMF-Leitlinie als auffällig (ebd., S. 25). Anschließend wird der Anteil der auffälligen Werteverteilungen an allen überprüften Datenelementen berechnet.

Die Berechnung kann sowohl für einzelne Beobachtungseinheiten als auch für einen Datenbestand durchgeführt werden. Ein hoher berechneter Wert deutet auf schlechte Datenqualität hin. Auffällige Werteverteilungen sollten gemäß TMF-Leitlinie zu eingehenden Untersuchungen der einbezogenen Datenwerte führen, da auffällige Verteilungen Hinweise auf Probleme bei verwendeten Untersuchungsverfahren, Analysemethoden wie auch auf Datenfälschung sein können. Als Schwellenwert wurden $> 10\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61). Die TMF-Leitlinie nennt zwei Quellen auf deren Basis der QI in die TMF-Leitlinie übernommen wurde (Hasford 1994; J. Schmidt et al. 1995). Beeinflussen lassen soll sich der Wert des QI durch die Standardisierung von Untersuchungsverfahren und Analysemethoden. Die Population der in das Register eingegangenen Patienten der Einrichtung spiele ebenso eine Rolle.

TMF-1007 - Bevorzugung bestimmter Endziffern

Eine Unterkategorie des vorangegangenen QI ist der QI „Bevorzugung bestimmter Endziffern“. Der berechnete Wert des QI beschreibt den Anteil der Datenelemente mit einer auffälligen Verteilung der Endziffern der Datenwerte des Datenelementes an der Gesamtheit der überprüften Datenelemente. Ein hoher Wert des QI ist ein Hinweis auf schlechte Datenqualität. Für die Berechnung werden die Häufigkeiten der Endziffern aus den jeweiligen Datenwerten der Datenelemente extrahiert und, im Hinblick auf eine angenommene Gleichverteilung der Endziffern, auf auffällige Verteilungen untersucht. Kann für das untersuchte Datenelement keine Gleichverteilung der Endziffern angenommen werden, sollte der QI nicht eingesetzt werden. Ursprung schlechter Datenqualität können, gemäß Leitlinie, Probleme bei verwendeten Analysemethoden und Untersuchungsverfahren zur Erfassung der Datenwerte sein. Zusätzlich gibt die TMF-Leitlinie auch Datenfälschung als Grund

für auffällige Unterschiede an (Nonnemacher et al. 2014, S. 27). Als Schwellenwert wurden $> 5\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61). Gemäß TMF-Leitlinie wurde der QI auf Basis zweier Quellen in die TMF-Leitlinie übernommen (Hasford 1994; O’Kelly 2004).

TMF-1009 - Werteverteilung der durch Untersucher erfassten Parameter

Bei technischen Untersuchungen kann es zu einer manuellen Erfassung als Zwischenschritt zwischen der Erzeugung der Messwerte und der Befundung kommen. Ergebnis der Berechnung des Wertes dieses QI ist der Anteil der Untersucher mit einer auffälligen Werteverteilung an der Gesamtheit der betrachteten Untersucher. Gemäß TMF-Leitlinie kann Auffälligkeit bei Abweichung eines Verteilungsparameters um einen bestimmten Betrag vorliegen. Dieser QI ist eine Subkategorie des QI TMF-1006 „Werteverteilung“. Für die Berechnung ist es erforderlich, zunächst eine technische Untersuchung, die mehrere Geräte einbeziehen kann, als Ziel der Berechnung auszuwählen. Des Weiteren ist, anhand eines festzulegenden Kriteriums, zu bestimmen, wann eine Verteilung als auffällig gilt. Anschließend erfolgt die Berechnung der Verteilungen für die Untersucher, die Untersuchung der Verteilungen auf Auffälligkeit und die abschließende Berechnung des QI für den Datenbestand. Höhere Werte des QI sind Zeichen schlechter Datenqualität.

Einflussfaktoren des QI sind sowohl die Kompetenz der Untersucher sowie die standardisierte Erfassung der Messwerte (Nonnemacher et al. 2014, S. 29). Es sei zudem auszuschließen, dass die Berechnung bei Einschluss mehrerer Geräte durch das Vorliegen von festen Untersucher-Geräte-Kombinationen beeinflusst wird. Berücksichtigt werden sollte zudem die Anzahl der eingehenden Untersuchungen pro Untersucher.

In die TMF-Leitlinie übernommen wurde der QI gemäß TMF-Leitlinie auf Basis eines TMF-Projekts (C. O. Schmidt et al. 2012c)

TMF-1010 - Werteverteilung der durch Geräte erfassten Parameter

Während der vorherige QI sich auf auffällige Werteverteilungen der Untersucher bei technischen Untersuchungen fokussiert, legt dieser QI sein Augenmerk auf die bei technischen Untersuchungen verwendeten Geräte. Die Berechnung des QI gibt Auskunft über den Anteil der Geräte mit auffälligen Werteverteilungen an der Gesamtzahl der überprüften Geräte. Gleich dem vorherigen QI bildet der QI eine Unterkategorie des QI „Werteverteilung“. Die TMF-Leitlinie nennt als Quelle für den QI das selbe TMF-Projekt wie bei dem vorherigen QI (ebd.).

Für die Berechnung müssen zunächst die zu überprüfenden Geräte ausgewählt und ein Kriterium für Auffälligkeit festgelegt werden. Auf Basis dieses Kriteriums sollen die Werteverteilungen der Geräte untersucht werden und in die Berechnung des QI für den Datenbestand eingehen. Ein hoher Wert des QI steht daher für schlechte Datenqualität. Zu

berücksichtigen sei im Rahmen der Interpretation des Wertes des QI die Anzahl der Untersuchungen pro Gerät (Nonnemacher et al. 2014, S. 30). Auch die Wartung der Geräte sowie ihre korrekte Handhabung und Kalibrierung sollen den Wert des QI beeinflussen.

TMF-1011 - Werteverteilung von Befunden

Der letzte Teilaspekt der Qualität von technischen Untersuchungen ist die Betrachtung der Qualität im Hinblick auf die befundende Person. Gleich den beiden vorangegangenen QI ist dieser QI eine Unterkategorie von „Werteverteilung“ und geht auf das selbe TMF-Projekt zurück (C. O. Schmidt et al. 2012c). Ergebnis der Berechnung des QI ist der Anteil der Befunder mit auffälligen Werteverteilungen an der Gesamtzahl der überprüften Befunder. Ein hoher Wert ist Zeichen einer schlechter Datenqualität des überprüften Datenbestandes. Ein Gerät oder mehrere Geräte als Teil einer technischen Untersuchung können den Betrachtungsrahmen für die Berechnung bilden. Wie bei den vorherigen QI ist auch hier die Festlegung eines Kriteriums für Auffälligkeit notwendig, anhand dessen die Werteverteilungen der Befunder überprüft werden. Zu berücksichtigen ist zudem die Anzahl der Befunde pro Befunder bei der Interpretation des Wertes des QI (Nonnemacher et al. 2014, S. 31). Beeinflusst wird der QI durch die Qualifikation des Befunders und eine regelmäßig stattfindende Zweitbefundung, in deren Rahmen eine Rückmeldung stattfindet.

TMF-1052 - Werteverteilung von Parametern zwischen Zentren

Um systematische Auffälligkeiten zwischen den Datenwerten von Einrichtungen feststellen zu können, bietet die TMF-Leitlinie den QI „Werteverteilung von Parametern zwischen Zentren“. Der Wert des QI beschreibt den Anteil der Einrichtungen mit auffälliger Werteverteilung an der Gesamtzahl der einbezogenen Einrichtungen eines Datenbestandes. Für die Berechnung werden zunächst die einzubeziehenden Einrichtungen ausgewählt und das Kriterium für Auffälligkeit festgelegt. Im Beispiel der TMF-Leitlinie ist dies die Abweichung eines Verteilungsparameters um einen bestimmten Betrag. Anschließend erfolgt die Berechnung des Verteilungsparameters und die Identifikation auffälliger Verteilungen. Höhere Werte des QI deuten auf schlechte Datenqualität hin. Genannte Einflussfaktoren sind hierbei die jeweiligen Gegebenheiten der Einrichtungen, unter anderem auch Schulungsvoraussetzungen und Gerätespezifika (ebd., S. 32).

Wie die vorherigen QI geht dieser QI gemäß TMF-Leitlinie auf ein TMF-Projekt zurück (C. O. Schmidt et al. 2012c).

TMF-1008 - Untersuchungen am Wochenende

Dieser QI prüft, ob Datumsangaben einer Untersuchung auf ein Wochenende oder einen gesetzlichen Feiertag fallen. Der Wert des QI ergibt sich aus dem Anteil dieser Datumsangaben an der Gesamtzahl der überprüften Datumsangaben.

Durchführbar ist die Berechnung für einzelne Datensätze, Beobachtungseinheiten oder den Datenbestand der Einrichtung. Es sollten nur Datenelemente einbezogen werden, in denen Datumsangaben von Untersuchungen erfasst wurden (Nonnemacher et al. 2014, S. 33).

Ein hoher Wert des QI steht für schlechte Datenqualität, darüber hinaus sei jeder Fehler Hinweis auf das Vorliegen eines Datenfehler oder Datenfälschung. Die TMF-Leitlinie gibt weiter zu bedenken, dass in manchen Einrichtungen Untersuchungen auch an Wochenenden oder Feiertagen möglich seien, diese seien nicht als Datenfehler zu erachten. GleichermäÙen sei zu prüfen, ob Feiertage auch am Ort der Datenentstehung als solche galten. Als Schwellenwert wurden $> 2\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61).

Einfluss auf den Wert des QI kann über automatische Plausibilitätsprüfungen während der Datenerfassung genommen werden. Die TMF-Leitlinie verweist auf eine Quelle, auf Basis derer der QI in die TMF-Leitlinie übernommen wurde (Hasford 1994).

TMF-1012 - Fehlende Module

Module können in einem Datensatz mehrere Datenelemente zusammenfassen, beispielsweise nach thematischer Zugehörigkeit. Dieser QI beschreibt den Anteil der fehlenden Module an der Gesamtzahl der überprüften Module einer einzelnen Beobachtungseinheit oder des Datenbestands und prüft somit die Vollständigkeit der vorliegenden Module. Ein hoher Wert des QI zeugt von schlechter Datenqualität, beeinflusst wird der Wert laut TMF-Leitlinie durch die Compliance der Studienteilnehmer, deren Motivation sowie ihre Belastung durch die durchgeführten Untersuchungen (Nonnemacher et al. 2014, S. 34). Zusätzlich nehmen das Studiendesign und die Verfügbarkeit von Geräten für die notwendigen Untersuchungen Einfluss auf den Wert.

Der QI wurde gemäß TMF-Leitlinie auf Basis eines TMF-Projektes in die TMF-Leitlinie übernommen (C. O. Schmidt et al. 2012c).

TMF-1013 - Fehlende Werte bei Datenelementen

Um den Anteil fehlender Datenwerte bei Datenelementen an der Gesamtzahl der überprüften Datenwerte zu bestimmen kann dieser QI genutzt werden. Die Berechnung kann für einzelne Datensätze, Beobachtungseinheiten oder den Datenbestand durchgeführt werden. Der sich ergebende Wert lässt bei höheren Werten auf schlechte Datenqualität schließen. Der QI bezieht sowohl mandatorische als auch optionale Datenelemente mit in die Berechnung ein.

Als Einflussfaktoren des Wertes des QI werden EDC-Systeme genannt, die Einträge bei mandatorischen Datenelementen erzwingen können, sowie die Möglichkeit bei der Datenerfassung zu vermerken, dass das Datenelement nicht auf den Datensatz anwendbar war oder nicht erhoben werden konnte (Nonnemacher et al. 2014, S. 35).

Der QI wurde auf der Basis der Ergebnisse des TMF-Projektes V020-04 für die TMF-Leitlinie etabliert um eine ganzheitliche Betrachtung von fehlenden mandatorischen und optionalen Datenelementen zu ermöglichen.

TMF-1014 - Fehlende Werte bei mandatorischen Datenelementen

Der QI „Fehlende Werte bei mandatorischen Datenelementen“ bildet eine Unterkategorie des vorherigen QI und ist gemäß TMF-Leitlinie „wichtig zur korrekten Schätzung von Inzidenzen und Prävalenzen“ (ebd., S. 36). Die TMF-Leitlinie betont an dieser Stelle die Wichtigkeit der Vollständigkeit der Daten für die späteren Auswertungen. Der QI findet gemäß TMF-Leitlinie mehrfach Erwähnung in der Literatur (Aronsky und Haug 2000; Gaus 2003; D. Parkin et al. 1992; Swart et al. 2005). Die Unterkategorisierung in mandatorische und optionale Datenelemente durch die TMF-Leitlinie geht dabei ebenso auf Literaturquellen zurück und ist mit unterschiedlichen Schwellwerten des Wertes der QI begründet (Aronsky und Haug 2000). Andere genannte Quellen schlagen gemäß TMF-Leitlinie lediglich eine Fokussierung auf mandatorische Datenelemente vor (Gaus 2003).

Der Wert des QI berechnet sich aus dem Anteil der fehlenden Datenwerte bei mandatorischen Datenelementen an der Gesamtzahl der untersuchten Datenwerte bei mandatorischen Datenelementen. Ein hoher berechneter Wert ist folglich ein Zeichen schlechter Datenqualität. Ein Schwellenwert von $> 5\%$ wurde in der Implementierungshilfe von Version 1.0 der TMF-Leitlinie empfohlen (Nonnemacher et al. 2007, S, 61). Die Berechnung kann für einzelne Datensätze, Beobachtungseinheiten oder den Datenbestand der Einrichtung durchgeführt werden. Darüber hinaus gibt die TMF-Leitlinie noch eine alternative Berechnungsmöglichkeit mit Betrachtung der fehlenden Datenwerte pro Patient an.

Beeinflusst werden kann der Wert des QI durch die Ermöglichung der Angaben „nicht anwendbar“ oder „nicht bestimmt“.

TMF-1015 - Fehlende Werte bei optionalen Datenelementen

Ähnlich dem QI für fehlende Datenwerte bei mandatorischen Datenelementen bestimmt die Berechnung dieses QI den Anteil der fehlenden Datenwerte von optionalen Datenelementen an der Gesamtzahl der überprüften Datenwerte der optionalen Datenelemente. Die TMF-Leitlinie wiederholt die Anmerkung zum vorherigen QI, dass dieser QI „wichtig zur korrekten Schätzung von Inzidenzen und Prävalenzen“ sei (Nonnemacher et al. 2014, S. 38). Der QI bildet eine Unterkategorie von TMF-1013 und stammt gemäß TMF-Leitlinie aus den selben Quellen wie der vorherige QI (Aronsky und Haug 2000; Gaus 2003; D. Parkin et al. 1992; Swart et al. 2005). Die TMF-Leitlinie gibt die Empfehlung aus der Literatur wieder, fehlende Werte bei optionalen Datenelementen zu vernachlässigen (Gaus 2003). Der QI unterliegt den selben Einflussfaktoren wie sein Vorgänger. Gleiches gilt für die Aussagen, die anhand des Wertes des QI getroffen werden können. Als Schwellenwert wurden $> 5\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen

(Nonnemacher et al. 2007, S. 61). Rahmen der Berechnung sind ebenso einzelne Datensätze, Beobachtungseinheiten oder der Datenbestand der Einrichtung.

TMF-1016 - Datenelemente mit Wert unbekannt o.ä.

Für mandatorische Datenelemente findet im Rahmen dieses QI die Betrachtung von Datenwerten im Hinblick auf Merkmalsausprägungen wie „unbekannt o.ä.“ statt. Die Berechnung soll gemäß TMF-Leitlinie nur für mandatorische Datenelemente erfolgen, bei denen eine vorgegebene Merkmalsliste vorhanden ist, diese aber auch „unbekannt o.ä.“ umfasst (Nonnemacher et al. 2014, S. 39). Der QI steht daher in Verbindung zum QI TMF-1025 „Datenelemente mit unspezifischen Werten“. Rahmen der Berechnung kann die Betrachtung von einzelnen Datensätzen, Beobachtungseinheiten oder des Datenbestands sein. Für die Berechnung müssen zunächst die einzuschließenden Bezeichnungen („unbekannt o.ä.“) definiert werden. Anschließend werden die Datenwerte erfasst, die diese Merkmalsausprägung besitzen. Der Wert des QI entspricht der Anzahl dieser Datenwerte an der Gesamtzahl der betrachteten Datenwerte. Schlechte Datenqualität spiegelt sich daher in einem hohen Wert des QI wieder. Beeinflussbar ist der Wert des QI durch die Art der Erhebung sowie durch die Motivation des Eingabepersonals.

Der QI wurde gemäß TMF-Leitlinie auf Basis eines TMF-Projektes in die TMF-Leitlinie übernommen (C. O. Schmidt et al. 2012c).

TMF-1017 - Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten

Die Berechnung dieses QI bestimmt den Anteil von Datenelementen mit Datenwerten in allen Beobachtungseinheiten an allen überprüften Beobachtungseinheiten eines Datenbestands. Zur Berechnung des QI werden die einzubeziehenden Datenelemente ausgewählt, anschließend wird überprüft, ob für diese Datenelemente Datenwerte bei allen Beobachtungseinheiten existieren. Ein niedriger Wert des QI steht demzufolge für schlechte Datenqualität. Die Berechnung kann der TMF-Leitlinie zufolge für mandatorische als auch optionale Datenelemente durchgeführt werden. Vorab sei allerdings die Handhabung von Merkmalsausprägungen wie „unbekannt“ o.ä. festzulegen. Des Weiteren führt die TMF-Leitlinie aus, dass der Wert des QI von vielen Einflussfaktoren abhängt und „Aspekte anderer Kennzahlen bündelt“ (Nonnemacher et al. 2014, S. 40). Die Betrachtung des Wertes des QI sollte daher nicht einzeln, sondern mit Blick auf die anderen QI aus der Kategorie „fehlende Einträge“ erfolgen.

Der QI wurde gemäß TMF-Leitlinie auf Basis eines TMF-Projektes in die TMF-Leitlinie übernommen (C. O. Schmidt et al. 2012b).

TMF-1018 - Ausreißer bei stetigen Datenelementen

Bei stetigen Datenelementen, beispielsweise Laborwerten, kann es zu Ausreißern kommen, die außerhalb eines definierten Intervalls von erwarteten oder plausiblen Werten liegen. Besonders Vital- und Laborparameter sieht die TMF-Leitlinie als Ansatzpunkt für den QI (Nonnemacher et al. 2014, S. 41). Die TMF-Leitlinie übernimmt zu dieser Thematik den QI „Ausreißer bei stetigen Datenelementen“ unter Verweis auf zwei Quellen (Gaus 2003; Swart et al. 2005). Der QI berechnet den Anteil der festgestellten Ausreißer an der Gesamtzahl der überprüften Datenwerte. Die Berechnung kann für einzelne Datensätze, Beobachtungseinheiten oder den Datenbestand durchgeführt werden.

Für die Berechnung ist es notwendig, vorab ein Intervall zu definieren, bei dessen Verlassen Datenwerte als Ausreißer gelten. Die TMF-Leitlinie nennt als Beispiel für die Definition eines Intervalls den 1,5-fachen Interquartilsabstand des Box-Whisker-Plots. Darüber hinaus sei zu beachten, dass nicht jeder Ausreißer auch ein Datenfehler sei, da besonders bei schwer erkrankten Personen Werte außerhalb der Intervalle auftreten könnten. Eine hohe Anzahl an Ausreißern ist ein Zeichen schlechter Datenqualität. Beeinflusst werden kann der Wert des QI laut TMF-Leitlinie durch die Standardisierung von verwendeten Untersuchungsverfahren und Analysemethoden. Als Schwellenwert wurden $> 10\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61).

TMF-1019 - Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten

Im Rahmen von Untersuchungen werden technische Verfahren eingesetzt, deren Genauigkeit physikalischen Grenzen unterliegt. Der Wert dieses QI entspricht dem Anteil der Datenwerte, die außerhalb der Messbarkeitsgrenzen liegen, an der Gesamtzahl der überprüften Datenwerte. Für die Berechnung des QI ist es gemäß TMF-Leitlinie erforderlich, Datenwerte identifizieren zu können, welche die Messbarkeitsgrenzen unter- oder überschreiten (Nonnemacher et al. 2014, S. 43). Die Messbarkeitsgrenzen der spezifischen Verfahren müssen dabei von Relevanz für die Praxis sein. Es kommen zudem nur technische Verfahren für die Berechnung in Betracht. Ziel der Berechnung können einzelne Datensätze, Beobachtungseinheiten oder ein Datenbestand sein.

Die Berechnung selbst erfolgt zunächst durch die Auswahl des jeweiligen Verfahrens mit der anschließenden Suche nach Datenwerten, welche die Messbarkeitsgrenzen dieses Verfahrens unter- oder überschreiten. Abschließend erfolgt die Berechnung des Wertes des QI anhand der ermittelten Datenwerte. Hohe Werte des QI stehen für schlechte Datenqualität, der Wert des QI kann gemäß TMF-Leitlinie durch korrekte Indikationsstellung sowie eine verfahrensgerechte Durchführung der Untersuchungen beeinflusst werden.

Auf Basis eines TMF-Projektes wurde der QI in die TMF-Leitlinie übernommen (C. O. Schmidt et al. 2012b).

TMF-1020 - Werte aus Standards

Datenelemente können die Möglichkeit bieten, neben standardisierten Auswahllisten auch die Ergänzung der Erfassung durch einen Freitext anzubieten oder gänzlich auf Freitextangaben zu setzen. Ziel der Berechnung dieses QI ist es festzustellen, wie hoch der Anteil der Datenwerte mit Bezeichnungen aus standardisierten Vokabularen an der Gesamtzahl der überprüften Datenwerte ist. Viele nicht-standardisierte Datenwerte resultieren entsprechend in einem niedrigen Wert des QI und sind somit ein Hinweis auf schlechte Datenqualität. Bei der Betrachtung des Wertes des QI sei zu berücksichtigen, wieso auf die Nutzung eines kontrollierten Vokabulars für das jeweilige Datenelement verzichtet wurde (Nonnemacher et al. 2014, S. 44). Die Berechnung sollte nur für Freitextfelder oder Felder erfolgen, deren Text ergänzt werden kann.

Für die Berechnung ist zunächst die Festlegung der Vokabulare der Datenelemente notwendig. Die Datenwerte des in die Berechnung eingehenden Datenbestands werden durch einen Textvergleich normalisiert; das Verfahren zum Textvergleich sollte auf Anraten der TMF-Leitlinie eingehend geprüft sein. Abschließend werden alle Datenwerte erfasst, die im vorher festgelegten Vokabular vorhanden sind. Diese Datenwerte werden in den Zähler der Berechnungsformel des QI eingebracht.

Der Wert des QI hängt dabei von der Verfügbarkeit von kontrollierten Vokabularen für die betrachteten Datenelemente ab. Die TMF-Leitlinie nennt ein TMF-Projekt als Quelle auf dessen Basis der QI in die TMF-Leitlinie übernommen wurde (C. O. Schmidt et al. 2012b).

TMF-1021 - Unerlaubte Werte bei qualitativen Datenelementen

Bei qualitativen Datenelementen kann es möglich sein, unerlaubte Datenwerte anzugeben. Der Wert dieses QI beschreibt den Anteil der unerlaubten Datenwerte an der Gesamtzahl der überprüften Datenwerte. Für die Berechnung werden zunächst die erlaubten Datenwerte festgelegt, anschließend werden alle Datenwerte nach unerlaubten Datenwerten durchsucht und beide Anzahlen in Zähler und Nenner in die Berechnung eingebracht. Ein hoher Wert des QI ist Zeichen schlechter Datenqualität. Berechnungsrahmen kann die Betrachtung einzelner Datensätze, Beobachtungseinheiten oder des Datenbestands sein. Der QI betrachtet dabei nur die formale, nicht jedoch die inhaltliche Korrektheit (Nonnemacher et al. 2014, S. 45).

Die Berechnung wird nur empfohlen, wenn es technisch möglich sei, unerlaubte Datenwerte zu erfassen. Als Einflussfaktor nennt die TMF-Leitlinie daher automatische Bereichsprüfungen der Datenwerte während der Datenerfassung.

Der QI wurde gemäß TMF-Leitlinie auf Basis mehrerer Quellen in die TMF-Leitlinie übernommen (Gaus 2003; Swart et al. 2005; M. Winter et al. 2003).

TMF-1022 - Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings

Neben der Betrachtung von unerlaubten Datenwerten bei qualitativen Datenelementen betrachtet die TMF-Leitlinie auch unerlaubte Datenwerte bei qualitativen Datenelementen zur Kodierung von Missings. Gleich dem vorherigen QI bilden einzelne Datensätze, Beobachtungseinheiten oder der Datenbestand den Berechnungsrahmen. Voraussetzung für die Berechnung des QI ist das Auftreten und das Bestehen der Möglichkeit der Kodierung von Missings (Nonnemacher et al. 2014, S. 47).

Der Wert des QI setzt sich aus dem Anteil der unerlaubten Datenwerte zur Kodierung von Missings an der Gesamtzahl der überprüften Datenwerte zusammen. Für die Berechnung ist die vorausgehende Identifikation der Datenwerte notwendig, welche eine unerlaubte Kodierung von Missings darstellen. Ein hoher Wert des QI steht für schlechte Datenqualität, beeinflussbar sei diese durch die automatische Kodierung von Missings mit vordefinierten Kodierungen.

Auf Basis eines TMF-Projektes wurde der QI in die TMF-Leitlinie übernommen (C. O. Schmidt et al. 2012c).

TMF-1023 - Unerlaubte Werte zur Kodierung von fehlenden Modulen

Neben der Kodierung von Missings kann auch die Möglichkeit bestehen, das Fehlen von Modulen zu erfassen. Bei dieser Erfassung kann es wiederum zur Angabe unerlaubter Datenwerte zur Kodierung von fehlenden Modulen kommen. Der Wert dieses QI beschreibt den Anteil der unerlaubten Datenwerte zur Kodierung von fehlenden Modulen an der Gesamtzahl der überprüften Datenwerte. Die Berechnung kann für einzelne Datensätze, Beobachtungseinheiten oder einen Datenbestand ausgeführt werden, ein hoher Wert ist Zeichen schlechter Datenqualität.

Für die Berechnung des QI ist es zunächst erforderlich, die Menge der erlaubten Datenwerte festzulegen, um unerlaubte Datenwerte identifizieren zu können. Anschließend werden alle Datenwerte, die nicht Teil der erlaubten Menge sind, als „unerlaubt“ markiert und als solche in die Berechnung einbezogen. Wie beim vorherigen QI kann der Wert des QI durch die automatische Kodierung von fehlenden Daten beeinflusst werden.

Als Quelle des QI nennt die TMF-Leitlinie ein TMF-Projekt (ebd.).

TMF-1024 - Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen

Bei quantitativen Datenelementen kann es möglich sein, Unter- oder Überschreitungen von Messbarkeitsgrenzen zu kodieren. Der Wert dieses QI entspricht dem Anteil der zur Kodierung verwendeten, unerlaubten Datenwerte an der Gesamtzahl der überprüften Datenwerte. Für die Berechnung ist es erforderlich festzulegen, welche Datenwerte zur Kodierung erlaubt sind. Auf Basis dieser Wertemenge werden anschließend alle Datenwerte

des Datensatzes, der Beobachtungseinheit oder des Datenbestands überprüft. Unerlaubte Datenwerte seien Datenfehler und somit ein hoher Wert des QI Zeichen schlechter Datenqualität (Nonnemacher et al. 2014, S. 49). Die Berechnung kann für einzelne Datensätze, Beobachtungseinheiten oder einen Datenbestand ausgeführt werden. Dieser QI ist gemäß TMF-Leitlinie verwandt mit dem QI TMF-1019 „Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten“. Die automatische Kodierung von Datenwerten, die Messbarkeitsgrenzen unter- oder überschreiten, kann den Wert des QI positiv beeinflussen. Auf Basis eines TMF-Projektes wurde der QI in die TMF-Leitlinie übernommen (C. O. Schmidt et al. 2012c).

TMF-1025 - Datenelemente mit unspezifischen Werten

Unspezifische Datenwerte zeugen von schlechter Datenqualität, da hier eine mangelnde Informationsdichte sichtbar wird (Nonnemacher et al. 2014, S. 50). Dieser QI betrachtet den Anteil der Datenelemente mit unspezifischen Angaben an der Anzahl der überprüften Datenelemente. Für die Berechnung ist es notwendig, die einzubeziehenden Datenelemente festzulegen und eine Kategorie „unspezifisch“ für diese Datenelemente zu definieren. Die Berechnung kann für einzelne Beobachtungseinheiten oder einen Datenbestand ausgeführt werden. Eine Berechnung auf der Ebene einzelner Datensätze könne gemäß TMF-Leitlinie gegebenenfalls ignoriert werden, da durch die Zusammenführung von mehreren Datensätzen unspezifische Angaben durch spezifische ersetzt werden könnten.

Einflussfaktoren des QI sind gemäß TMF-Leitlinie die Genauigkeit der Diagnostik sowie Vollständigkeit und Korrektheit der primären Datenquellen.

Die TMF-Leitlinie nennt zwei Quellen für den QI (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a).

TMF-1026 - Beobachtungseinheiten mit unbekanntem Primärtumor

Gemäß TMF-Leitlinie stammt der QI „Beobachtungseinheiten mit unbekanntem Primärtumor“ aus dem Bereich der Krebsregister (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a). Der QI betrachtet den Anteil der Beobachtungseinheiten mit unbekanntem Primärtumor an der Gesamtheit der überprüften Beobachtungseinheiten. Der Betrachtungsrahmen kann aus einer einzelnen Beobachtungseinheit oder einem Datenbestand bestehen. In beiden Fällen ist für die Berechnung die Identifikation der Beobachtungseinheiten mit unbekanntem Primärtumor Ausgangspunkt der Berechnung (Nonnemacher et al. 2014, S. 52). Der berechnete Wert des QI steht mit steigendem Wert für schlechtere Datenqualität. Die TMF-Leitlinie gibt als Grenze einen akzeptablen Anteil von 5% an. Einflussfaktoren für den Wert sind die Genauigkeit der Diagnostik sowie Vollständig- und Korrektheit der Primärquellen der Daten.

TMF-1027 - Nachweis bekannter Korrelationen

Zwischen zwei oder mehreren quantitativen Datenelementen können statistische Beziehungen existieren. Dieser QI betrachtet die Nachweisbarkeit von Korrelationen der quantitativen Datenelemente eines Datenbestands. Hierfür ist es im Rahmen der Berechnung zunächst notwendig, die korrelierenden Datenelemente zu identifizieren und das statistische Verfahren zum Nachweis der Korrelation zu bestimmen, das im nächsten Schritt auf die Datenelemente angewendet wird. Der Wert des QI setzt sich aus der Anzahl der nicht nachweisbaren Korrelationen an der Gesamtzahl der überprüften Korrelationen zusammen. Ein hoher Wert steht daher für schlechte Datenqualität. Für die TMF-Leitlinie ist das Scheitern der Nachweisbarkeit ein Hinweis auf Datenfehler auf Datenfälschung (ebd., S. 53). Nicht nachweisbare Korrelationen sollten daher auf Anraten der TMF-Leitlinie einer weiteren Prüfung unterzogen werden. Als Schwellenwert wurden $> 10\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61). Die TMF-Leitlinie gibt keine Einflussfaktoren für diesen QI an und nennt eine Quelle auf Basis derer der QI in die TMF-Leitlinie übernommen wurde (Hasford 1994).

TMF-1050 - Umfang der Metadaten bei Untersuchungen

Bei übermittelten Metadaten zu technischen Untersuchungen kann die Abwesenheit von Parametern als Zeichen schlechter Datenqualität interpretiert werden. Der Wert dieses QI entspricht dem Anteil der tatsächlich vorhandenen Parameter einer Untersuchung an der Gesamtzahl der erwarteten Parameter zu dieser Untersuchung. Die Berechnung des QI sollte nur erfolgen, wenn das Weglassen von Parametern bei der technischen Übermittlung überhaupt möglich ist (Nonnemacher et al. 2014, S. 54). Ein niedriger Wert des QI steht für schlechte Datenqualität. Beeinflusst wird der Wert durch den Umfang der verwendeten Schnittstellen sowie durch die Einstellungen der verwendeten Geräte. Im Gegensatz zu den vorherigen QI werden für diesen QI nicht die Daten selbst betrachtet, sondern der Prozess der Übermittlung und somit die Prozessqualität.

Als Quelle des QI nennt die TMF-Leitlinie ein TMF-Projekt (C. O. Schmidt et al. 2012b).

Qualitätsindikatoren der Ebene Organisation

Während die Indikatoren der Ebene Richtigkeit nahezu ausschließlich die Strukturqualität der Daten betrachten, treffen die Indikatoren der Ebene Organisation ganzheitlich Aussagen zur Prozessqualität.

TMF-1028 - Aktualität der gespeicherten Daten

Die Aussagekraft von Auswertungen hängt, gemäß Leitlinie, unter anderem von der Aktualität der gespeicherten Daten ab (Nonnemacher et al. 2014, S. 55). Der QI „Aktualität der gespeicherten Daten“ betrachtet die Aktualität von veränderlichen Datenelementen einzelner Datensätze, Beobachtungseinheiten oder eines Datenbestands. Bei unveränderlichen

soziodemografischen Datenelementen, beispielsweise dem Geburtsdatum und -ort, ist der QI nicht anwendbar, gleichermaßen bei unveränderlichen krankheitsbezogenen Daten wie Krankenhausaufenthalten etc. Im Hinblick auf veränderliche, krankheitsbezogene Daten, beispielsweise Medikation, wird dem QI bedingte Anwendbarkeit zugewiesen, wenn diese Daten aus alten Quellen nachvollzogen werden müssen. Im Rahmen von Feasibilitystudien ist der QI gegebenenfalls als wichtig zu erachten, da nur bei aktuellen Daten über den Ein- und Ausschluss von Patienten entschieden werden kann. Die TMF-Leitlinie betont des Weiteren die Wichtigkeit des QI bei unerwünschten Ereignissen im Rahmen von Medikations- und Therapiestudien sowie bei meldepflichtigen Erkrankungen.

Als alternative Definition kann auch „Aktualität“, gemäß Leitlinie, bei Daten, die sich über die Zeit ändern, auch als die Frage nach dem Alter der Daten verstanden werden (Naumann und Rolker 2000; Wang et al. 1993). Die TMF-Leitlinie nennt mehrere Quellen auf deren Basis der QI in die TMF-Leitlinie übernommen wurde (Abate et al. 1998; Bobrowski et al. 1999; Jung und R. Winter 2000; Katalinic 2005; Laporte, Mazumdar et al. 1994; Naumann und Rolker 2000; Wang et al. 1993).

Die TMF-Leitlinie beschreibt zwei Berechnungsmöglichkeiten:

Zum einen die Berechnung im Hinblick auf den Eingang der Daten in das Register (Katalinic 2005). Hierfür wird zunächst eine akzeptable Zeitdifferenz zwischen Eingang der Daten in das Register und dem Zeitpunkt des Entstehungszeitpunktes der Daten festgelegt. Anschließend wird die Zeitdifferenz für alle Werte berechnet. Werte, deren Zeitdifferenz das akzeptable Maß überschreiten, werden im Zähler in die Berechnung der Rate eingebracht. Im Nenner befindet sich die Anzahl der überprüften Werte. Ein hoher Wert zeugt von schlechter Datenqualität.

Die alternative Berechnung betrachtet die Aktualität der Daten im Hinblick auf die in das Register eingegangenen Meldungen eines festzulegenden Zeitintervalls, gemessen anhand der von Inzidenzen/Prävalenzen erwarteten Anzahl an Meldungen (Laporte, Mazumdar et al. 1994). Der hieraus berechnete Wert wird als schlecht erachtet, wenn er unter der akzeptablen Aktualitätsrate liegt.

Als Einflussfaktoren sind die personelle Kapazität der Einrichtung, die Qualität der technischen Infrastruktur sowie die Güte der Datenquelle selbst genannt. Als Schwellenwert wurden $> 10\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61).

TMF-1029 - Dubletten im Datenbestand

In den Daten von Registern können sogenannte Dubletten auftreten, die TMF-Leitlinie definiert diese als „zwei oder mehr Informationseinheiten, die dasselbe Subjekt oder Objekt beschreiben und bei denen die Identität nachgewiesen werden kann oder - sofern der Nachweis nicht möglich ist - mit hinreichender Wahrscheinlichkeit angenommen werden kann.“ (Nonnemacher et al. 2014, S. 57).

Um Dubletten festzustellen, werden nach Empfehlung der TMF-Leitlinie Sets von Datenelementen gebildet. Als Beispiel nennt die TMF-Leitlinie ein Set aus Geschlecht, Geburtsdatum, Wohnort, Infektionsdatum und Prozeduren. Anhand dieses Sets werden Beobachtungseinheiten verglichen, bei einer übereinstimmenden Mindestzahl an Datenelementen des Sets geht die Informationseinheit als Dublette in die weitere Berechnung des Wertes des QI ein. Der Wert des QI entspricht dabei der Anzahl der erkannten Dubletten an der Gesamtzahl der überprüften Sets. Ein hoher Wert steht daher für schlechte Datenqualität. Die TMF-Leitlinie bezeichnet den als akzeptabel vorgeschlagenen Wert von $<0,01\%$ der genannten Literaturquelle des QI als zu rigide (Swart et al. 2005). Als realistischer wird der Wert $2,5\%$ auf Basis von Berechnungen des Kompetenznetz Parkinson erachtet.

Die genannten Einflussfaktoren für Dubletten sind vielfältig, besonders hervorgehoben werden die Übernahme von Daten aus verschiedenen Quellen sowie Patienten, die in mehreren Einrichtungen vorstellig werden oder die Einrichtung wechseln. Hinzukommen kann ein hoher Durchlauf an Patienten in der Einrichtung, so dass in der Einrichtung nicht ersichtlich ist, welche Patienten bereits erfasst wurden. In Version 1.0 der TMF-Leitlinie wurde $> 5\%$ als Schwellenwert empfohlen (Nonnemacher et al. 2007, S. 61).

TMF-1030 - Rekrutierungsrate

Die Rekrutierungsrate eines Registers beschreibt die Anzahl der in das Register aufgenommenen Personen, welche die Einschlusskriterien erfüllen, in einem bestimmten Zeitintervall gemessen an der Anzahl der Personen, die in diesem Zeitintervall in den überprüften Datenbestand hätten eingehen können (Nonnemacher et al. 2014, S. 59). Das Zeitintervall ist gemäß der vorgeschlagenen Berechnung der TMF-Leitlinie frei zu wählen. Die TMF-Leitlinie bietet eine alternative Berechnungsmöglichkeit an, unter Einbezug einer festzulegenden möglichen Abweichung der Rekrutierungsrate, welche durch den Vergleich der Rekrutierungsraten zweier Zeitintervalle berechnet wird. Wenn die Abweichung zu groß ausfällt, sollte dies als Hinweis auf Rekrutierungsprobleme interpretiert werden. Im Zuge der ersten Berechnungsvariante ist ein hoher Wert des QI als gute Datenqualität zu verstehen. Die TMF-Leitlinie merkt an dieser Stelle zudem an, dass sowohl Schwere der Erkrankung als auch Schwierigkeiten im Hinblick auf die Einwilligung des Patienten, beispielsweise durch aufwändige Untersuchungen, zur Erhöhung der Aussagekraft des QI berücksichtigt werden müssten. Diese Punkte werden daher auch als Einflussfaktoren des Wertes genannt, hinzu kommt die Kapazität der jeweiligen Einrichtung, Personal für die Rekrutierung bereitzustellen. Als Schwellenwert wurden $< 90\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61).

In der TMF-Leitlinie werden mehrere Literaturquellen für den QI genannt (Goldberg et al. 1980; Hasford 1994; Laporte, Mazumdar et al. 1994; Pogash et al. 2001; Pollock 1994; Svolba und Bauer 1999). Aus diesen Quellen wurden zwei weitere alternative Berechnungsmöglichkeiten in die TMF-Leitlinie übernommen. Zum einen die Schätzung der Vollständigkeit

der Registerpopulation anhand der Extrapolation einer Stichprobe des Registers auf das gesamte Register (Goldberg et al. 1980). Zum anderen die Schätzung der zu erwartenden Fallzahl über bekannte Inzidenz- und Prävalenzraten mit Abgleich der zu erwartenden Patientenzahl und der berechneten Zahl; die festgestellte Differenz ist als Maß der Vollständigkeit der Registerpopulation zu interpretieren (Goldberg et al. 1980; Laporte, Mazumdar et al. 1994).

TMF-1051 - DCO-Rate (Death Certificate Only)

Eine Unterkategorie des vorherigen QI ist der QI „DCO-Rate“ der den Anteil der gemeldeten Krebserkrankungen, für die nur eine Todesbescheinigung mit Hinweis auf die Krebserkrankung vorliegt, an der Gesamtpopulation des Registers berechnet. Die TMF-Leitlinie hebt den QI als „wichtigsten Indikator für epidemiologische Krebsregister“ hervor und verweist auf mehrere Forschungsprojekte und Datensammlungen, welche diesen QI als maßgeblichen Wert zur Entscheidung hinsichtlich der Datenübernahme aus einer Quelle verwenden (Nonnemacher et al. 2014, S. 61).

Ein hoher Wert ist Zeichen schlechter Datenqualität, die TMF-Leitlinie gibt anhand der genannten Primärquellen des QI eine DCO-Rate von unter 5% als akzeptabel an (Hentschel und Katalinic 2008). Im Zuge der Betrachtungen sollte die Referenzrate gemäß TMF-Leitlinie jedoch für jeden Tumor spezifisch gewählt werden. Des Weiteren sei zu bedenken, dass junge Register naturgemäß zunächst eine hohe DCO-Rate aufweisen, da nicht alle versterbenden Patienten auch zu Beginn der Erkrankung im Register registriert wurden. Einfluss auf den Wert des QI nehmen daher die Meldebereitschaft der jeweiligen Stellen und die fehlerhafte Zusammenführung von Meldungen.

Die TMF-Leitlinie verweist weiterhin auf ein TMF-Projekt auf dessen Basis der QI in die TMF-Leitlinie übernommen wurde (C. O. Schmidt et al. 2012a).

TMF-1031 - Verweigerungsrate von Untersuchungen

Bei Registern, welche die Verweigerung von Untersuchungen dokumentieren, kann der Anteil der verweigerten Untersuchungen an der Gesamtzahl der überprüften Untersuchungen bestimmt werden. Ergebnis dieser Berechnung ist der Wert dieses QI, der bei hohen Beträgen als Zeichen schlechter Datenqualität zu interpretieren ist. Der QI sollte nur angewendet werden, wenn für die einbezogenen Untersuchungen eine Verweigerung dokumentiert werden kann (Nonnemacher et al. 2014, S. 63). Die Berechnung des QI kann sowohl für einzelne als auch zusammengefasst für mehrere Untersuchungen erfolgen. Den Berechnungsrahmen des QI bildet immer ein Datenbestand. Der Wert des QI unterliegt dabei dem Einfluss des Studiendesigns, den Anreizen für die Studienteilnehmer sowie den Belastungen, denen sie ausgesetzt sind.

Als Quelle des QI nennt die TMF-Leitlinie ein TMF-Projekt (C. O. Schmidt et al. 2012c).

TMF-1032 - Verweigerungsrate von Modulen

Gleich der Verweigerungsrate von Untersuchungen kann ebenso die Verweigerungsrate von Modulen bestimmt werden. Die Berechnung des Wertes des QI erfolgt für einen Datenbestand. Betrachtet wird der Anteil der verweigerten Module an der Gesamtzahl der überprüften Module. Ein hoher Wert des QI zeugt von schlechter Datenqualität. Wie beim vorherigen QI können nur Module berücksichtigt werden, deren Verweigerung explizit dokumentiert werden kann (Nonnemacher et al. 2014, S. 64). Die Verweigerungsrate der Module ist dabei über Studiendesign und Anreize für die Teilnehmer beeinflussbar.

Der QI geht gemäß TMF-Leitlinie auf ein TMF-Projekt zurück (C. O. Schmidt et al. 2012c).

TMF-1033 - Verweigerungsrate von einzelnen Datenelementen

Eine feingranularere Betrachtung als die Verweigerungsrate von Modulen erlaubt der QI „Verweigerungsrate von einzelnen Datenelementen“. Für die Berechnung des Wertes des QI wird die Anzahl der Datenwerte mit Verweigerung an der Gesamtzahl der untersuchten Datenwerte eines Datenbestands berechnet. Gleich den beiden vorherigen QI ist die Berechnung nur möglich, wenn die Verweigerung tatsächlich dokumentiert wurde (Nonnemacher et al. 2014, S. 65). Ein hoher Wert des QI zeugt von schlechter Datenqualität. Die TMF-Leitlinie merkt an, dass die zu überprüfenden Datenelemente größtenteils aus Befragungen stammen werden und weniger aus Datenelementen, die beispielsweise in Zusammenhang mit technischen Untersuchungen stehen. Beeinflusst wird der Wert des QI durch die Art der Befragung und die Schulung des jeweiligen Interviewers. Zudem spielen die Sensibilität der Datenelemente eine Rolle.

Die TMF-Leitlinie nennt eine Quelle auf Basis derer der QI in die TMF-Leitlinie übernommen wurde (C. O. Schmidt et al. 2012c).

TMF-1034 - Vorzeitig aus dem Register ausscheidende Beobachtungseinheiten

Patienten können vorzeitig aus Registern ausscheiden, etwa durch Zurücknahme der Einwilligungserklärung, Abbrechen des Kontakts mit der Einrichtung (lost-to-follow-up) oder das Ableben des Patienten. Besonders bei Verlaufsregistern und Registern, welche für Feasibilitätsstudien genutzt werden, ist eine möglichst niedrige Drop-out-rate wichtig (Nonnemacher et al. 2014, S. 66). Der QI ist laut TMF-Leitlinie nur für diese beiden Registerformen relevant. Dieser QI beschreibt den Anteil der Patienten, die innerhalb eines Zeitintervalls vorzeitig aus dem Register ausgeschiedenen sind, an der in diesem Zeitintervall in das Register aufgenommenen Patienten. Wenn zu viele Krankheitsverläufe unvollständig erfasst werden, sinkt die Datenqualität und damit die Aussagekraft der Daten.

Der Indikator kann unter dem Aspekt des Ablebens des Patienten auch medizinische Aussagekraft besitzen. Eine hohe Dunkelziffer bei Todesfällen sei in diesem Fall bei Alterskrankheiten zu erwarten. Ein weiterer Aspekt ist die Interpretation des Wertes des QI im

Hinblick auf die „Einwilligungsschwere“ und das für die Patienten belastende Studiendesign.

Die Berechnung des QI erfolgt durch die Festlegung eines Zeitintervalls als Betrachtungsrahmen. Aus dem Datenbestand werden alle in diesem Zeitintervall vorzeitig aus dem Register ausscheidenden Patienten extrahiert. Der Anteil dieser Patienten an der Gesamtzahl der in diesem Zeitintervall neu in das Register eingegangenen Patienten bildet den Wert des QI. Der Wert kann durch engeren Kontakt zum Patienten, eingehende Information des Patienten zur Wichtigkeit des Forschungsvorhabens und seiner Teilnahme daran sowie die Verfügbarkeit von Personal in der Einrichtung beeinflusst werden.

Alternativ können für die Berechnung des QI Todesfälle aus dem Zähler ausgeschlossen werden. Die Drop-out-Rate ist als problematisch zu erachten, wenn ein festzulegender Schwellenwert überschritten wird. Als Beispiel für eine Drop-out-Rate wurde aus dem Kompetenznetz Parkinson 0,013% berechnet. In Version 1.0 der TMF-Leitlinie wurde $> 1\%$ als Schwellenwert empfohlen (Nonnemacher et al. 2007, S. 61).

Für den QI nennt die TMF-Leitlinie keine Literaturquellen, die Expertise des IMIBE wird als Quelle genannt.

TMF-1036 - Synonyme

Im Zusammenhang mit Patientenregistern definiert die TMF-Leitlinie Synonyme als Patientenidentifikatoren, die sich lexikalisch unterscheiden, jedoch dem selben Patienten zugeordnet sind (Nonnemacher et al. 2014, S. 68). Synonyme können zur Mehrfachzählung von Patienten und somit zur falschen Einschätzung von Inzidenzen und Prävalenzen führen. Der QI geht laut TMF-Leitlinie auf zwei Quellen zurück (D. Parkin et al. 1992; M. Winter et al. 2003). Die TMF-Leitlinie greift die Empfehlung einer der Quellen auf, die Beziehungen zwischen Synonymen zu dokumentieren (M. Winter et al. 2003).

Der Wert des QI setzt sich aus der Anzahl der festgestellten Synonyme an der Gesamtzahl der überprüften Patienten zusammen. Ein hoher Wert steht entsprechend für schlechte Datenqualität. In Version 1.0 der TMF-Leitlinie wurde $> 5\%$ als Schwellenwert empfohlen (Nonnemacher et al. 2007, S. 61). Für die Berechnung ist es erforderlich, ähnlich des QI „TMF-1029 - Dubletten im Datenbestand“, über ein Set von Datenelementen mehrfach im Register gespeicherte Patienten zu erkennen. Als Alternative zu diesem Vorgehen werden Verfahren des Record Linkage genannt.

Beeinflusst werden kann der Wert durch die Nutzung von Algorithmen zur Erzeugung der Patientenidentifikatoren sowie die Qualität der technischen Infrastruktur, besonders im Hinblick auf Vorgänge wie Einrichtungswechsel und parallele Betreuung der Patienten in mehreren Einrichtungen.

TMF-1037 - Homonyme

Die TMF-Leitlinie definiert Homonyme als Patientenidentifikatoren, die mehr als einem Patienten zugeordnet sind (Nonnemacher et al. 2014, S. 70). Der Wert des QI entspricht dem Anteil der Homonyme an der Gesamtzahl der überprüften Patienten eines Datenbestands. Für die Berechnung des Wertes des QI ist es zunächst erforderlich, die Kriterien zu definieren, an denen die Homonyme erkannt werden sollen. Als Beispiele genannt werden die zur Generierung von Patientenidentifikatoren genutzten Datenelemente sowie die Erkennung von nicht plausiblen Bezügen zwischen Datenelementen. Beispielsweise existiert kein Bezug zwischen Therapie und Diagnose. Anhand der Kriterien sind anschließend Homonyme im Rahmen der Berechnung zu identifizieren. Ein hoher Wert des QI ist Zeichen schlechter Datenqualität. Ein Schwellenwert von $> 2\%$ wurde in Version 1.0 der TMF-Leitlinie als Schwellenwert empfohlen (Nonnemacher et al. 2007, S. 61). Dies liegt darin begründet, dass Homonyme zur Zusammenführung von Daten verschiedener Patienten führen können und somit widersprüchliche oder falsche Informationen darstellen. Homonyme nehmen daher auch Einfluss auf Inzidenzen und Prävalenzen. Der Wert des QI kann durch Algorithmen zur Erzeugung von Patientenidentifikatoren beeinflusst werden. Die Literaturquellen zum QI „Homonyme“ entsprechen laut TMF-Leitlinie denen des QI „Synonyme“ (D. Parkin et al. 1992; M. Winter et al. 2003).

TMF-1038 - Einzelmeldungen pro Beobachtungseinheit

Für Krebsregister ist es wichtig festzustellen, wie viele Beobachtungseinheiten nur von einer meldenden Stelle gemeldet wurden, da dies einen Hinweis auf unvollständige oder fehlende Meldungen und somit mangelnde Datenqualität geben kann (Nonnemacher et al. 2014, S. 72). Für den Wert des QI werden daher alle Beobachtungseinheiten mit nur einer Meldung identifiziert und deren Anteil an der Gesamtzahl der überprüften Beobachtungseinheiten berechnet. Ein hoher Wert des QI ist Zeichen schlechter Datenqualität. Die Berechnung des QI kann für einzelne Beobachtungseinheiten oder einen Datenbestand durchgeführt werden.

Beeinflusst wird der Wert durch die Motivation der meldenden Stellen sowie die Vollständigkeit der Meldungen selbst. Der QI geht dabei laut TMF-Leitlinie auf zwei Quellen zurück (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a).

TMF-1039 - Solitäre Meldungen von Pathologen

Eine Spezialisierung der vorherigen QI ist der QI „Solitäre Meldungen von Pathologen“, der den Anteil der Beobachtungseinheiten, für die nur eine solitäre Meldung eines Pathologen vorliegt, an der Gesamtzahl der überprüften Beobachtungseinheiten betrachtet. Die Berechnung des QI sollte für einen Datenbestand erfolgen. Gleich dem vorherigen QI steht ein hoher Wert des QI bei der Überprüfung des Datenbestandes für schlechte Datenqualität. Die TMF-Leitlinie führt aus, dass dies an wichtigen Informationen zur Erkrankung der

Patienten liegt, die in den Meldungen der Pathologen oftmals fehlen (Nonnemacher et al. 2014, S. 73). Beeinflusst wird der Wert des QI durch die Motivation der meldenden Stellen sowie die Vollzähligkeit der Meldungen selbst. Der QI geht gemäß TMF-Leitlinie auf dieselben Quellen wie der übergeordnete QI TMF-1038 zurück (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a).

TMF-1040 - Zurückgewiesene Meldungen

Bei fehlerhaften Meldungen kann es vorkommen, dass die Verarbeitung der Meldung aufgrund der Fehler nicht erfolgen konnte und die Meldung somit zurückgewiesen werden musste. Der Wert dieses QI entspricht dem Anteil der zurückgewiesenen Meldungen an der Gesamtzahl der überprüften Meldungen. Ein hoher Wert ist daher ein Zeichen schlechter Datenqualität, die Berechnung des QI sollte für einen Datenbestand erfolgen (Nonnemacher et al. 2014, S. 74). Nach Korrektur der fehlerhaften Daten der Meldung kann diese erneut vorgenommen werden. Einflussfaktoren auf den Wert des QI sind sowohl Vollständig- als auch Korrektheit der primären Datenquelle der Meldung als auch die Motivation der meldenden Stelle.

Die TMF-Leitlinie nennt zwei Quellen für den QI (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a).

TMF-1041 - Datenquellen pro Beobachtungseinheit

Meldungen zu Tumorfällen können aus mehreren Datenquellen vorliegen. Die TMF-Leitlinie geht an dieser Stelle davon aus, dass nicht alle Meldungen zwangsläufig vollständig sind und somit wichtige Daten möglicherweise in manchen Datenquellen fehlen (Nonnemacher et al. 2014, S. 75). Entsprechend deutet nur eine Datenquelle pro Beobachtungseinheit auf schlechte Datenqualität hin. Der Wert des QI setzt sich daher aus der Anzahl der Beobachtungseinheiten mit nur einer Datenquelle an der Gesamtzahl der überprüften Beobachtungseinheiten zusammen. Die Berechnung des QI sollte für einzelne Beobachtungseinheiten oder einen Datenbestand erfolgen. Maßgebliche Einflussfaktoren des Wertes sind sowohl die Motivation der meldenden Stelle als auch der Zeitpunkt der Berechnung des QI, da alle Meldungen nicht zwangsläufig zeitgleich eintreffen. Der QI geht laut TMF-Leitlinie auf zwei Quellen zurück (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a).

TMF-1042 - Beobachtungseinheiten mit Follow-up

Der Anteil der Beobachtungseinheiten mit Follow-up an der Gesamtheit der überprüften Beobachtungseinheiten eines Datenbestands liegt nach der Berechnung des QI „Beobachtungseinheiten mit Follow-up“ vor. Für die Aufnahme in den Zähler ist lediglich ein vorhandenes Follow-up erforderlich, ein geringer Wert des QI steht für schlechte Datenqualität. Gemäß TMF-Leitlinie sind anhand des Wertes des QI Aussagen zur Vollzähligkeit des Follow-ups möglich (Nonnemacher et al. 2014, S. 76). Das Vorhandensein der Follow-ups

sei wichtig für die Untersuchung von Krankheitsverläufen. Die Berechnung sollte für einen Datenbestand erfolgen. Der QI wurde gemäß TMF-Leitlinie aus dem Bereich der Krebsregister in die TMF-Leitlinie übernommen (Hentschel und Katalinic 2008; C. O. Schmidt et al. 2012a). Sowohl Mobilität der Patienten als auch die Bereitschaft der Meldestellen zur Meldung werden als Einflussfaktoren genannt.

Qualitätsindikatoren der Ebene Richtigkeit

Anhand der sechs QI der Ebene Richtigkeit lassen sich Aussagen zur Ergebnisqualität machen. Die QI dieser Ebene finden vor allem im Rahmen der Source Data Verification Anwendung.

TMF-1043 - Genauigkeit der Registerdaten

Für den Wert dieses QI wird der Anteil der im Register vorhandenen, korrekten Datenwerte an der Anzahl der überprüften Datenwerte bestimmt. Als „wahr“ sollten Datenwerte aus einer unabhängigen Datenquelle betrachtet werden, deren Datenwerte nicht Teil der Datenerhebung des Register sind (Nonnemacher et al. 2014, S. 77). Zu berücksichtigen sei zudem, dass ein Vergleich der Registerdaten mit den Ursprungsdaten der Quellen nicht Teil der Betrachtung dieses QI ist, sondern unter die QI „Übereinstimmung der Registerdaten mit den Originaldaten,, (TMF-1044, TMF-1045) fällt. Auf Basis einer der angegebenen Literaturquellen des QI wird Genauigkeit als das Ausmaß definiert, in dem die Registerdaten der Wahrheit entsprechen (Arts et al. 2002).

Die Berechnung des QI kann mit dem Fokus auf einzelne Datensätze, Beobachtungseinheiten oder den Datenbestand erfolgen. Betrachtet werden sollten grundsätzlich nur Datenelemente. Ein hoher Wert des QI ist ein Zeichen guter Datenqualität. Die TMF-Leitlinie verweist an dieser Stelle auf die Literatur für eine Einordnung des berechneten Wertes. So wird guter Datenqualität ein berechneter Wert von $> 95\%$ zugewiesen (Prins et al. 2000). Mäßig seien noch Werte zwischen 80% und 95% . Werte unter 80% seien hingegen schlecht. In der Implementierungshilfe der TMF-Leitlinie 1.0 war ein Schwellenwert von $> 95\%$ verzeichnet (Nonnemacher et al. 2007, S. 61).

Die Qualifikation und Verfügbarkeit des Erhebungspersonals und dessen verfügbare Zeit sind Einflussfaktoren des QI. Die Qualität der Datenquellen, die für den Abgleich genutzt werden, gehört ebenfalls zu den Einflussfaktoren.

Die TMF-Leitlinie nennt auf Basis der Literaturrecherche zwei weitere alternative Definitionen des QI. Zum einen sei Genauigkeit die Kombination der Indikatoren „Korrektheit“ und „Vollständigkeit“ (Brennan und Stead 2000; Hüasers 2004). Sie sei zudem definiert als Anteil der Personen, welche ein Merkmal faktisch aufweisen an der Gesamtheit der Personen im Register, denen dieses Merkmal dort zugewiesen ist (Hassey et al. 2001). Die TMF-Leitlinie verweist zudem auf weitere Quellen auf deren Basis der QI in die TMF-Leitlinie

übernommen wurde (Barrie und Marsh 1992; Jung und R. Winter 2000; Kühn-Stoffers 2004; Lindquist 2004; Massey und Hoffman 1989).

TMF-1044 - Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente

Der QI „Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente“ bildet einen Teil der SDV (Nonnemacher et al. 2014, S. 79). Im Zuge der Berechnung des QI werden die im Register gespeicherten Daten mit den Originaldaten im Hinblick auf Diskrepanzen verglichen. Der Wert des QI ist der Anteil der differierenden Datenwerte an der Gesamtheit der überprüften Datenwerte. Rahmen der Berechnung ist ein Datenbestand. Ein hoher Wert deutet daher auf schlechte Datenqualität hin, welche durch Probleme bei der Erfassung oder Erhebung der Registerdaten entstanden sein kann. Ein Schwellenwert von $> 5\%$ wurde in der Implementierungshilfe von Version 1.0 der TMF-Leitlinie festgelegt (Nonnemacher et al. 2007, S. 61). Die Qualifikation und Verfügbarkeit des Erhebungspersonals und dessen verfügbare Zeit sind Einflussfaktoren des QI. Zusätzlich kann der Wert des QI durch Zweiterfassungen beeinflusst werden.

Diskrepanzen zwischen Original- und Registerdaten werden als Datenfehler interpretiert. Die TMF-Leitlinie gibt zu bedenken, dass keine Aussagen über den Wahrheitsgehalt der Daten zu treffen seien, da nur die Übereinstimmungen zwischen Registerdaten und Originaldaten überprüft würden. Ebenso wird die Nähe dieses QI zu den QI „Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten“, „Konkordanz“ und „Genauigkeit der Registerdaten“ betont.

Die in der TMF-Leitlinie genannten Quellen sind vielfältig (Arts et al. 2002, 2001; Califf et al. 1997; Cnattingius et al. 1990; Gissler et al. 1996; Håkansson et al. 2001; Jensen et al. 2002; R. Khosla et al. 2000; Lin et al. 2004; Lu et al. 1995; Maruszewski et al. 2005; Mullooly 1990; Nielsen et al. 1996; Ose et al. 2004; O’Sullivan et al. 1996; Pogash et al. 2001; J. Schmidt et al. 1995; Teperi 1993; Vestberg et al. 1997; Weiss 1998; Weiss et al. 1993)

TMF-1045 - Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten

Während bei der Berechnung des vorherigen QI der Fokus auf Datenelementen lag, wird der Fokus im Rahmen der Berechnung dieses QI auf Beobachtungseinheiten erweitert. Der QI ist ebenso Teil der SDV. Diskrepanzen zwischen Originaldaten und Registerdaten werden somit als Datenfehler interpretiert. Ergebnis der Berechnung des QI ist der Anteil der Beobachtungseinheiten mit einem Fehler an der Gesamtheit der überprüften Beobachtungseinheiten eines Datenbestands. Ein hoher Wert des QI ist somit Zeichen schlechter Datenqualität. Hohe Werte sind laut TMF-Leitlinie ein Hinweis auf eine problembehaftete Erhebung oder Erfassung von Registerdaten (Nonnemacher et al. 2014, S. 82). Einfluss auf den QI nehmen Qualifikation und Verfügbarkeit des Eingabepersonals sowie dessen

zugestandene Zeit für die Datenerhebung. Zudem kann der Wert des QI durch Zweiterfassungen beeinflusst werden. Als Schwellenwert wurden $> 5\%$ in der Implementierungshilfe zur TMF-Leitlinie in Version 1.0 empfohlen (Nonnemacher et al. 2007, S. 61).

Die TMF-Leitlinie gibt zu bedenken, dass keine Aussagen über den Wahrheitsgehalt der Daten zu treffen seien, da nur die Übereinstimmungen zwischen Registerdaten und Originaldaten überprüft würden. Ebenso wird die Nähe dieses QI zu den QI TMF-1044 „Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente“, TMF-1002 „Konkordanz“ und TMF-1043 „Genauigkeit der Registerdaten“ betont.

Der QI entstammt gemäß TMF-Leitlinie aus denselben Quellen wie der vorherige QI (Arts et al. 2002, 2001; Califf et al. 1997; Cnattingius et al. 1990; Gissler et al. 1996; Håkansson et al. 2001; Jensen et al. 2002; R. Khosla et al. 2000; Lin et al. 2004; Lu et al. 1995; Maruszewski et al. 2005; Mullooly 1990; Nielsen et al. 1996; Ose et al. 2004; O’Sullivan et al. 1996; Pogash et al. 2001; J. Schmidt et al. 1995; Teperi 1993; Vestberg et al. 1997; Weiss 1998; Weiss et al. 1993).

TMF-1046 - Vollständigkeit der Registerdaten

Die TMF-Leitlinie definiert Vollständigkeit als das Ausmaß in dem Datenwerte, welche in das Register hätten eingehen können, tatsächlich eingegangen sind (Nonnemacher et al. 2014, S. 83). Die TMF-Leitlinie verweist an dieser Stelle auf die Literatur als Quelle dieser Definition (Arts et al. 2002). Der Wert des QI entspricht demgemäß dem Anteil der im Register vorhandenen Datenwerte an der Summe aus vorhandenen und fehlenden Datenwerten; die TMF-Leitlinie verwendet diese Berechnungsformel auf Basis einer der Quellen des QI (Naumann und Rolker 2000). Die Berechnung des QI ist nur für Datenelemente durchzuführen. Als Betrachtungsrahmen für die Berechnung können einzelne Datensätze, Beobachtungseinheiten oder der Datenbestand gewählt werden. Ein hoher berechneter Wert des QI steht für gute Datenqualität. Ein Schwellenwert von $< 95\%$ wurde in der Implementierungshilfe von Version 1.0 der TMF-Leitlinie empfohlen (Nonnemacher et al. 2007, S. 61). Die TMF-Leitlinie beschreibt zudem die Möglichkeit der Berechnung des QI im Hinblick auf die Fälle, die in das Register hätten eingehen sollen, über der Anzahl der Fälle, die tatsächlich im Register vorhanden sind. Die Verfügbarkeit und Qualifikation des Erhebungspersonals sowie deren verfügbare Zeit werden als Einflussfaktoren genannt. Zusätzlich unterliegt der Wert des QI der Vollständigkeit der Originaldokumente, auf deren Basis die Datenwerte in das Register eingegangen sind.

Die TMF-Leitlinie betont die Beziehung des QI zu anderen QI der Leitlinie, namentlich zu TMF-1002 „Konkordanz“ und TMF-1043 „Genauigkeit der Registerdaten“. Um die Vollständigkeit korrekt schätzen zu können, sei es notwendig, Datenquellen zu identifizieren um feststellen zu können, ob deren Datenwerte in das Register hätten eingehen können. Als Beispiele hierfür werden andere Register, Todesbescheinigungen der Patienten oder die

Abrechnungsdaten von Krankenkassen genannt. Der QI sei zudem wichtig zur korrekten Schätzung von Inzidenzen und Prävalenzen.

Der QI wird gemäß TMF-Leitlinie vielfach in der Literatur genannt (Arts et al. 2002; Barrie und Marsh 1992; Bobrowski et al. 1999; Goldberg et al. 1980; Hassey et al. 2001; Hogan und Wagner 1997; Jensen et al. 2002; Jung und R. Winter 2000; Katalinic 2005; Laporte, Mazumdar et al. 1994; Lindquist 2004; Logan et al. 2001; Naumann und Rolker 2000; Nielsen et al. 1996; D. Parkin et al. 1992; Teppo et al. 1994; Topp et al. 1997; Vestberg et al. 1997).

TMF-1047 - Übereinstimmung mit Verfahrensregeln

Auf Basis des Studienplans eines Registers oder einer Kohortenstudie können Abweichungen in der Behandlung oder dem Follow-up von Patienten festgestellt werden. Der Wert dieses QI entspricht der Anzahl der Abweichungen vom Studienplan an der Gesamtzahl der überprüften Patienten. Die TMF-Leitlinie hebt an dieser Stelle die Übereinstimmung mit dem Studienplan als Basis für die Nutzbarkeit der Registerdaten hervor (Nonnemacher et al. 2014, S. 85).

Für die Berechnung des QI ist es zunächst erforderlich, die relevanten Datenelemente des zu betrachtenden Datenbestands und Verfahren zur Messung von Abweichungen dieser Datenelemente vom Studienplan zu identifizieren. Anhand dieser Rahmenbedingungen werden alle Abweichungen vom Studienplan erfasst und in die Berechnung eingebracht. Die TMF-Leitlinie gibt als alternative Berechnungsmöglichkeit zudem noch die mittlere Anzahl von Abweichungen pro Patient an.

Hohe Werte des QI sind Zeichen schlechter Datenqualität. Ein Schwellenwert von $> 5\%$ wurde in der Implementierungshilfe von Version 1.0 der TMF-Leitlinie empfohlen (Nonnemacher et al. 2007, S. 61). Der Wert unterliegt gemäß der TMF-Leitlinie vielen Einflussfaktoren. Neben der Motivation von Eingabekräften und Patienten spielt auch die Klarheit des Studienplans sowie dessen Form und Inhalt eine Rolle. Des Weiteren werden die Qualität der technischen Infrastruktur als auch die Strukturiertheit der Datenerhebung genannt.

Die TMF-Leitlinie nennt mehrere Quellen auf deren Basis der QI in die TMF-Leitlinie übernommen wurde (Gaus 2003; Hoffmann et al. 2004; Whitney et al. 1998).

TMF-1048 - Repräsentativität der Registerdaten

Die Verallgemeinerbarkeit von Auswertungsergebnissen ist abhängig von der Repräsentativität der Registerdaten (Nonnemacher et al. 2014, S. 87). Die Repräsentativität ist dabei gemäß TMF-Leitlinie nur relevant, wenn nicht die gesamte Population erfasst werden soll, sondern nur eine Subpopulation, die in ihren wesentlichen Einflussgrößen der Gesamtpopulation entspricht. Für die Berechnung selbst ist es erforderlich, wesentliche Einflussgrößen der Repräsentativität festzulegen sowie diese für die Registerpopulation zu bestimmen.

Die aus dem Datenbestand bestimmten Einflussgrößen sind anschließend mit den Einflussgrößen der Gesamtpopulation zu vergleichen, die sich beispielsweise aus Inzidenzen und Prävalenzen festlegen lassen.

Der Wert des QI ergibt sich aus dem Anteil der Einflussgrößen mit erwarteter Verteilung an der Gesamtzahl der überprüften Einflussgrößen. Die Repräsentativität ist daher mit hohen Werten des QI als besser einzuschätzen. Einfluss auf den Wert des QI nehmen die Art der Einrichtungen sowie ihre räumliche Verteilung, die rekrutierten Patientenkollektive der jeweiligen Einrichtungen spielen zudem eine Rolle. Die Empfehlung in Version 1.0 der TMF-Leitlinie ist ein Schwellenwert von $< 100\%$ (Nonnemacher et al. 2007, S, 61).

Die TMF-Leitlinie nennt keine Literaturquellen sondern beruft sich auf Projektpartner der TMF hinsichtlich der Herkunft des QI.

2.7.3 Qualitätsscore

Die TMF-Leitlinie definiert ein beispielhaftes Vorgehen zur Ermittlung eines Qualitätsscores anhand der vorgestellten QI (Nonnemacher et al. 2014, S. 89). Die Anwendung eines Scores geht dabei gemäß TMF-Leitlinie auf zwei Literaturquellen zurück (Holle 1995; Tolonen et al. 2006). Die Komponenten des Scores ergeben sich aus der Auswahl der anzuwendenden QI sowie deren Gewichtung. Auf eine definitive Gewichtung der QI seitens der TMF-Leitlinie wurde verzichtet, da weder ein adäquater Datenbestand, noch ein „externes Kriterium zur Datenqualität“ zur Verfügung stand (Nonnemacher et al. 2014, S. 90). Die Gewichte seien daher normativ-analytisch festgelegt worden. Die TMF-Leitlinie weist der Ebene Richtigkeit im Hinblick auf die SDV, gefolgt von den Ebenen Organisation und Integrität, die größte Relevanz zu. Den QI werden im Zuge der Berechnung beispielhafte Gewichte von 1-9 zugewiesen. Die Ebene Richtigkeit beinhaltet Gewichte von 7 bis 9, Organisation von 4 bis 6 und letztlich die Ebene Integrität von 1 bis 3. Die beispielhafte Zuweisung der Gewichte zu den QI, innerhalb der einzelnen Ebenen, wurden auf Basis der Meinung von Experten vorgenommen.

Die Ermittlung des Qualitätsscore erfolgt zunächst durch den Vergleich der ermittelten Werte der QI mit dem jeweiligen Schwellwert des QI. Bei Auffälligkeit geht der QI mit Wert 1 in die Berechnung ein, ansonsten mit Wert 0. Die Werte der QI werden anschließend mittels einer Transformation zur Standardisierung auf eine Skala von 0 bis 100 umgelegt. Für diese Transformation werden die Werte der QI mit dem jeweiligen Gewicht des QI multipliziert. Diese berechneten Werte werden über alle QI aufsummiert und durch die Summe aller Gewichte dividiert. Zuletzt wird das Ergebnis der Division mit 100 multipliziert und bildet somit den Gesamt-Scorewert auf einer Skala von 0 bis 100.

Auf Basis dieses transformierten Wertes wird die Datenqualität in fünf Wertegruppen der Größe 20 auf der genannten Skala beurteilt, wobei 0 sehr schlechter und 100 sehr guter Datenqualität entspricht. Für die Beurteilung des Scores sei zudem der Anteil der QI der

jeweiligen Ebenen zu berücksichtigen, da je nach Auswahl der QI bestimmte Ebenen über- oder unterrepräsentiert sein können. Im Zuge dessen sei eine Anpassung der Gewichte vorzunehmen. Dabei sollten nicht alle QI verwendet werden.

2.7.4 Registerempfehlungen

Die TMF-Leitlinie bietet zudem ein Kapitel zu Anwendungsempfehlungen für Register (Nonnemacher et al. 2014, S. 110). Anhand des Nationalen Registers für angeborene Herzfehler wird beispielhaft dargestellt, wie in diesem Register ausgewählte QI zum Einsatz kommen. Jeder dieser 20 QI wird mit einem anwendungsbezogenen Kommentar versehen und allgemeine Eigenschaften und Abläufe des Registers werden erläutert. Ebenso werden Informationen zum Datensatz und dessen Design gegeben. Hervorgehoben werden die Aspekte der Machbarkeit bei der Sicherstellung der Datenqualität und die Nutzung der TMF-Leitlinie während der Planung von Registern um adäquate Datenqualität und adäquates Datenmanagement derselben sicherzustellen.

2.8 Verwandte Arbeiten

Sowohl als Basis des Ansatzes der TMF-Leitlinie als auch des Datenqualitätsmanagementprozesses der CERTAIN-Register können die Empfehlungen zur Datenqualität von Arts et al. angesehen werden, siehe Abbildung 8 (Arts et al. 2002).

Während die TMF-Leitlinie jedoch den Fokus auf Maßnahmen zur Verbesserung der künftigen Datenqualität legt, stehen beim Ansatz der CERTAIN-Registerplattform die Prüfungen während der Dateneingabe im Vordergrund. Dies liegt zum einen an den begrenzten Ressourcen des Registers, zum anderen folgt dieser Ansatz den Empfehlungen von Gassman et al., Rückfragen an die Einrichtung möglichst zeitnah zu stellen, da es den Eingabekräften so einfacher möglich wird, angeforderte Informationen aufzufinden (Gassman et al. 1995). Der Ansatz der automatisierten Generierung von Rückfragen hat sich gemäß Gassman in zwei Publikationen bewährt (Amoroso et al. 1989; Newhouse 1985). Gassman et al. empfehlen sowohl proaktive Mechanismen zur Sicherung der Datenqualität als auch regelmäßige Berichte zur Datenqualität an die teilnehmenden Einrichtungen.

Beretta et al. führten eine Analyse von 499 Patienten ihrer Datenbank für Kopfverletzungen durch und implementierten ihre Datenqualitätssicherungsmaßnahmen als Teil der Datenbank (Beretta et al. 2007). Die Autoren heben die Verhinderung von falschen Eingaben während der Datenerhebung hervor und geben eine Fehlerrate von 0,44% an, mit dem Hinweis, dass aufgrund der niedrigen Fehlerraten keine weiteren Versuche zur programmatischen Verbesserung der Datenqualität auf Basis des als gering eingeschätzten Kosten-Nutzen-Verhältnis angestrebt wurden. Die Autoren sehen es weiterhin als essentiell

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 compose minimum set of necessary data items define data & data characteristics in data dictionary draft a data collection protocol define pitfalls in data collection compose data checks create user friendly case record forms create quality assurance plan In case of new participating sites perform site visit train new participants Continuously motivate participants communicate with local sites In case of changes (e.g., in data set) adjust forms, software, data dictionary, protocol, training material, etc. communicate with local sites	At the onset of participating in the registry assign a contact person check developed software for data entry and for extraction check reliability and completeness of extraction sources standardise correction of data items Continuously train (new) data collectors motivate data collectors make data definitions available place date & initials on completed forms keep completed case record forms data collection close to the source and as soon as possible use the registry data for local purposes In case of changes (e.g., in data set) adjust data dictionary, forms, software, etc. communicate with data collectors
Detection during data collection	
During import of data into the central database perform automatic data checks Periodically and in case of new participants perform site visits for data quality audit (registry data <> source data) and review local data collection procedures Periodically check inter- and intraobserver variability perform analyses on the data	Continuously visually inspect completed forms perform automatic data checks check completeness of registration
Actions for quality improvement	
After data import and data checks provide local sites with data quality reports control local correction of data errors After data audit or variability test give feedback of results and recommendations resolve causes of data errors	After receiving quality reports check detected errors correct inaccurate data & fill in incomplete data resolve causes of data errors After receiving feedback implement recommended changes communicate with personnel

Abbildung 8: Prozeduren zur Sicherung der Datenqualität in medizinischen Registern (Arts et al. 2002)

an, das Vorhandensein kritischer Daten, beispielsweise des Outcomes, über eine generell gute Vollständigkeit zu stellen.

Hlaing et al. etablierten eine Datenvalidierung für ein Traumaregister auf Basis von Beschränkungen in der Datenbank des Registers, mit dem Ziel den Eingang fehlerhafter Daten in die Datenbank zu verhindern (Hlaing et al. 2006). Die Autoren geben zudem einen Überblick über die eingesetzten Validierungsprüfungen, festgestellte Fehlerraten der untersuchten Daten und schließen mit der Forderung, Datenvalidierungsmechanismen zu veröffentlichen und zu standardisieren, um die Validierung aller Traumaregister zu ermöglichen.

Bautista et al. betonen in ihrer Beschreibung eines Modells zur Validierung von Daten in Arthroplastik-Registern die Wichtigkeit strikter Eingabeformulare für gute Datenqualität,

da Datenerhebung keine Stärke der meisten Ärzte sei (Bautista et al. 2017). Zudem seien periodische Audits der Datenqualität notwendig, um systematische oder zufällige Fehler der elektronischen oder manuellen Datenerfassung zeitnah erkennen zu können.

In einem systematischen Literaturreview zur Datenqualität von Traumaregistern von O'Reilly et al. benennen die Autoren die Entwicklung von effizienten Feedbackmechanismen als besten Ansatz zur Verbesserung der Datenerhebung (O'Reilly et al. 2016). Die Autoren stellen zudem fest, dass es keine einfache Checkliste gäbe, um ein Traumaregister mit hoher Datenqualität zu etablieren.

Im Zuge einer Untersuchung der Datenqualität einer Datenbank mit Fokus auf Operationen am Herzen sollte festgestellt werden, ob Genauig- und Vollständigkeit der Datenbank durch Monitorierung, Validierung und Feedback an die teilnehmenden Einrichtungen verbessert werden könnte (Fine et al. 2003). Durch eine Kombination aus Validierung, Monitoring und Feedback an die Einrichtungen über einen Zeitraum von fünf Monaten konnte die Datenqualität verbessert werden, blieb aber dennoch hinter den Erwartungen zurück. Die Autoren ziehen den Schluss, dass ein kontinuierlicher Zyklus von externer Monitorierung und Validierung in Verbindung mit der Validierung aller Datenbanken sinnvoll sei. Für Auswertungen zu medizinischen Fragen genutzte Daten seien vor der Veröffentlichung nochmals gesondert zu prüfen.

Das Programm zur Verbesserung der Datenqualität des National Cardiovascular Data Registry (NCDR) basierte ursprünglich rein auf Prüfungen der Vollständigkeit der Daten, entwickelte sich jedoch mit der über die Zeit gestiegenen Bedeutung des Registers zu einem Programm mit drei Komponenten (Messenger et al. 2012). Zum einen ist es ein Bericht zur Datenqualität, der die Vollständigkeit der Datensätze zwischen den Einrichtungen vergleicht. Zum anderen sind es interne Protokolle zur Sicherung der Qualität, welche jedoch nur bei einem Export der Daten zu Forschungszwecken durchgeführt werden. Das jährliche Audit prüft die Übereinstimmung der Daten mit den Originaldokumenten. Die Autoren sehen durch die gestiegene Nutzung der Registerdaten einen gestiegenen Bedarf für die Validierung der Daten und planen den Umfang der Vollständigkeitskontrollen des Berichts zu steigern.

3 Methoden und Werkzeuge

Durch die Auswertungen des CERTAIN-Datenbestandes für wissenschaftliche Fragestellungen wurde deutlich, dass immer wieder offensichtlich falsche Daten in das Register eingepflegt wurden. Diese Falschangaben bedingten aufwändige Rückfragen durch den Datenqualitätsmanager an die Zentren, aus denen die Daten stammten. Durch die erforderlich gewordenen Rückfragen wurde zudem ersichtlich, dass die automatische Validierung des Datenbestandes bisher nicht den notwendigen Umfang zur ganzheitlichen Qualitätssicherung des Datenbestandes erreicht hatte. Für die Registerzentrale des CERTAIN-Registers war es deshalb von Bedeutung, die Qualität der Daten weiter zu verbessern. Aufgrund der Entstehung der beiden neuen Register, CERTAIN-LI und KRhOKo, sollten alle künftigen Maßnahmen zur Verbesserung der Datenqualität auch auf diese beiden Register anwendbar sein. Da sich alle drei Register dieselbe Codebasis und Architektur teilen, sollten hierdurch Mehrfachimplementierungen vermieden sowie bestehende Komponenten genutzt werden.

3.1 Entwicklung des Ansatzes

Um geeignete Werkzeuge zur Verbesserung der Datenqualität der drei Register zu identifizieren, wurde, nach Sichtung der Literatur und in Absprache mit Fachexperten, die TMF-Leitlinie zur Datenqualität der Technologie- und Methodenplattform für die vernetzte medizinische Forschung als Startpunkt gewählt. Ein wichtiger Grund war das umfassende Literaturreview, welches die Basis der TMF-Leitlinie bildet.

Die TMF-Leitlinie scheint vor allem deshalb besonders geeignet, weil sich der Ansatz der Leitlinie, das adaptive Feedback mit Ziel der Lenkung in Richtung einer Verbesserung der künftigen Dateneingabe und der CERTAIN-Ansatz, dreistufiger Qualitätssicherungsprozess mit rigider, automatischer Validierung mit dem Ziel der Verhinderung von Falscheingaben während der Dateneingabe, ergänzen. Die TMF-Leitlinie favorisiert das adaptive Feedback gegenüber dem festgelegten Goldstandard der SDV, aufgrund der besseren Umsetzbarkeit und Praktikabilität des adaptiven Feedbacks. Da sich automatische Validierung und adaptives Feedback ergänzen, aber auch ohne einander existieren können, sollte zunächst festgestellt werden, ob der Ansatz der automatischen Validierung auch für das KRhOKo-Register problemlos anwendbar wäre, oder ob die existierenden Maßnahmen der CERTAIN-Validierung durch eine Änderung des Datensatzes an ihre Grenzen stoßen.

3.2 Analyse der automatischen Validierung des CERTAIN-Ansatzes

Die bisher papierbasierte Erhebung der Daten für das Forschungsprojekt „Kerndokumentation“ des Deutschen Rheuma-Forschungszentrums sollte, im Rahmen der parallel zu die-

ser Dissertation verlaufenden Umsetzung des Projektes KRhOKo („Kinderrheumatologie online - Kerndokumentation“), auf eine webbasierte Erhebung umgestellt werden. Im Zuge dieser Umstellung wurden, analog zu den Validierungsregeln in CERTAIN, KRhOKo-Validierungsregeln in das Datenmodell der Onlineerhebung implementiert, die bisher bei einer papierbasierten Erhebung nicht unmittelbar geprüft werden konnten. Da der Datenbestand der papierbasierten Erhebung im Auftrag des DRFZ durch eine Firma in manueller Arbeit durch Übertragung der Datenwerte in ein ASCII-Format digitalisiert wird, war es möglich, die Datenqualität aller im Jahre 2014 papierbasiert erfassten Daten anhand der KRhOKo-Validierungsregeln zu prüfen, um so die Effektivität der Validierungsregeln abschätzen zu können.

Für diese Analyse wurde zunächst gemäß den Empfehlungen der TMF, auf Basis der während der Source Data Verification benötigten Zahl an zu untersuchenden Beobachtungseinheiten, die benötigte Anzahl an Fragebogen berechnet. Zur Berechnung der unadjustierten Fallzahl n_0 wurde auf die Berechnungsformel der TMF-Leitlinie zurückgegriffen:

$$n_0 = \frac{p(1-p)}{\delta^2} \cdot z_{1-\alpha/2}^2$$

In die Berechnung gehen der Anteil p von Beobachtungseinheiten mit mindestens einem fehlerhaften Eintrag in den Daten, die absolute halbe Breite δ des Konfidenzintervalls und das Quantil der Standard-Normalverteilung $z_{1-\alpha/2}$ ein. Die Fallzahl wird adjustiert, da n_0 größer sein kann als die Gesamtfallzahl einer Einrichtung und entsprechend eine Reduktion der Fallzahl für die Anzahl der verfügbaren Beobachtungseinheiten der Einrichtung notwendig ist. Aufgrund von Aussagen des Registerpersonals des DRFZ zur beobachteten Datenqualität wurde die Datenqualität als sehr gut angenommen. Daher wurde auf Basis der Empfehlung der TMF-Leitlinie für die absolute halbe Breite δ des Konfidenzintervalls der Wert $\delta = 0,05$ gewählt. Für den Anteil p von Beobachtungseinheiten mit mindestens einem fehlerhaften Eintrag in den Daten wurde die Empfehlung der TMF-Leitlinie für den ersten Zyklus der Leitlinienanwendung mit $p = 0,05$ übernommen. Das Quantil der Standard-Normalverteilung $z_{1-\alpha/2}$ wurde auf Basis der Leitlinie, mit $\alpha = 0,05$ für den vorgegebenen Fehler 1. Art, mit 1,96 gesetzt. Für die unadjustierte Fallzahl wurde daher $n_0 = 72,9904$ berechnet. Zur Berechnung der adjustierten Fallzahl n wurde auf die Berechnungsformel der TMF-Leitlinie zurückgegriffen:

$$n = \frac{n_0 \cdot N}{n_0 + N}$$

Mit $N = 11706$ für die Gesamtzahl der Beobachtungseinheiten, ergab sich eine adjustierte Fallzahl von $n = 72,5381033$. Daher wurde beschlossen 80 Arzt-, 80 Eltern- und 80 Jugendlichenbogen für die Analyse auszuwerten. Diese Fragebogen wurden durch das DRFZ per Zufall aus den vier teilnahmestärksten Einrichtungen ausgewählt, welche an der Kerndokumentation des DRFZ in der bisher papierbasierten Form teilnahmen. Durch

einen Fehler beim Abzählen der Fragebogen wurde ein Elternbogen zu viel nach Heidelberg übersandt, so dass letztlich insgesamt 241 Fragebogen der papierbasierten Kerndokumentation analysiert wurden. Die Papierversionen der, zufällig aus allen eingesandten Fragebogen der Einrichtungen ausgewählten, 80 Arzt-, 81 Eltern- und 80 Jugendlichenbogen wurden zunächst auf Diskrepanzen zwischen Papierfragebogen und digitalisierter Version untersucht. Hierfür wurden alle Datenelemente jedes einzelnen Fragebogens mit dem korrespondierenden Datenelement der digitalen Version abgeglichen und alle Unterschiede erfasst. Die aufgetretenen Fehler wurden in vier Kategorien unterteilt. Zunächst Fehler, bei denen Datenwerte während der Digitalisierung geändert wurden, beispielsweise im Papierbogen 0 („problemlos“), in der digitalisierten Version jedoch 1 („leicht erschwert“). Zudem wurden Fehler durch fehlende Datenwerte festgestellt, bei denen der Datenwert des Papierbogens in der digitalisierten Version nicht mehr vorhanden war. Ebenso gab es Fehler, bei denen im Papierbogen keine Angabe vorhanden war, die digitalisierte Version für dieses Datenelement jedoch einen Datenwert aufwies. Alle übrigen, festgestellten Fehler wurden in der Kategorie „Sonstige“ zusammengefasst. Um die digitalisierten Fragebogen gegen die Validierungsregeln der Onlineerhebung prüfen zu können, wurde eine Importfunktion entwickelt, welche die Daten der digitalisierten Fragebogen auf das Datenmodell der Onlineerhebung überträgt. Nach Überführung der digitalisierten Fragebogen in das Datenmodell der Onlineerhebung wurden die Fragebogen gegen die neuen Validierungsregeln des KRhOKo-Registers geprüft, um festzustellen, inwieweit die neuen Validierungsregeln Fehler durch den Digitalisierungsprozess sowie durch falsche Angaben während der Datenerhebung erkannt hätten. Um etwaige Unterschiede beim Ausfüllen der Fragebogen zwischen den Benutzergruppen Eltern und Jugendliche hinsichtlich der Vollständigkeit der Fragebogen feststellen zu können, wurden in der Programmiersprache Scala spezielle Methoden entwickelt. Anhand dieser Methoden konnte die Vollständigkeit von Eltern- und Jugendlichenbogen auf Basis des aktuellen KRhOKo-Datenmodells berechnet werden. Für Datenelemente, die zwar im Papierbogen, aber nicht mehr in der neuen Onlineerhebung vorhanden waren, wurde die Vollständigkeit manuell durch Kontrolle jedes einzelnen Fragebogens erfasst. Die berechnete Vollständigkeit auf Basis des KRhOKO-Datenmodells und die manuell festgestellte Vollständigkeit der restlichen Datenelemente des Papierbogens wurden für jeden Fragebogen zu einem Gesamtwert der Vollständigkeit zusammengefasst. Anschließend wurde die Vollständigkeit der beiden Stichproben, 81 Eltern- und 80 Jugendlichenbogen, durch einen Wilcoxon-Mann-Whitney-Test verglichen. Hierdurch sollte festgestellt werden, ob zwischen den Benutzergruppen Unterschiede bestehen, so dass für die Benutzergruppen spezielle Hilfen oder motivierende Informationen während der Dateneingabe angeboten werden könnten.

3.3 Entwicklung des allgemeinen Vorgehens

Um die TMF-Leitlinie zur Verbesserung der Datenqualität der drei Register zu nutzen, war zunächst die Entwicklung eines allgemeinen Vorgehens zur Anwendung der QI der TMF-Leitlinie im Rahmen einer automatischen Validierung notwendig. Hierfür wurden sowohl die TMF-Leitlinie als auch die in der TMF-Leitlinie referenzierte Literatur auf Ansatzpunkte zur Formalisierung und Vereinfachung der Anwendung der QI hin untersucht. Nach der Identifizierung der relevanten Ansatzpunkte wurde ein Vorgehen zur Auswahl der für das jeweilige Register relevanten QI entwickelt. Dieses Vorgehen stützt sich auf eine Einschätzung der Anwendbar- und Implementierbarkeit sowie Aussagekraft des jeweiligen QI auf einer Likert-Skala. Auf diesem Vorgehen setzte die weitere Entwicklung eines Verfahrens zur Festlegung von Prioritäten für die QI auf, um ein schrittweises Vorgehen bei der Anwendung zu ermöglichen. Sowohl die identifizierten Ansatzpunkte als auch das Verfahren zur Festlegung der Prioritäten bildeten die Basis der Entwicklung eines Vorgehens zur Ermittlung der Gewichte der QI im Rahmen der Berechnung des Gesamtscores der Datenqualität.

Auf der Basis des entwickelten, allgemeinen Vorgehens wurden die QI der TMF-Leitlinie im Hinblick auf ihre Anwendbar- und Implementierbarkeit sowie ihre Aussagekraft für die drei CERTAIN-Registern systematisch analysiert. Die Beurteilung der QI stütze sich auf Erfahrungen, die während der Entwicklung der drei Register der CERTAIN-Registerplattform und während medizinischer Auswertungen der Daten dieser Register gemacht worden waren. Teil der Beurteilung war gleichermaßen die Analyse der QI im Hinblick auf die Datensätze der drei Register und die Architektur der Register-Plattform. Diese Einschätzungen wurden in Gesprächen mit dem Personal der Registerzentralen überprüft. Die Beurteilung der QI wurde zur Priorisierung der QI für die drei CERTAIN-Register genutzt und das Ergebnis der Priorisierung wiederum zur Festlegung der Gewichte der QI. Die zu den QI zugewiesenen Prioritäten und Gewichte wurden wiederum durch Gespräche mit den Registerzentralen bestätigt. Abbildung 9 fasst diese Vorgänge zusammen.

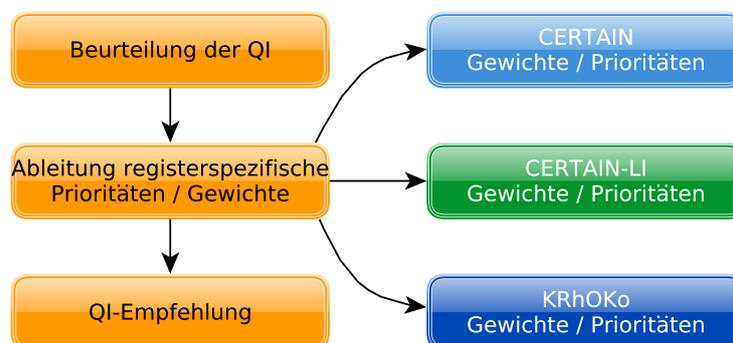


Abbildung 9: Entwicklung der allgemeinen und registerspezifischen QI-Empfehlungen

Zudem wurde auf Basis dieser Einschätzungen eine QI-Empfehlung für elektronische Register entwickelt, welche die Anwendung der TMF-Leitlinie für elektronische Register bei Nutzung des entwickelten Vorgehens vereinfachen sollte.

3.4 Implementierung von Modulen zur Qualitätsprüfung

Durch die vorhergehende systematische Analyse der QI konnten die QI identifiziert werden, die durch Software-Komponenten zur Überprüfung von automatischen Validierungsregeln abgebildet werden können. Diese QI wurden in generischen Validierungsregeln implementiert, welche diese QI entweder vollständig oder in einem Teilaspekt abdecken. Diese neuen Validierungsregeln ergaben sich entweder direkt aus den QI der TMF-Leitlinie oder wurden durch Gespräche mit dem Personal der jeweiligen Register identifiziert und sollten in allen drei CERTAIN-Registern anwendbar sein. Als höchste Priorität wurde hierbei die Erkennung von Falschangaben oder möglichen Fehlern für die Labor- und Medikationsangaben der beiden Transplantationsregister identifiziert. Um diese Validierungsregeln umsetzen zu können, wurde durch das medizinische Personal der CERTAIN-Registerzentrale ein Datenvalidierungsplan erstellt, der Wertebereiche in Form von harten und weichen Grenzen für künftig zu validierende Datenwerte enthielt. Auf Basis dieses Datenvalidierungsplans und den QI, welche als umsetzbar in automatischen Validierungsregeln identifiziert wurden, wurden Software-Komponenten in Form von Java-Modulen für diese neuen Validierungsregeln implementiert. Diese neuen Software-Komponenten sollten die bisherigen Validierungsregeln der drei Register komplementieren oder ersetzen. Die Software-Komponenten orientieren sich an und nutzen JSR-349 und die JSR-349 Implementierung, welche in Hibernate in Form des Hibernate Validators umgesetzt wurde. Als Entwicklungsumgebung für alle Implementierungsarbeiten wurde IntelliJ genutzt. Die initiale Verifizierung der Funktionalität aller neuen Software-Komponenten erfolgte anhand von JUnit-Tests unter Zuhilfenahme der Test-Bibliothek Mockito.

3.5 Auswertung der Effektivität der neuen Validierungsmodule

Nach Implementierung und Verifizierung der neuen Software-Komponenten wurden die neuen Validierungsregeln der Software-Komponenten durch Validierung des existierenden Gesamtdatenbestand des Nierentransplantationsregister CERTAIN auf ihre Richtigkeit und Funktionalität hin überprüft. Auftretende Softwarefehler der Software-Komponenten aufgrund von bisher nicht getesteten Datenkonstellationen wurden korrigiert und zusätzliche Tests implementiert. Beobachtete Auffälligkeiten in den Ergebnissen der neuen Validierungsregeln wurden betrachtet und nach Rücksprache mit dem medizinischen Personal bereinigt.

Zudem wurden die Wertebereiche des Validierungsplans von Medikation und Labor eingehend durch speziell programmierte Analysemethoden untersucht, welche den Gesamtdatenbestand des CERTAIN-Registers einbezogen. Diese Analysemethoden lieferten Informationen über Datenwerte, welche die harten Validierungsgrenzen verletzten. Zusätzlich analysierten sie die Datenwerte eines Datenelements aus allen Visiten des Datenbestandes, um die Werteverteilung eines jeden Datenelements betrachten und durch zusätzliche Skripte der Programmiersprache R visualisieren zu können. Auch lieferten diese Analysemethoden Informationen über inkompatible Einheiten, die nicht durch die neuen Software-Komponenten konvertiert werden konnten. Auffälligkeiten dieser Analysen wurden durch Korrekturen der Software-Komponenten, der neuen Validierungsregeln oder des Datenbestandes adressiert. Anschließend wurden die Software-Komponenten und die auf der Basis der Analysen optimierten Validierungsregeln in den Produktivbetrieb übernommen und ab diesem Zeitpunkt für die automatische Validierung der Datensätze des CERTAIN-Registers verwendet.

Zur Überprüfung des Einflusses der optimierten Validierungsregeln auf die Datenqualität wurden spezielle Analysemethoden in der Programmiersprache Scala entwickelt, die jede Visite eines Datenbestands einzeln mittels des Hibernate *Validators* gegen die Gesamtheit der Validierungsregeln des CERTAIN-Registers prüfen. Die pro Visite festgestellten Fehler und Warnungen werden anschließend zu einem Gesamtbild des Status der Datenqualität des jeweiligen Datenbestandes aggregiert.

Die Auswirkungen der optimierten Validierungsregeln wurden mittels dieser Analysemethoden anhand von mehreren Datenbeständen untersucht. Diese Datenbestände wurden zu unterschiedlichen Zeitpunkten aus dem CERTAIN-Register entnommen, siehe Abbildung 10.

Zunächst wurde am 17.11.2016 der Datenbestand D1 aus dem Produktivbetrieb der Version R1 des CERTAIN-Registers entnommen. Dieser Datenbestand unterlag nur den initialen Validierungsregeln V1, welche vor dieser Dissertation existierten. Dieser Datenbestand wurde gegen die zu diesem Zeitpunkt aktiven, initialen Validierungsregeln V1 geprüft. Anschließend wurden mit der Aktualisierung des Registers auf Version R2 die optimierten Validierungsregeln V2 aktiviert. Nach der Aktualisierung der Register-Version und der Durchführung von Einheitenkorrekturen auf der Datenbank des Registers wurde der Datenbestand D2 aus der Datenbank entnommen. Zur Feststellung der initialen Datenqualität wurde der Datenbestand D1 gegen die Validierungsregeln V1 geprüft. Parallel fand eine Prüfung des Datenbestandes D2 gegen die optimierten Validierungsregeln V2 statt. Diese beiden Validierungsergebnisse wurden anschließend zur Begutachtung der Datenqualität und des Einflusses der optimierten Validierungsregeln V2 auf die Datenqualität miteinander verglichen.

Der Einfluss der optimierten Validierungsregeln V2 sollte anschließend ein Jahr lang beobachtet werden. Im Laufe dieses Jahres wurden kleinere Anpassungen an den Validie-

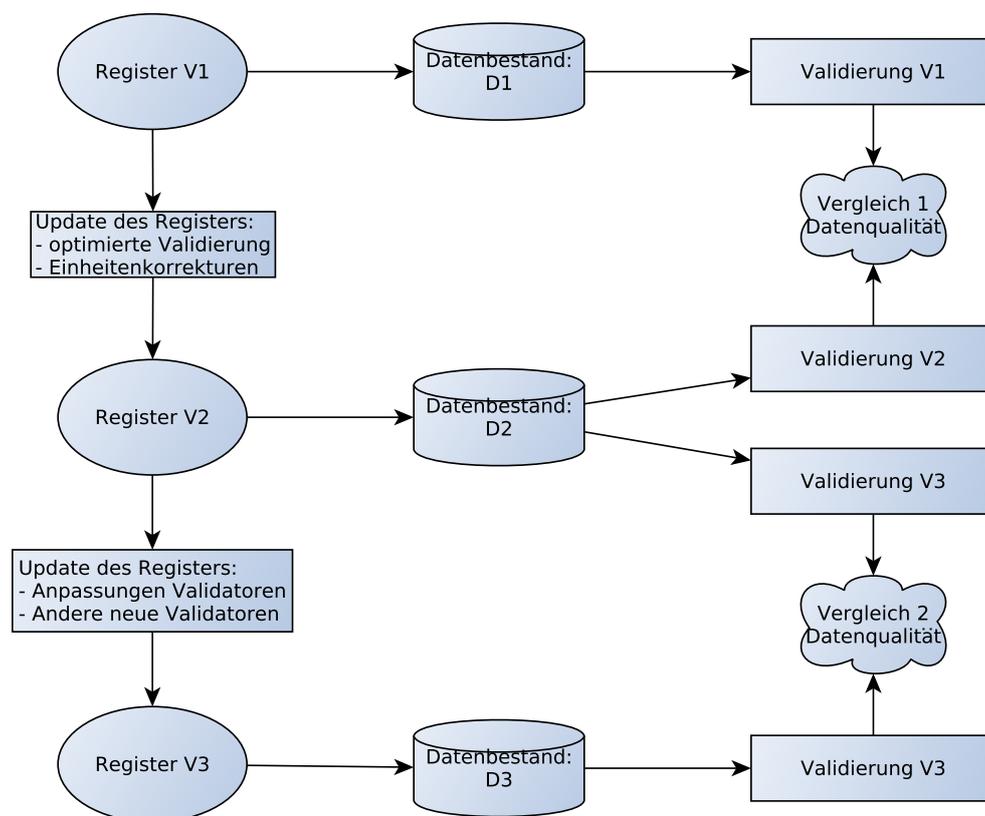


Abbildung 10: Auswertung der Effektivität der Validierung

Validierungsregeln V2 auf Basis der Rückmeldung der Eingabekräfte vorgenommen. Es wurden zusätzlich zwei weitere Software-Komponenten zur Überprüfung von Validierungsregeln durch den Programmierer des CERTAIN-Registers etabliert. Diese Anpassungen bilden die Validierungsregeln V3. Die Register-Software wurde das Jahr hinüber weiterentwickelt und erhielt neue Funktionalitäten sowie weitere Studienmodule. Diese Register-Version im November 2017 wird im Folgenden als Version R3 bezeichnet.

Zur Überprüfung der Datenqualität nach einem vergangenen Jahr wurde der Datenbestand D3 am 17.11.2017 aus dem CERTAIN-Register Version R3 entnommen. Der Datenbestand D3 wurde gegen die am 17.11.2017 aktiven Validierungsregeln V3 geprüft. Der Datenbestand D2 wurde ebenfalls gegen die Validierungsregeln V3 geprüft, um beide Datenbestände gegen die aktuellste Version der Validierungsregeln zu prüfen und Vergleichbarkeit der Validierungsergebnisse herstellen zu können. Die Validierungsergebnisse der Validierung der beiden Datenbestände wurden anschließend zur Einschätzung der Entwicklung der Datenqualität innerhalb des vergangenen Jahres miteinander verglichen.

Gesondert betrachtet wurden hierbei die Software-Komponenten und deren Validierungsregeln, die im Rahmen dieser Arbeit etabliert wurden. Die Validierungsergebnisse der Datenbestände D2 und D3 wurden daher zusätzlich um alle Fehler bereinigt, die nicht durch

Validierungsregeln entstanden welche im Rahmen dieser Arbeit etabliert wurden. Diese Validierungsergebnisse wurden anschließend miteinander verglichen.

Ein weiterer Punkt der Betrachtung war die Analyse des Einflusses von neuen Validierungsregeln auf die Datenqualität bereits existierender Daten. Durch den mehrstufigen Datenqualitätsmanagementprozess der CERTAIN-Registerplattform und das nachträgliche Einfügen von neuen Validierungsregeln unterlagen bereits existierende Visiten der Datenbestände nicht zwangsläufig allen Validierungsregeln, bis die Visiten in den Zustand „Akzeptiert“ überführt wurden. Es sollte daher festgestellt werden, inwieweit eine Verbesserung der Datenqualität bei Altdaten aufgrund des Feedbacks durch neue Validierungsregeln zu erwarten ist. Es konnten für diese Betrachtung mehrere relevante Untergruppen der Visiten des Datenbestands D3 identifiziert werden. Zunächst die Gruppe der Visiten, die nach Inbetriebnahme der Validierungsregeln V2 nicht erneut bearbeitet worden waren. Zudem die Gruppe der Visiten, welche vor Inbetriebnahme der neuen Validierungsregeln V2 angelegt und nach der Inbetriebnahme erneut editiert worden waren. Im Zuge der Speicherung der editierten Daten dieser Visiten wurde dem Eingabepersonal das Ergebnis der Überprüfung des Datensatzes gegen die neuen Validierungsregeln V2 und V3 angezeigt. Das Eingabepersonal hatte daher die Möglichkeit eventuelle Fehler zu korrigieren. Die dritte Gruppe bilden Visiten, die innerhalb des zwischen den Entnahmezeitpunkten der beiden Datenbestände D2 und D3 verstrichenen Jahres neu angelegt und akzeptiert worden waren. Diese Visiten unterlagen während des gesamten Dateneingabe- und Datenqualitätsmanagementprozesses den Validierungsregeln V2 und V3 und somit durchgehend strikteren Regeln als die bisherigen Visiten.

Da die neuen Validierungsregeln eine Hürde für das Freigeben von Visiten und somit das Eingehen der Visite in wissenschaftliche Auswertungen darstellen, sollte zudem untersucht werden, wie festgestellte Verletzungen der Validierung von Wertebereichen bei Datenwerten in den Bereichen Labor und Medikation durch die Eingabekräfte korrigiert wurden. Durch gesondert entwickelte Analysemethoden wurden alle Ausreißer aus dem Datenbestand D2 extrahiert, welche die harten Validierungsgrenzen der Validierung von Wertebereichen von numerischen Datenelemente verletzen. Das Vorhandensein und der jeweilige aktuelle Datenwert dieser Ausreißer wurden anschließend im Datenbestand D3 untersucht. Hierfür wurde jeder korrigierte Datenwert dahingehend überprüft, ob der Datenwert gelöscht wurde oder der Datenwert nunmehr soweit herabgesetzt wurde, dass der Datenwert nur noch ein möglicher Fehler ist, oder der Datenwert nicht mehr als Fehler oder möglicher Fehler erkannt wurde.

Zudem sollte im Hinblick auf neu angelegte Visiten untersucht werden, ob die Werteverteilungen der validierten Labor- und Medikationsangaben auffällige Werteanhäufungen (Clustering) in Nähe der Validierungsgrenzen zeigten. Hierfür wurden alle Datenwerte jedes validierten Datenelementes des Datenbestands D3 pro Datenelement aggregiert und diese Werteverteilungen hinsichtlich des Auftretens von lokalen Maxima in der Dichtefunktion in Nähe der Validierungsgrenzen untersucht.

4 Überprüfung des Ansatzes der automatischen Validierung

Auf der Basis der Ausführungen des vorherigen Kapitels zum Vorgehen zur Überprüfung des Ansatzes der automatischen Validierung sollen in diesem Kapitel die Ergebnisse dieser Überprüfung präsentiert werden. Zur Überprüfung des Ansatzes der automatischen Validierung wurden die digitalisierten Versionen von 80 Arzt- und Jugendlichenbogen sowie 81 Elternbogen der Papierversion der Kerndokumentation zunächst auf Fehler hin untersucht, welche auf Probleme während der Digitalisierung der Papierbogen zurückzuführen waren. Anschließend wurden die in einem ASCII-Format digitalisierten Fragebogen gegen die neuen Validierungsregeln der elektronischen Version der Kerndokumentation (KRhOKo) geprüft und auf Fehler untersucht, welche durch die neuen Validierungsregeln hätten verhindert werden können.

4.1 Überprüfung der Digitalisierung der Fragebogen

Bei Überprüfung der Qualität des Digitalisierungsprozesses wurden bei einigen Fragebogen Diskrepanzen zwischen der Papierversion und dem Ergebnis der Digitalisierung auf Basis des bereits beschriebenen Vorgehens festgestellt, siehe Tabelle 2. So zeigte sich bei insgesamt sechs Arztbogen jeweils ein Fehler in der digitalisierten Version der Fragebogen, der auf den Digitalisierungsprozess zurückgeführt werden konnte. Bei den Elternbogen waren von solchen Fehlern 18 Fragebogen betroffen, bei den Jugendlichenbogen 17. Die beiden Patientenbogen wiesen eine durchschnittlich höhere Anzahl an Digitalisierungsfehlern pro Fragebogen mit Digitalisierungsfehlern auf als die betroffenen Arztbogen. Ebenso wiesen die Patientenbogen jeweils einen höheren Anteil an von Digitalisierungsfehlern betroffenen Fragebogen als der Arztbogen auf.

	Arztbogen ($n_1=80$)	Elternbogen ($n_2=81$)	Jugendlichenbogen ($n_3=80$)
ASCII-Bogen mit Digitalisierungsfehlern	6 (7,5%)	18 (22,2%)	17 (21,3%)
Durchschnittliche Anzahl Digitalisierungsfehler pro Bogen mit Digitalisierungsfehlern	1	1,45	1,65

Tabelle 2: Dokumente mit Fehlern durch den Digitalisierungsprozess

Die erfassten Fehler des Digitalisierungsprozesses wurden zur differenzierteren Betrachtung in verschiedenen Typen kategorisiert, Tabelle 3.

	Wert gelöscht	Wert erzeugt	fehlerhafte Transkription	sonstige Fehler	Summe
Arztbogen	2	3	1	–	6
Elternbogen	8	5	11	2	26
Jugendlichenbogen	8	6	14	–	28

Tabelle 3: Fehleranzahl im Digitalisierungsprozess aufgeteilt nach Fehlertyp

Da die Arztbogen nur wenige Fehler aufwiesen, werden diese nicht näher betrachtet. Die häufigste Fehlerart in den Patientenbogen waren Datenelemente, die im Papierbogen Datenwerte enthielten, im Zuge der Digitalisierung aber eine andere Merkmalsausprägung annahmen. Der Verlust von Datenwerten war die zweithäufigste Fehlerart, gefolgt von fälschlich transkribierten Datenwerten, welche im Papierbogen nicht existierten. Die zwei Fehler der Kategorie „Sonstige“ waren auf Datenelemente zurückzuführen, welche im Papierbogen doppelt gekreuzt waren, beispielsweise Skalen von 0 bis 10, aber nur ein Kreuz erhalten dürften und somit auch nur eine Zahl in das Format der digitalisierten Version übernommen werden konnte.

4.2 Überprüfung des Ansatzes der automatischen Validierung

Im nächsten Schritt wurden die digitalisierten Fragebogen gegen die Validierungsregeln des KRhOKo-Registers geprüft, siehe Tabelle 4.

	Arztbogen ($n_1=80$)	Elternbogen ($n_2=81$)	Jugendlichenbogen ($n_3=80$)
Bogen mit Verstößen gegen Validierungsregeln	14 (17,5%)	12 (14,8%)	7 (8,8%)
Bogen mit Verstößen durch Digitalisierungsfehler	3	2	4

Tabelle 4: Validierungsergebnis der digitalisierten Fragebogen

Im Zuge der Validierung der digitalisierten Fragebogen zeigte sich, dass die festgestellten Verstöße gegen die Validierungsregeln nur bei wenigen Fragebogen auf Digitalisierungsfehler zurückgeführt werden konnten. Im einzelnen wurde ersichtlich, dass 14 von 80 Arztbogen

die Validierungsregeln in mindestens einem Punkt verletzt. Bei drei Arztbogen gingen Verstöße gegen die Validierung auf Fehler in der Digitalisierung des jeweiligen Fragebogens zurück. Die häufigste Art von Verstößen in den zwölf problembehafteten Elternbogen waren mit elf Verstößen fehlende Werte, gefolgt von sechs Verstößen durch widersprüchliche Angaben, welche durch die Validierungsregeln erfasst wurden. Digitalisierungsfehler resultierten in zwei Elternbogen in Verstößen gegen die Validierungsregeln. In sieben problembehafteten Jugendlichenbogen wurden acht fehlende Werte gefunden. Zudem gab es drei Verstöße durch widersprüchliche Angaben, die durch Plausibilitätsprüfungen erfasst wurden. In vier Jugendlichenbogen wurden Verstöße gegen die Validierungsregeln festgestellt, welche auf Digitalisierungsfehlern beruhten.

4.3 Überprüfung der Vollständigkeit der Fragebogen

Im Zuge der Überprüfung wurde auch untersucht, inwieweit sich die Vollständigkeit von Eltern- und Jugendlichenbogen unterscheiden, um gegebenenfalls gesonderte Maßnahmen zur Sicherung der Vollständigkeit zu ergreifen. Dies liegt darin begründet, dass die Teilnehmer beim Ausfüllen der Online-Fragebogen nicht direkt durch das Personal der Einrichtung gebeten werden können, den Bogen zu vervollständigen. Das Ergebnis dieser Überprüfung wird in Tabelle 5 dargestellt. Die Vollständigkeit der Arztbogen wurde nicht untersucht, da Rückfragen durch die Registerzentrale an die jeweilige Einrichtung jederzeit möglich sind und die Arztbogen daher im Zweifelsfall auf Basis von gezielten Rückfragen vervollständigt werden.

	Elternbogen ($n_2=81$)	Jugendlichenbogen ($n_3=80$)
Durchschnittliche Vollständigkeit	92,2%	97,2%
Standardabweichung	$\pm 4\%$	$\pm 8\%$

Tabelle 5: Vollständigkeit der Patientenbogen

Da beide Verteilungen keine Normalverteilungen sind, wurde ein Wilcoxon-Mann-Whitney-Test zur Überprüfung der Gleichheit zwischen den beiden Gruppen herangezogen. Für den Test wurde $\alpha = 0,05$ gewählt. Der Test wurde mit R Version 3.4.3 durchgeführt und führte zu einem p-Wert von $4.095e^{-13}$ und $W = 5379.5$. Die Nullhypothese gleicher Vollständigkeit in den beiden Gruppen wurde entsprechend verworfen.

Auf der Basis dieses Ergebnisses wurde für die webbasierte Erfassung der Patientenbogen als erste Maßnahme eine automatisch berechnete Anzeige der Vollständigkeit des Bogens entwickelt. Diese Anzeige wurde mit dem Ziel entwickelt, die Vollständigkeit der Elternbogen zu erhöhen. Die Vollständigkeit des Fragebogens wird beim Ausfüllen des Fragebogens

zu jeder Zeit für die Eltern dargestellt. Zudem erhalten die Eltern zum Abschluss des Fragebogens einen Hinweis, wie vollständig der Fragebogen ausgefüllt wurde und die zusätzliche Information, dass ein vollständigerer Fragebogen für die Auswertung der Daten und Verbesserung der Versorgungslage hilfreicher ist. Diese Maßnahmen wurden zudem auch für den Jugendlichenbogen übernommen. Für die Jugendlichenbogen wurden keine weiteren spezifischen Maßnahmen zur Erhöhung der Vollständigkeit unternommen, da diese Bogen bereits nahezu vollständig ausgefüllt wurden.

5 Vorgehen zur Verbesserung der Datenqualität in Registern

Die Ergebnisse der vorangegangenen Überprüfung des Ansatzes der automatischen Validierung haben gezeigt, dass eine automatische Validierung ein wichtiger Teil des Datenqualitätsmanagementprozesses ist und Vorteile für die Datenqualität der CERTAIN-Register erwarten lässt. Dieser Ansatz ist prinzipiell kompatibel mit der TMF-Leitlinie. Die Leitlinie bietet jedoch keine Hinweise, wie die entsprechenden QI geeignet implementiert werden können. Daher war zunächst die Entwicklung eines Vorgehens zur Anwendung der QI der TMF-Leitlinie im Rahmen des Ansatzes der automatischen Validierung notwendig. Dieses Vorgehen wird in diesem Kapitel vorgestellt.

5.1 QI-Kategorien

Zur Entwicklung des allgemeinen Vorgehens wurden sowohl die TMF-Leitlinie als auch die in der TMF-Leitlinie referenzierte Literatur auf Möglichkeiten und Ansatzpunkte zur Formalisierung und Vereinfachung der Anwendung der QI hin untersucht. Die Leitlinie verweist bei einigen QI auf Plausibilitätsprüfungen zur Verbesserung der Datenqualität. Bei anderen QI verweist sie darauf, dass die Berechnung des jeweiligen QI nur notwendig sei, wenn derartige Fehler überhaupt auftreten könnten.

In der Arbeit von Arts et al. (Arts et al. 2002) werden drei Zeitpunkte für den Einsatz von Prozeduren zur Sicherung der Datenqualität benannt, siehe Abbildung 8. So werden Maßnahmen vor Beginn der Datenerfassung im Register, wie die Entwicklung benutzerfreundlicher Eingabemasken, unterschieden von Maßnahmen während der Dateneingabe, wie periodische Überprüfungen der Daten, sowie von Maßnahmen zur Verbesserung der Datenqualität, beispielsweise in Form von Reports an die Einrichtungen.

Auf der Basis dieser Zeitpunkte und der Anmerkungen der Leitlinie wurden vier Kategorien identifiziert, in welche die QI der TMF-Leitlinie eingeteilt werden können.

Die erste QI-Kategorie „Registerplanung“ bilden QI, hinsichtlich derer bereits durch eine schlüssige Definition von Datensatz und geeignete Implementierung von Eingabemasken während der Entstehung des Registers eine gute Datenqualität gewährleistet werden kann. Eine Berechnung der QI ist bei diesem Vorgehen nicht mehr erforderlich, da fehlerhafte und widersprüchliche Eingaben gar nicht erst in das Register eingegeben werden können und somit die durch die QI zu erfassenden Fehler nicht auftreten können. Beispielsweise kann auf die Berechnung des QI „Unerlaubte Werte bei qualitativen Datenelementen“ verzichtet werden, wenn sichergestellt werden kann, dass unerlaubte Datenwerte nicht eingegeben werden können. Dies kann durch die Festlegung von Auswahllisten und technischen Maßnahmen zur Verhinderung von Freitext-Eingaben abseits der Auswahllisten erreicht werden. Durch gewissenhaftes Ausschließen aller Fehlerquellen im Zuge dieses präventiven Ansatzes

zur Verhinderung der Eingabe von fehlerhaften Daten für alle QI dieser QI-Kategorie sind keine Fehler mehr zu erwarten. Auf eine Berechnung dieser QI für den Gesamtscore der Datenqualität kann in diesem Fall verzichtet werden.

Die zweite QI-Kategorie lautet „Automatische Validierung“. Für die Berechnung der QI dieser QI-Kategorie finden Überprüfungen von Datenwerten während der Dateneingabe statt, zum Teil durch Plausibilitätsprüfungen aber auch durch Prüfungen auf das Vorhandensein von Datenwerten (Vollständigkeit). Diese Überprüfungen können bereits automatisiert während der Eingabezeit zur Verhinderung von fehlerhaften Datenwerten erfolgen. Bei konsequenter Nutzung der Überprüfungen während der Eingabezeit können fehlerhafte Datenwerte oder Widersprüche in den Daten eingeschränkt und somit die Integrität der Daten verbessert werden. Eine Berechnung der QI dieser QI-Kategorie für das adaptive Feedback ist dann nicht erforderlich. Dieser Ansatz bedingt einen mehrstufigen Datenqualitätsmanagementprozess, in dem alle durch die Plausibilitätsprüfungen festgestellten Fehler beseitigt werden müssen, bevor die Daten in medizinische Auswertungen des Datenbestands des Registers eingehen können. Ohne diese gestaffelten Vorgänge würden die festgestellten Fehler den Eingabekräften zwar angezeigt, blieben aber ohne Konsequenzen. Ein Beispiel für einen solchen Datenqualitätsmanagementprozess bieten die Abläufe der CERTAIN-Registerplattform, dargestellt in Kapitel 2.4.3 Datenqualitätsmanagementprozess.

In der dritten QI-Kategorie „Report“ finden sich die QI, die im Rahmen des adaptiven Feedbacks im Sinne eines Reportings an die Einrichtungen Anwendung finden. Dies beinhaltet QI, deren berechneter Wert zumeist keine direkte Aussage über die Datenqualität eines Datensatzes erlaubt, viel mehr kommt deren Wert einer Übersichtsgröße gleich. QI dieser Kategorie ermöglichen beispielsweise den Vergleich von Einrichtungen untereinander. Abschließend beinhaltet die vierte QI-Kategorie „Datenverifizierung“ die verbleibenden QI. QI dieser QI-Kategorie erfordern für ihre Berechnung entweder den Zugriff auf die Originaldaten oder aber eine weitere Datenquelle, die zum Abgleich der Daten des Registers herangezogen werden kann. In beiden Fällen ist aufgrund von Medienbrüchen und Hürden für den Zugriff auf die Daten mit hohem Aufwand bei der Berechnung der QI zu rechnen. Die Berechnung von QI dieser QI-Kategorie sollte nur in Erwägung gezogen werden, wenn die entsprechenden Ressourcen vorhanden sind oder die Möglichkeiten zur Verbesserung der Datenqualität auf Basis der anderen QI-Kategorien ausgeschöpft wurden.

Tabelle 6 fasst die vier QI-Kategorien zusammen.

Die dargelegten QI-Kategorien orientieren sich an den von Arts et al. vorgestellten Prozeduren zur Verbesserung der Datenqualität, siehe Abbildung 8. Die QI-Kategorie „Registerplanung“ setzt an den von Arts et al. angemarkten Prozeduren der Definition des Datensatzes und der Entwicklung von Eingabemasken während der Planung des Registers an. Die QI-Kategorie „Automatische Validierung“ entspricht der Forderung nach automatischen Datenprüfungen während der Dateneingabe. Die QI-Kategorie „Report“ korrespon-

Registerplanung	Beinhaltet QI, die durch Maßnahmen auf Ebene des Datensatzes und der Eingabemasken abgelöst werden können
Automatische Validierung	Beinhaltet QI, die durch ganzheitliche Plausibilitätsprüfungen während der Eingabezeit abgelöst werden können
Report	Beinhaltet QI, die über den gesamten Datenbestand (einer Einrichtung) berechnet werden
Datenverifizierung	Beinhaltet QI zur Bewertung der Datenqualität im Rahmen einer durchzuführenden Datenverifizierung von Original- oder Referenzdaten

Tabelle 6: QI-Kategorien des entwickelten Vorgehens

diert mit der Forderung nach Datenqualitätsreports für die teilnehmenden Einrichtungen zur Verbesserung der Datenqualität. Die QI-Kategorie „Datenverifizierung“ geht zurück auf die Forderung nach Besuchen bei den teilnehmenden Einrichtungen und den Abgleich der dort vorliegenden Daten mit den Registerdaten.

Zudem konnten einige QI in der QI-Kategorie „Report“ identifiziert werden, die aufgrund ihres spezifischen Betrachtungsrahmens nur bei bestimmten Registern zum Einsatz kommen können, beispielsweise bei Turmorregistern. Für diese QI muss für das jeweilige Vorhaben entschieden werden, ob eine Anwendbarkeit gegeben ist.

5.2 Auswahl der relevanten QI

Die TMF-Leitlinie betont, dass nicht zwingend alle QI in allen Registern anzuwenden seien, sondern dass eine Auswahl der QI getroffen werden sollte. Zu diesem Zweck müssen sich Anwender der TMF-Leitlinie des Bezug jedes einzelnen QI zu ihrem Register bewusst werden. Das im Rahmen dieser Dissertation erarbeitete allgemeine Vorgehen schlägt vor, die QI auf Basis der drei Kriterien Anwendbarkeit, Implementierbarkeit und Aussagekraft zu beurteilen. Für die Bewertung der Anwendbarkeit auf das jeweilige Register sollten die Anwendungsszenarien des QI, die zu erwartende Breite der Anwendung und Möglichkeiten zur Reduktion des Umfangs der Berechnung des QI für das jeweilige Register analysiert werden. Die Implementierbarkeit der technischen Umsetzung sollte anhand des Umfangs der für die Berechnung erforderlichen Informationen und der Komplexität der Berechnung bewertet werden. Ebenso sollte für die Implementierbarkeit betrachtet werden, ob die Berechnung für das jeweilige Register automatisiert erfolgen kann oder ob manuelle Zwischenschritte erforderlich sind. Die Bewertung der Aussagekraft eines QI sollte aufgrund der in der TMF-Leitlinie beschriebenen Methoden zur Feststellung der Datenfehler und durch die Betrachtung der Schwere der festgestellten Fehler sowie ihres Einflusses auf die Datenqualität des jeweiligen Vorhabens erfolgen.

Die Bewertung der drei Kriterien selbst kann beispielsweise in Form einer fünfstufigen Likert-Skala erfolgen, die schlechteste Klasse kann durch - - symbolisiert werden, die Beste durch + +. QI die als nicht anwendbar für das jeweilige Register identifiziert wurden, sollten aus der Analyse der QI gestrichen werden. Bei diesen QI kann es sich beispielsweise um QI handeln, die nur für Tumorregister anwendbar sind. Wurden alle QI hinsichtlich der dargelegten Gesichtspunkte analysiert liegt eine registerspezifische Analyse der QI vor. Diese registerspezifische Analyse enthält nur noch QI, die für das jeweilige Vorhaben relevant sind, sowie eine Bewertung dieser QI im Hinblick auf die Anwendung dieser QI für das Vorhaben.

5.3 Priorisierung der Umsetzung der QI

Anhand dieser Bewertung der QI hinsichtlich ihrer Anwendbarkeit, Implementierbarkeit und Aussagekraft für das jeweilige Vorhaben kann eine Priorisierung der QI für das jeweilige Register vorgenommen werden. Durch die vergebenen Prioritäten kann eine gestaffelte Umsetzung der QI erfolgen, bei der die QI höchster Priorität zuerst umgesetzt werden sollten. Hierdurch soll es ermöglicht werden, zunächst die QI zu identifizieren und umzusetzen, die im jeweiligen Vorhaben maßgebliche Aussagen zur Datenqualität erlauben. Den QI einer jeden QI-Kategorie werden zu diesem Zweck Prioritäten von 1-3 zugewiesen, wobei Priorität 1 den höchsten Stellenwert einnimmt. Für die Auswahl der Werte zur Priorisierung sollte der Aussagekraft des QI der höchste Stellenwert zugemessen werden. Der Implementierbarkeit sollte ein geringfügig höherer Stellenwert als der Anwendbarkeit zugestanden werden, da eine Umsetzung eines QI nur sinnvoll erscheint, wenn diese praktikabel ist und die Berechnung bestenfalls automatisch erfolgen kann. Die Priorisierung sollte insgesamt die drei Kriterien berücksichtigen und auch im Hinblick auf die vergebenen Prioritäten der anderen QI derselben QI-Kategorie erfolgen. In die Überlegungen zur Priorisierung können darüber hinaus auch weitere Eigenheiten des Registers oder besondere Anforderungen der Benutzer einbezogen werden.

5.4 Festlegung der Gewichte der QI

Die erfolgte Priorisierung kann zudem zur Festlegung der registerspezifischen Gewichte der QI zur Berechnung des Gesamtscores der Datenqualität genutzt werden.

In der TMF-Leitlinie wird im Zuge der Berechnung des Gesamtscores jedem QI ein Gewicht zugewiesen. Dieses Gewicht wird mit dem berechneten Wert des QI multipliziert und geht anschließend in die Berechnung des Gesamtscores ein, siehe Kapitel 2.7.3 Qualitätsscore. Die TMF-Leitlinie beinhaltet beispielhaft vergebene Gewichte, betont aber die Notwendigkeit der Anpassung der Gewichte je nach Vorhaben, da nicht in jedem Vorhaben alle QI Anwendung finden.

Durch die Nutzung der QI-Kategorien und der innerhalb der QI-Kategorien vergebenen Prioritäten kann die Anzahl der zu gewichtenden QI zur Berechnung des Gesamtscores reduziert werden. So sind für die QI der Kategorien „Registerplanung“ und „Automatische Validierung“ keine Gewichtungen vorzunehmen, da diese QI ganzheitlich durch konkrete Maßnahmen zur Sicherung der Datenqualität nicht berechnet werden müssen. Die QI würden daher immer als „unauffällig“ in die Berechnung eingehen und den Gesamtscore positiv beeinflussen. Die Werte dieser QI dieser Kategorien sind für die Berechnung nicht mehr interessant, da sie nicht weiter optimiert werden können und somit nicht zu einer weiteren Verbesserung der Datenqualität beitragen können. Werden diese beiden Kategorien nicht mit in die Berechnung einbezogen, steigt gleichermaßen der Einfluss der QI der verbleibenden Kategorien „Report“ und „Datenverifizierung“ auf den Gesamtscore und somit im Verhältnis auch ihre Aussagekraft. Die TMF-Leitlinie hebt im Rahmen der Berechnung des Gesamtscores die Bedeutung der Source Data Verification hervor. Die QI der SDV, und somit die in der TMF-Leitlinie beschriebene Ebene Richtigkeit, sollten gemäß TMF-Leitlinie die höchsten Gewichte erhalten, um Situationen zu vermeiden, in denen das Ergebnis der SDV und der Gesamtscore zu sehr voneinander abweichen. Da die Kategorie „Datenverifizierung“ die mit der SDV assoziierten QI enthält, sollten dieser Kategorie demgemäß die höchsten Gewichte zugewiesen werden. Die TMF-Leitlinie schlägt Gewichte von 7-9 für die QI der Ebene Richtigkeit vor, in der die QI der SDV enthalten sind.

Durch die vorgenommene Priorisierung der QI innerhalb der QI-Kategorien wird die Gewichtung vereinfacht, da die zugewiesenen Prioritäten als Richtwert für die Gewichte herangezogen werden können. So können den QI-Kategorien generelle Gewichte zugewiesen werden, beispielsweise 7-9 für die QI-Kategorie „Datenverifizierung“ und 4-6 für die QI-Kategorie „Report“. Innerhalb der QI-Kategorien werden den QI anschließend die Gewichte gemäß der zugewiesenen Prioritäten zugeteilt. Den QI der Priorität 1 wird das höchste Gewicht der Kategorie zugewiesen, QI der Priorität 2 das nächsthöchste Gewicht und QI der Priorität 3 das geringste Gewicht. Ob der QI-Kategorie „Datenverifizierung“ oder der QI-Kategorie „Report“ die höheren Gewichte zugewiesen werden sollen, hängt vom jeweiligen Vorhaben ab. Wenn beispielsweise eine Berechnung der QI der QI-Kategorie „Datenverifizierung“ aufgrund der benötigten Ressourcen nicht angestrebt wird, müssen für diese QI-Kategorie keine Gewichte zugewiesen werden. Durch die in der TMF-Leitlinie beschriebene Normalisierung auf einer Skala von 0-100 bleiben die nach diesem Ansatz berechneten Gesamtscores mit Gesamtscores vergleichbar, die auf Basis des adaptiven Feedbacks der TMF-Leitlinie berechnet wurden.

Sollten Anwender des Ansatzes dennoch die QI der QI-Kategorien „Registerplanung“ und „Automatische Validierung“ für den Gesamtscore berechnen wollen, können den nunmehr vier QI-Kategorien Gewichtungen in Dreiergruppen von 1-12 zugewiesen werden. Die jeweils drei Gewichte der Gruppen entfallen dann nach dem beschriebenen Schema auf die zugewiesenen Prioritäten der QI. Die Wahl der jeweiligen Intervalle bleibt dem Anwender

überlassen. Eine Überlappung der Intervalle der Gewichte zwischen den QI-Kategorien ist möglich.

5.5 Allgemeines Vorgehen zur Anwendung des Ansatzes

Abbildung 11 fasst die durchzuführenden Schritte zur Bestimmung der Gewichte der QI zusammen.

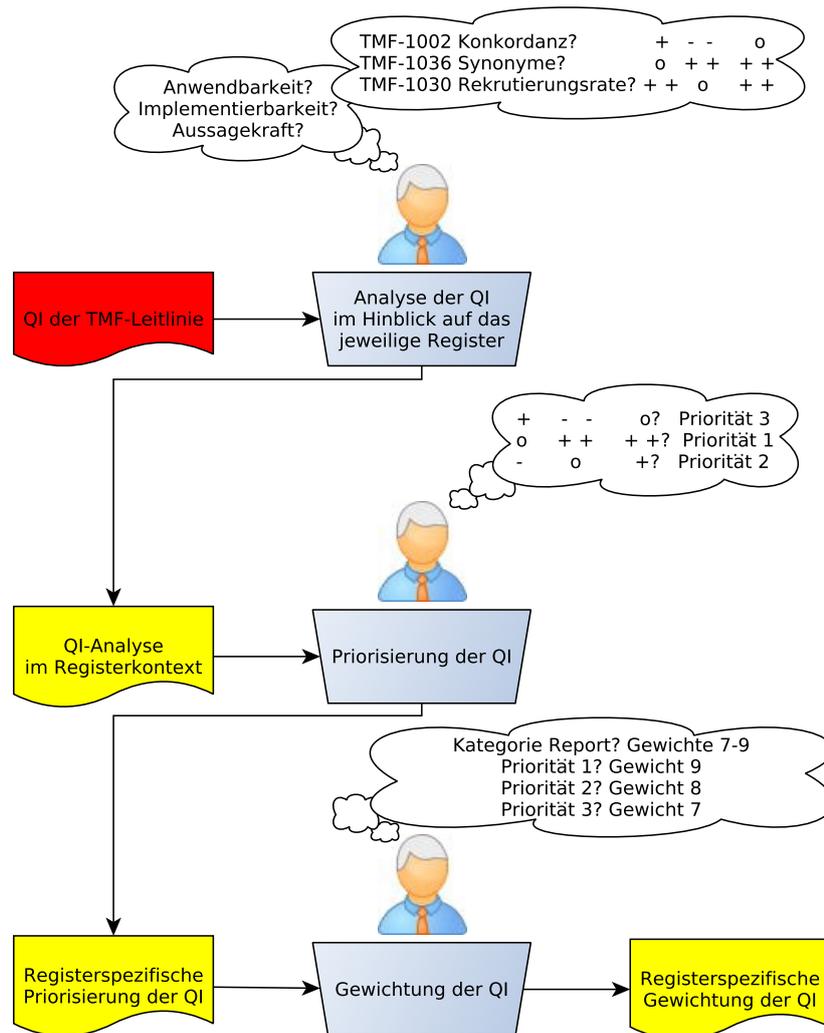


Abbildung 11: Zusammenfassung Ansatz zur Ermittlung der QI Gewichte

6 Registerempfehlung CERTAIN/-LI und KRhOKo

Das im vorherigen Kapitel entwickelte Vorgehen sollte genutzt werden um die Datenqualität drei CERTAIN-Register zu verbessern. Hierfür war es zunächst notwendig die QI der TMF-Leitlinie im Rahmen der drei Register zu analysieren, besonders zur späteren Umsetzung der QI der QI-Kategorie „Automatische Validierung“ in Form von Software-Komponenten. Zur Ermöglichung dieser Analyse mussten die technischen Anforderungen an die neuen Software-Komponenten der automatischen Validierung erarbeitet werden. Zunächst wurden jedoch die grundlegenden Gemeinsamkeiten und Unterschiede der drei Register ermittelt, um diese bei der Analyse der QI zu berücksichtigen. Die festgestellten Anforderungen und Rahmenbedingungen gingen in die Analyse der QI ein.

6.1 Vergleich der Register

Während sich alle drei Register eine gemeinsame Basis von Programmfunktionen teilen, ist die Ausrichtung der Datensätze und deren Aufbau zuweilen sehr unterschiedlich. So ist der Datensatz der beiden Transplantationsregister im Vergleich mit zu dem Datensatz im KRhOKo-Register deutlich größer. Es existieren aufgrund der klinischen Ausrichtung dieser beiden Register in ihren Datensätzen deutlich mehr numerische Datenelemente als in KRhOKo. KRhOKo legt sein Augenmerk auf qualitative Datenelemente, die zu Großteil in Form von Checkboxen umgesetzt sind. Die beiden Transplantationsregister nutzen für nicht-numerische Datenelemente hauptsächlich Auswahllisten, die bei wenigen Datenelementen durch Freitext ergänzt werden können. Auswahllisten, sowie die Möglichkeit der Ergänzung dieser Listen, existieren auch in KRhOKo, jedoch in noch geringerer Zahl als in den Transplantationsregistern. Alle drei Register bieten die Möglichkeit der fortlaufenden Dokumentation von Medikationsgaben, siehe Kapitel 2.4.2 Datenmodell. Darüber hinaus bieten die beiden Transplantationsregister die Möglichkeit der fortlaufenden Dokumentation auch für Laborwerte.

Die Register unterscheiden sich zudem im Ansatz der Datenerhebung. Während KRhOKo als Register mit bewusst niedrig gewählter Hürde zur Teilnahme aufgebaut ist, liegt das Augenmerk bei beiden Transplantationsregistern auf der Vollständigkeit der Dokumentation und der umfassenden Dokumentation des Therapieverlaufs der Patienten. So existieren im KRhOKo-Register nur wenige Pflichtfelder, um es den Eingabekräften leichter zu ermöglichen einen Patienten einzugeben, mit dem Ziel möglichst alle Patienten zu erfassen. In den beiden Transplantationsregistern übersteigt der Anteil der Pflichtfelder den der optionalen Felder, da eine möglichst breit aufgestellte Basisdokumentation angestrebt wird, um spätere optionale Studien zu erleichtern. Im CERTAIN-LI-Register wurde auf Wunsch von Projektpartnern im Rahmen einer Studie für viele Datenelemente die Auswahlmöglichkeit „unbekannt o.ä.“ nachträglich ergänzt.

Die Erhebungszeitpunkte der Visiten unterscheiden sich ebenfalls in den drei Registern. Während das CERTAIN-Register sechs Visiten im ersten Jahr und je zwei Visiten in den Folgejahren vorschreibt, sind es beim Schwesterregister CERTAIN-LI nur fünf Visiten im ersten Jahr und je eine Visite in den Folgejahren. KRhOKo wiederum rät zur jährlichen Erhebung der Fragebogen. Es ist den Einrichtungen jedoch freigestellt, die Fragebogen beliebig oft und nach eigens definierten Intervallen zu erheben.

Im Hinblick auf externe Datenquellen ergibt sich ebenso ein heterogenes Bild. Das CERTAIN-Register besitzt bidirektionale Schnittstellen zu Eurotransplant und zur Collaborative Transplant Study. CERTAIN-LI und KRhOKo hingegen besitzen keine weiteren Datenquellen als die eingehenden Einrichtungen selbst.

Weiterhin unterstützen CERTAIN und CERTAIN-LI die Dateneingabe der Einrichtungen durch die Möglichkeit des Importes von Laborwerten über einen Dateiupload. KRhOKo bietet die Möglichkeit, Daten von Patienten aus dem Vorjahr als Basis für das Anlegen des Patienten zu nutzen, so dass einige Parameter aus der Vorjahresdokumentation übernommen werden. Ebenso werden gezielt ausgewählte Daten beim Anlegen eines neuen Fragebogens aus dem vorherigen Bogen übernommen oder abgeleitet.

Als einziges der drei Register erhebt KRhOKo Patient-reported-outcome und ermöglicht es den Patienten, ihre Fragebogen direkt über die Weboberfläche in das Register einzugeben.

6.2 Technische Anforderungen

Trotz der unterschiedlichen Datensätze sollten sich die neuen Validierungsregeln auf alle drei Register anwenden lassen. Die neuen Validierungsregeln sollten zudem als Basis die existierende Validierung der CERTAIN-Register nutzen und diese erweitern oder mit verbesserter Funktionalität ersetzen. Die Anwendbarkeit auf alle drei Register erfordert einen generischen Ansatz, der sich nicht nur speziell auf die drei Register anwenden sollte, sondern auch anderen Registern als Beispiel für eine automatische Validierung dienen kann. Als Basis sollten daher nach wie vor der Hibernate Validator und JSR-303 dienen, durch deren technische Möglichkeiten gleichzeitig auch maßgeblich die Rahmenbedingungen der Implementierung festgelegt wurden. Neue und überarbeitete Funktionalitäten sollten gleichermaßen möglichst stark auf Standards und Programmibliotheken Dritter setzen, um unnötige Neuentwicklungen zu vermeiden. Weiterhin musste die Möglichkeit bestehen, alle Validierungsergebnisse auch künftig den Eingabekräften anzeigen zu können, damit diese die Verstöße gegen die Validierungsregeln zeitnah korrigieren und somit die Datenqualität verbessern können.

6.3 Analyse der TMF-Qualitätsindikatoren im Rahmen der CERTAIN-Register

Die Analyse der QI für die drei CERTAIN-Register erfolgte anhand der Kriterien allgemeine Anwendbarkeit, Implementierbarkeit der technischen Umsetzung sowie der Aussagekraft des berechneten Wertes des QI. Die Bewertungen der Kriterien der QI selbst erfolgte in Form von fünf Klassen. Die schlechteste Klasse wird durch - - symbolisiert, die Beste durch + +. In den folgenden Abschnitten wird das Ergebnis der Analyse der QI nach den dargelegten Kriterien beschrieben. Die Reihenfolge orientiert sich dabei nicht an den in Kapitel 5.1 QI-Kategorien vorgestellten Kategorien des allgemeinen Vorgehens, sondern nach der Reihenfolge, in der die QI in der TMF-Leitlinie angeordnet sind, damit sie dort besser nachgeschlagen werden können. Die hier dargestellte Analyse ist zu großen Teilen so allgemein gehalten, dass sie auch für andere Register zur Beurteilung der Kriterien genutzt werden kann. Die Übertragung auf die CERTAIN-Register soll als Beispiel dienen.

6.3.1 Qualitätsindikatoren der Ebene Integrität

Im Folgenden werden die Eigenschaften der QI der Ebene Integrität für die drei CERTAIN-Register ausgeführt.

TMF-1001 - Übereinstimmung mit Vorwert

Der QI „Übereinstimmung mit Vorwert“ soll gemäß Leitfaden genutzt werden, um den Anteil an Abweichungen zum vorherigen Datenwert bei unveränderlichen oder veränderlichen Merkmalen feststellen zu können. Gemäß TMF-Leitlinie ist die Berechnung des QI nur bei mehrfach erhobenen Datenelemente sinnvoll.

Anwendbarkeit

Wie in der TMF-Leitlinie beschrieben, kann der QI durch die nur einmalige Erfassung von unveränderlichen Merkmalen beeinflusst werden.

Der Aufwand zur Berechnung dieses QI lässt sich daher bereits während der Planung des Datensatzes des Registers durch eine sinnvolle Definition des Datensatzes reduzieren, indem unveränderliche Merkmale nur einmal erhoben werden. Im Falle der Berechnung des QI für ein existierendes Register kann die Identifikation von relevanten Datenelementen für die Berechnung genutzt werden, um mehrfach erhobene, unveränderliche Datenelemente zu identifizieren. Der Datensatz kann anschließend im Hinblick auf diese Mehrfacherhebung angepasst werden. Die Vermeidung von Mehrfacherhebung unveränderlicher Datenelemente sollte zudem die Akzeptanz beim Eingabepersonal erhöhen. Als unveränderliche, mehrfach erhobene Datenwerte wurden im Datensatz der beiden CERTAIN-Transplantationsregister nur die HLA-Werte des Organempfängers sowie seine ethnische Herkunft und sein Migrationshintergrund identifiziert. Aufgrund der stringenten Definition des Datensatzes findet

der QI daher bei unveränderlichen Datenelementen in den CERTAIN-Registern kaum Anwendung.

Aufgrund unterschiedlicher Registerpopulationen muss für jedes Forschungsvorhaben einzeln festgelegt werden, welche Datenelemente unveränderlich sind und somit für die Betrachtung des QI infrage kommen. Das in der TMF-Leitlinie aufgeführte Beispiel der im Rahmen der Messungenauigkeit unveränderlichen Körpergröße greift beispielsweise bei den drei CERTAIN-Registern nicht, da die Patienten noch erheblich wachsen können.

Bei veränderlichen Merkmalen sollte, wie in der TMF-Leitlinie beschrieben, der Betrag der akzeptablen Abweichung definiert werden, der eine Abweichung bei numerischen Merkmalen in der Ordnung von Messfehlern erlaubt. Bei Datenelementen, die starken Schwankungen unterworfen sein können, wie Laborwerten oder Medikationsgaben, ist die Anwendung des QI nur bedingt sinnvoll. Die Anwendbarkeit wird weiter eingeschränkt, wenn große Zeiträume zwischen den Beobachtungspunkten liegen, da hierdurch große Schwankungen auftreten können, beispielsweise wenn sich Laborwerte nach schwerer Erkrankung wieder normalisieren. Die Anwendbarkeit des QI wird bei den beiden auf klinische Datenwerte ausgerichteten Transplantationsregistern mit großen Abständen zwischen den Beobachtungseinheiten daher reduziert. Für die Abweichungen sollten pro Datenelement Referenzwerte aus der Literatur oder den Erfahrungen des medizinischen Personals des Registers herangezogen sowie das Datenelement im Kontext der Registerpopulation betrachtet werden. Auf der Basis dieser Überlegungen ist es beispielsweise sinnvoll, für die Körpergröße bei pädiatrischen Registern lediglich eine maximale negative Abweichung festzulegen. Die Festsetzung einer plausiblen, maximalen positiven Abweichung der Körpergröße wäre bei beiden Transplantationsregistern aufgrund der Diversität der Population schwierig. Dies liegt zum einen an den verschiedenen sozioökonomischen Verhältnisse der Patienten, welche das Wachstum beeinflussen (Mei et al. 1998). Zum anderen durchlaufen die Patienten während ihrer Kindheit unterschiedliche Wachstumsphasen was die Definition von akzeptablen Abweichungen weiter erschwert.

Die Feststellung von Abweichungen kann zusätzlich in ihrer Komplexität steigen, wenn mehrere Datenelemente in Zusammenhang stehen, da in diesen Fällen Wissen über die Menge der Merkmalsausprägungen und den Bezug zwischen den Merkmalen für die Feststellung von Abweichungen notwendig ist. Beispielsweise wird im Falle der Fragen der Anamnese des KRhOKo-Registers nach dem Auftreten bestimmter Erkrankungen gefragt, die Antwortmöglichkeiten der Erkrankungen bestehen aus „aktuell“, „jemals“ und „nie“, wobei „aktuell“ und „jemals“ zeitgleich angegeben werden können. Sollte zum ersten Erhebungszeitpunkt „aktuell“ gewählt werden müsste dieser Datenwert zum nächsten Erhebungszeitpunkt als „jemals“ angegeben werden um mit dem Vorwert übereinzustimmen. Da „aktuell“ und „jemals“ zeitgleich angegeben werden können und somit nicht in den sel-

ben Datenelementen gespeichert werden, ist für die Anwendung des QI zwingend Wissen über den Bezug zwischen den Datenelementen erforderlich.

Implementierbarkeit

Ausgehend von den diskutierten Beispielen der Körpergröße und dem Auftreten von Erkrankungen kann die Berechnung je nach betrachtetem Datenelement eine hohe Komplexität annehmen, da mitunter eine Vielzahl von Faktoren und Konstellationen berücksichtigt werden muss. Zudem muss die Berechnung für jedes identifizierte Datenelement einzeln durchgeführt werden. Die Realisierung der Berechnung in einer objektorientierten Programmiersprache bietet sich daher an, da hier gute Wiederverwendbarkeit und Abstraktion der Berechnungsmethoden gegeben sind. Dies wird durch die Möglichkeit der Verwendung der Programmier Technik Reflexion zur Generalisierung der Berechnungsmethoden weiter unterstrichen. Insgesamt scheint die Realisierung der Berechnung des QI für die CERTAIN-Register mit erhöhtem Aufwand verbunden.

Aussagekraft des QI

Gemäß TMF-Leitlinie sind nicht plausible Abweichungen zum Vorwert ein Hinweis auf einen Datenfehler. Zur besseren Einschätzung der Schwere der festgestellten Datenfehler kann es sinnvoll sein, den QI mehrfach mit unterschiedlichen akzeptablen Abweichungen zu berechnen. Anhand der Schwere der Datenfehler kann eine Einordnung in die Kategorien „Sichere Widersprüche/Fehler“ und „Mögliche Widersprüche/Warnungen“ vorgenommen werden. Anhand dieser Einordnung ist es möglich, einen genaueren Überblick über die Schwere der Fehler und somit die Datenqualität zu erhalten, als wenn nur eine Fehlerkategorie verwendet wird. Als sichere Widersprüche könnten beispielsweise physiologisch unmögliche Abweichungen zum Vorwert gelten, nicht plausible Abweichungen als mögliche Widersprüche.

Der Wert des QI bietet insgesamt eine gute Aussagekraft zur Datenqualität der drei CERTAIN-Register, da der Anteil potentieller Datenfehler eines Datenelements anhand objektiver Kriterien identifiziert werden kann.

Die zusammenfassende Bewertung des QI findet sich in Tabelle 7.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
o	-	+	automatische Validierung

Tabelle 7: Bewertung TMF-1001 - Übereinstimmung mit Vorwert

TMF-1002 - Konkordanz

Der Indikator „Konkordanz“ misst den Anteil der nicht mit einer Referenzquelle übereinstimmenden Datenwerte an der Gesamtzahl der überprüften Datenwerte.

Anwendbarkeit

Im Falle von Referenzquellen, die nicht direkt unter der Kontrolle der Registerzentrale stehen, beispielsweise einem Krankenhausinformationssystem, gestaltet sich die Überprüfung aufgrund von mangelnden Zugriffsrechten des Registerpersonals und der Datenschutzproblematik daher schwierig. Gerade bei rein elektronischen Registern, wie den drei CERTAIN-Registern, gestaltet sich die Anwendung des QI daher schwierig. Bei Datenübermittlung mit Medienbrüchen, zum Beispiel der Digitalisierung von Papierbogen mit anschließendem Import des Digitalisierungsergebnisses in das Register, kann der QI genutzt werden, um Aussagen über die Qualität des Digitalisierungsprozesses zu machen. Dies kann erreicht werden, indem die Datenwerte aus dem Ergebnis des Digitalisierungsprozesses mit den Datenwerten der Papierbogen abgeglichen werden und die Berechnung entsprechend der Definition des QI durchgeführt wird. Dies ist besonders für das KRhOKo-Register relevant, da während der Entwicklung des Registers Planungen stattfanden, digitalisierte Papierbogen in das elektronische Register zu überführen.

Im Falle einer direkten, elektronischen Datenübermittlung an das Register aus dem Primärsystem der eingebenden Einrichtung entspricht dieser QI, gemäß Leitlinie, den QI „Übereinstimmung der Registerdaten mit den Originaldaten“ (TMF-1044, TMF-1045), da Referenzquelle und Quelle der Datenerhebung identisch sind. Letztere QI sind Teil der SDV und würden aus den vorher genannten Gründen eine Überprüfung vor Ort bedingen, die im Hinblick auf die limitierten Ressourcen der Registerzentrale und die Anzahl der teilnehmenden Einrichtungen nur schwer durchführbar sein kann. Im Falle der drei CERTAIN-Register beispielsweise stehen hierfür keine Ressourcen zur Verfügung. Die Überprüfung der Konkordanz bei rein elektronischen Datenübermittlungen kann helfen, Fehler in der Datenharmonisierung der unterschiedlichen Datensätze von Referenzquelle und Register sichtbar zu machen, wenn sich der originale Datenwert der Referenzquelle und der im Register gespeicherte Datenwert nach der Übertragung unterscheiden. Für CERTAIN besitzt dieser Betrachtungswinkel aufgrund der Schnittstellen zu Eurotransplant und zur Collaborative Transplant Study Relevanz.

Für die Sinnhaftigkeit der Berechnung dieses QI ist der Datenerhebungsprozess des Registers maßgeblich. Bei einer Datenerhebung mit Medienbrüchen ist der QI als sehr sinnvoll zu erachten, da durch Medienbrüche Fehler entstehen können und die Berechnung des QI den Anteil dieser Fehler beziffern kann. Im Fall einer rein elektronischen Erhebung, etwa über eine webbasierte Plattform, entspricht der QI zwei QI aus der SDV, da hier die Regis-

terdaten mit den Originaldaten verglichen werden würden. Der QI kann daher in diesem Fall gegenüber den QI der SDV zurückgestellt werden.

Implementierbarkeit

Der Aufwand des Abgleichs der zwei Datenbestände hängt von der Menge der zu untersuchenden Datenelemente ab, variiert entsprechend und lässt sich daher nicht allgemeingültig schätzen. Für die beiden Transplantationsregister ist aufgrund des Umfangs des Datensatzes ein hoher Aufwand zu erwarten, sollten alle Datenelemente mit einer Referenzquelle verglichen werden. Wenn über eine Schnittstelle, beispielsweise Eurotransplant, nur eine Teilmenge des Datensatzes ausgetauscht wird, scheint die Überprüfung der Konkordanz besser durchführbar. Im Hinblick auf eine automatische Berechnung des QI für elektronische Importfunktionen bleibt festzuhalten, dass für diese Überprüfung eine weitere Importfunktion notwendig wäre, um die Eingangsdaten automatisch mit dem Ergebnis der Importfunktion vergleichen zu können. Wenn die Konkordanz über Medienbrüche hinweg untersucht werden soll, ist jedoch ein hoher Arbeitsaufwand zu erwarten, da hier ein manueller Abgleich stattfinden müsste.

Aussagekraft des QI

Da für diesen Indikator nur zwei Repräsentationen der Wirklichkeit verglichen werden, stellt sich die Frage nach den Vorteilen der Überprüfung mittels dieses Indikators im Hinblick auf die SDV, welche die Daten des Registers gegen die Originaldaten prüft. Der QI ließe sich daher am besten als Indikator während einer SDV anwenden, da hier die Möglichkeit besteht, die Registerdaten gegen die Referenzquelle und diese wiederum gegen die Originaldaten zu prüfen. Lediglich im Hinblick auf Transformationen im Rahmen von Datenübertragungen oder Medienbrüchen scheint die Aussagekraft im Verhältnis zum Aufwand zu stehen.

Unter diesen Gesichtspunkten findet sich die abschließende Bewertung des QI in Tabelle 8.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+	- -	o	Datenverifizierung

Tabelle 8: Bewertung TMF-1002 - Konkordanz

TMF-1003 - Widerspruchsfreiheit

Der QI „Widerspruchsfreiheit“ berechnet den Anteil der Datenwerte, die mindestens eine Plausibilitätsprüfung verletzen an allen überprüften Datenwerten. Die Berechnung kann für einzelne Datensätze, Beobachtungseinheiten oder für den gesamten Datenbestand durchgeführt werden, wobei der Wert des QI nur für Datenelemente berechnet werden sollte.

Anwendbarkeit

Die in der TMF-Leitlinie beschriebene Beeinflussung des Wertes des QI durch automatische Plausibilitätsprüfungen zur Eingabezeit, beispielsweise die in der TMF-Leitlinie genannte geschlechtsabhängige Überprüfung von Schwangerschaft, sind ein effektives Werkzeug zur Verbesserung der Datenqualität, da die Benutzer zeitnahes Feedback erhalten und die Eingabe direkt korrigieren können. Besonders bei der mehrfachen Wiederverwendung von Modulen, etwa für Medikationsangaben, können Plausibilitätsprüfungen in Form von automatisierten Logikprüfungen effizient umgesetzt werden. In allen drei CERTAIN-Registern finden Plausibilitätsprüfungen im Rahmen der automatisch durchgeführten Datenvalidierung bereits Anwendung.

Neben der Beeinflussung des Wertes des QI durch die automatisierte Ausführung der Plausibilitätsprüfungen während der Datenerfassung kann auch die Abbildung der Plausibilitätsprüfungen über die Eingabemasken positive Effekte auf den Wert des QI haben. So können widersprüchliche Angaben verhindert werden, wenn Datenelemente in der Eingabemaske aufgrund von vorherigen Angaben ausgeblendet werden. Für das Beispiel der TMF-Leitlinie könnten beispielsweise Fragen zu einer Schwangerschaft ausgeblendet werden, wenn als Geschlecht „männlich“ angegeben wird. Dynamische Eingabemasken können Plausibilitätsprüfungen sinnvoll ergänzen. Dieser Ansatz wird in allen drei CERTAIN-Registern so oft wie möglich genutzt um Falschangaben von vornherein ausschließen zu können.

Aufgrund von Importfunktionen, die Daten an der Oberfläche vorbei unverifiziert aus Quellen übernehmen, sind Plausibilitätsprüfungen jedoch in jedem Fall zu implementieren. Eine Beschränkung der Eingabemöglichkeiten in der Eingabeoberfläche kann Frustration bei Eingabekräften vermeiden, da fehlerhafte Angaben verhindert werden. Hierdurch müssen im Zweifelsfall keine Fehler angezeigt werden, welche die Eingabekräfte später erneut überprüfen müssten. Bei geringen verfügbaren Ressourcen während der Entwicklung eines Registers, das keine Importfunktion bietet, kann die Beschränkung der Eingabemöglichkeiten über die Eingabemasken daher ein effizientes Mittel zur Sicherung der Datenqualität sein, bis Plausibilitätsprüfungen umgesetzt werden können.

Implementierbarkeit

Bei einer geringen Menge von einzubeziehenden Datenelementen lassen sich Plausibilitätsprüfungen einfach etablieren. Mit steigender Anzahl der einzubeziehenden Datenelemente für die Überprüfung einer Plausibilitätsprüfung steigt die Komplexität der Prüfung und somit der Aufwand der Umsetzung. Komplizierte Zusammenhänge, wie beispielsweise die Überprüfung der Plausibilität von verabreichten Medikamentenregimen, lassen sich nur schwierig durch Logikregeln abbilden, da hier unter anderem der Erhebungszeitpunkt des Datensatzes und damit die zum Zeitpunkt der Medikamentengabe gängige Behandlung eine zentrale Rolle spielen. In den beiden CERTAIN-Transplantationsregistern werden beispiels-

weise retrospektiv Transplantationen seit 1995 erfasst. Eine Überprüfung der Plausibilität der dokumentierten Medikamentenregime wäre sehr komplex. Für Register mit einer langen Laufzeit sind komplexe Überprüfungen daher aufgrund der Fülle der zu berücksichtigenden Konstellationen unter Umständen nicht effizient umsetzbar. Bei der Definition der Plausibilitätsprüfungen zur Sicherung der Widerspruchsfreiheit muss der Aufwand im Verhältnis zum Nutzen abgeschätzt werden, da mit steigender Komplexität und damit verbundenen steigendem Implementierungsaufwand alternative Maßnahmen zielgerichteter erscheinen. Eine gezielte Auswahl der umzusetzenden Plausibilitätsprüfungen in Zusammenarbeit von medizinischem Personal, Datenmanagement und IT ist dringend anzuraten, der Fokus sollte sich dabei auf Plausibilitätsprüfungen richten, die einfache Sachverhalte überprüfen, die im besten Fall in mehreren Modulen wiederholte Anwendung finden.

Aussagekraft des QI

Die Widerspruchsfreiheit der Registerdaten ist als eines der wichtigsten Ziele der Qualitätsprüfung zu erachten, da Widerspruchsfreiheit und Integrität elementar für die Glaubwürdigkeit der Daten sind. Der Wert des QI ist daher ein sehr wichtiger Indikator der Datenqualität des Registers, auch wenn er keine Aussagen über die tatsächliche Richtigkeit der Daten erlaubt. Da sowohl sichere als auch mögliche Widersprüche Teil der Betrachtung sind, ist die Aussagekraft als gut einzuschätzen. Eine hohe Anzahl von durch Plausibilitätsprüfungen festgestellten Widersprüchen kann ein Hinweis auf fehlerhafte Plausibilitätsregeln sein, wenn diese die Wirklichkeit nicht korrekt abbilden.

Tabelle 9 fasst die abschließende Bewertung des QI zusammen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	+	+	automatische Validierung

Tabelle 9: Bewertung TMF-1003 - Widerspruchsfreiheit

TMF-1035 - Endlosüberlebende

Dieser QI berechnet den Anteil von Patienten an der Gesamtpopulation des Registers, bei denen formal nie der Tod festgestellt wird. Gemäß TMF-Leitlinie lassen sich anhand des QI Aussagen über Vollständigkeit und Validität von Follow-ups treffen. Wenn Patienten die zu erwartende Überlebenszeit übersteigen, können fehlende Informationen oder mangelnde Nachsorge die Gründe sein.

Anwendbarkeit

Eine Überprüfung im Hinblick auf Endlosüberlebende ist umso wichtiger, je geringer die verbleibende Lebensspanne der Beobachtungseinheiten aufgrund ihrer Erkrankung einzuschätzen ist. Bei geringer Lebenserwartung der Beobachtungseinheiten haben Endlosüber-

lebende einen statistische größeren Einfluss und verzerren daher die durchschnittliche Überlebenszeit der Studienteilnehmer überproportional. Die Relevanz des QI hängt auch von dem durchschnittlichen Alter der Registerpopulation ab, da mit steigendem Durchschnittsalter der Patienten Endlosüberlebende immer unwahrscheinlicher werden. Im Rahmen der Anmerkungen der TMF-Leitlinie zum QI „Vorzeitig aus dem Register ausscheidende Beobachtungseinheiten“ gibt die TMF-Leitlinie bei Alterskrankheiten eine schwer zu schätzende Dunkelziffer bei Todesfällen zu bedenken (Nonnemacher et al. 2014, S, 66).

Grundsätzlich lässt sich der QI auf alle Register anwenden, die den Erkrankungsbeginn, das Geburtsdatum und das Ableben des Patienten dokumentieren. Der QI ist daher auf die beiden Transplantationsregister anwendbar. Die Ausnahme unter den drei Registern bildet das KRhOKo-Register, da hier keine Möglichkeit besteht den Tod des Patienten zu dokumentieren.

Implementierbarkeit

Die Identifizierung von Endlosüberlebenden ist durch den Vergleich der Überlebenszeit der Beobachtungseinheit mit der durchschnittlichen Überlebenszeit möglich. Es ist für die Berechnung notwendig, einen Wert für eine auffällige Überschreitung der durchschnittlichen Überlebenszeit festzulegen. Beide Punkte sind einfach zu adressieren, eine automatisierte Erfassung der Endlosüberlebenden ist daher mit geringem Aufwand möglich.

Aussagekraft des QI

Bei meist tödlich endenden Erkrankungen oder hohem Alter der Registerpopulation ist die Anwendung des QI aufgrund der genannten Szenarien sehr sinnvoll. Der Indikator besitzt hingegen geringe Relevanz, wenn es sich um Erkrankungen handelt, die in der Regel nicht tödlich enden oder wenn eine lange Überlebenszeit zu erwarten ist. Die Berechnung des QI ist nur sinnvoll, wenn der betrachtete Datenbestand aktuell ist, da sonst keine Aussage getroffen werden kann, ob der Patient nicht in der Zwischenzeit verstorben und somit kein Endlosüberlebender mehr ist.

Rheumaerkrankungen verlaufen in der Regel nicht tödlich, der QI wäre daher für KRhOKo zu vernachlässigen, wenn die Dokumentation eines Todeszeitpunktes nachträglich ermöglicht werden sollte. In den beiden Transplantationsregistern treten Todesfälle auf, je nach Transplantat kann die Zeit bis zum Transplantatsverlust oder Tod stark variieren. Im Nierentransplantationsregister muss der Transplantatsverlust nicht mit dem Tod einhergehen. Patienten, die einen Transplantatsverlust erlitten haben, können durch Dialyse am Leben erhalten werden. Aussagen über das weitere Überleben sind in diesen Fällen nicht möglich, da der betroffene Patient mit dem Transplantatsverlust aus dem Register ausscheidet. Im Falle des Lebertransplantationsregisters ist ein sehr kurzfristiges Überleben nach Transplantatsverlust möglich. Ein Transplantatsverlust geht in diesem Register in der Regel mit dem Tod des Patienten oder einer Retransplantation eines neuen Organs einher. Der etwaige Tod des Patienten kann in beiden Transplantationsregistern in der letzten Visite des

Patienten dokumentiert werden. Für die Registerzentralen ist es aber selbst nicht feststellbar, ob und wann Patienten versterben. Die Registerzentrale ist in diesem Fall auf korrekte Informationen der Einrichtung angewiesen. Für das CERTAIN-Register ist die Aussage des QI daher von geringem Wert einzuschätzen. Für das CERTAIN-LI-Register ist die Aussage des QI aufgrund der geringen Überlebenszeit bei Organverlust relevant.

Die zusammenfassende Bewertung des QI findet sich in Tabelle 10.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+	+ +	o	Report

Tabelle 10: Bewertung TMF-1035 - Endlosüberlebende

TMF-1004 - Sichere Widersprüche/Fehler

Der QI „Sichere Widersprüche/Fehler“ ist eine Unterkategorie des QI TMF-1003 „Widerspruchsfreiheit“ und berechnet den Anteil von Plausibilitätsprüfungen mit Ergebnis eines sicheren Widerspruchs oder Fehlers an der Anzahl der durchgeführten Plausibilitätsprüfungen. Ziel der Berechnung sind, identisch zu TMF-1003, einzelne Datensätze, Beobachtungseinheiten oder der gesamte Datenbestand.

Wie TMF-1003 sollte auch dieser QI, gemäß Leitlinie, durch die Durchführung von Plausibilitätsprüfungen auf den einbezogenen Daten berechnet werden.

Anwendbarkeit

Da dieser QI eine Unterkategorie von TMF-1003 ist, greifen die selben Überlegungen zur Anwendbarkeit wie bei TMF-1003. Eine Evaluation der Plausibilitätsprüfungen bereits zur Eingabezeit, sowie die Beschränkung der Eingabemasken auf Basis von logischen Zusammenhängen, kann entsprechend den Aufwand zur Berechnung des QI reduzieren und den Wert des QI positiv beeinflussen. Der Umfang der Berechnung des QI kann weiter reduziert werden, indem nur Plausibilitätsprüfungen einbezogen werden, die auch tatsächlich in einem sicheren Widerspruch oder Fehler resultieren können. Da Plausibilitätsprüfungen auf sichere Widersprüche und Fehler nicht an bestimmte Voraussetzungen oder Datentypen geknüpft sind, bieten Plausibilitätsprüfungen eine sehr gute Breite der Anwendung und sind in allen drei CERTAIN-Registern Teil der automatischen Datenvalidierung.

Implementierbarkeit

Gleichermaßen steht die Implementierbarkeit wie bei TMF-1003 im Verhältnis zur Komplexität der, anhand der zu etablierenden Plausibilitätsprüfungen, zu überprüfenden Zusammenhänge. Im Falle von sicheren Widersprüchen kann es jedoch aufgrund der Schwere der Widersprüche sinnvoll sein, mehr Zeit in eine Umsetzung der Plausibilitätsprüfungen zu investieren, um offensichtliche Fehler von vornherein ausschließen zu können. Aufgrund

der Nutzung der Java Bean Validation 1.0 (JSR 303) und des Hibernate Validators werden alle Ergebnisse von Plausibilitätsprüfungen in den drei CERTAIN-Registern als sichere Widersprüche und Fehler interpretiert. Die programmierten Plausibilitätsprüfungen sind entsprechend auf diesen Sachverhalt zugeschnitten. Komplexe Plausibilitätsprüfungen erfordern auch mit diesen Hilfsmitteln eigens entwickelte Programmteile, welche die jeweiligen Zusammenhänge zwischen Datenelementen berücksichtigen. Die Berechnung des QI kann automatisiert erfolgen und ist durch Nutzung des Hibernate Validators einfach möglich. Der QI bildet eine Unterkategorie von TMF-1003 „Widerspruchsfreiheit“, die Berechnung unterscheidet sich jedoch im Hinblick auf den Fokus der Berechnung von TMF-1003 auf Datenwerte und Plausibilitätsprüfungen in diesem QI. Auch dieser QI kann jedoch als Anteil der Datenwerte, die eine Plausibilitätsprüfung verletzen die in einem Fehler resultiert, an der Gesamtzahl der überprüften Datenwerte berechnet werden. Dies verbessert die Vergleichbarkeit der beiden QI.

Aussagekraft des QI

Ähnlich TMF-1003 bietet die Erkennung von sicheren Widersprüchen und Fehlern ein sehr gutes Potenzial zur Erkennung von mangelnder Datenqualität. Im Falle von sicheren Widersprüchen wird dies noch unterstrichen, da hier durch automatische Plausibilitätsprüfungen ohne Zutun des Datenqualitätsmanagements des Registers direkt Feedback an das Eingabepersonal gegeben werden kann. In der TMF-Leitlinie werden Umfang und Tiefe der Plausibilitätsprüfungen nur für TMF-1005 „Mögliche Widersprüche/Warnungen“ als Einflussfaktoren genannt. Da die Berechnung dieses QI auch anhand von Plausibilitätsprüfungen erfolgt, sind Umfang und Tiefe daher ebenfalls als Einflussfaktoren für TMF-1004 zu erachten. Die Aussagekraft des Wertes des QI kann zudem verbessert werden, wenn anstatt aller Plausibilitätsprüfungen nur solche Plausibilitätsprüfungen in die Berechnung eingehen, die auch tatsächlich in einem erkannten sicheren Widerspruch oder Fehler resultieren können.

Plausibilitätsprüfungen mit einem Ergebnis das als sicherer Widerspruch/Fehler gewertet wird, werden in den drei CERTAIN-Registern zur Validierung während der Eingabezeit genutzt und sind elementarer Bestandteil des Datenqualitätsmanagements der drei Register. Eine hohe Anzahl von durch Plausibilitätsprüfungen festgestellten Widersprüchen kann ein Hinweis auf fehlerhafte Plausibilitätsregeln sein, wenn diese die Wirklichkeit nicht korrekt abbilden. Der QI bietet eine sehr gute Aussagekraft zur Datenqualität der drei CERTAIN-Register aufgrund der Schwere der festgestellten Fehler und ist somit essenziell für die Betrachtung der Datenqualität.

Tabelle 11 fasst die abschließende Bewertung des QI zusammen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	+	+ +	automatische Validierung

Tabelle 11: Bewertung TMF-1004 - Sichere Widersprüche/Fehler

TMF-1005 - Mögliche Widersprüche/Warnungen

Der QI „Mögliche Widersprüche/Warnungen“ ist eine Unterkategorie des QI TMF-1003 „Widerspruchsfreiheit“ und identifiziert im Vergleich zum vorangegangenen QI Widersprüche eines geringeren Schweregrads. Ziel der Berechnung sind, gleich TMF-1003, einzelne Datensätze, Beobachtungseinheiten oder der gesamte Datenbestand.

Anwendbarkeit

Der QI verhält sich gleich dem vorherigen QI. Allein eine Beschränkung der Eingabemasken auf Basis der Erkennung von fehlerhaften Datenwerten durch Plausibilitätsprüfungen ist nicht möglich. Dies liegt darin begründet, dass als fehlerhaft erkannte Datenwerte zwar nicht plausibel, aber prinzipiell möglich sind und somit zugelassen werden müssen. Die Etablierung von Plausibilitätsprüfungen für Sachverhalte von geringerer Bedeutung für das Forschungsvorhaben ist dabei nicht an spezielle Datentypen geknüpft und kann somit auf viele Datenelemente und logische Zusammenhänge zwischen diesen Datenelementen angewendet werden. Aufgrund von vielen zusammenhängenden Datenelementen und Folgefragen in den CERTAIN-Registern sind Plausibilitätsprüfungen zu möglichen Widersprüchen breit anwendbar. Bisher gibt es in den drei CERTAIN-Registern keine Möglichkeit Plausibilitätsprüfungen zu etablieren, welche in einem möglichen Fehler oder einer Warnung resultieren.

Implementierbarkeit

In Gegensatz zu TMF-1004 sollten tendenziell nur Plausibilitätsprüfungen mit geringer Komplexität und somit geringerem Implementierungsaufwand umgesetzt werden, da Implementierungsaufwand und Aussagekraft des Ergebnisses der Plausibilitätsprüfung im Verhältnis zueinander stehen sollten. Die Komplexität der eigentlichen Berechnung des Wertes des QI anhand der Evaluation der Plausibilitätsprüfungen ist als gering zu erachten. Die Berechnung des QI kann mitunter vereinfacht werden, indem nur Plausibilitätsprüfungen einbezogen werden, die auch tatsächlich in einem möglichen Widerspruch oder einer Warnung resultieren können. Um Plausibilitätsprüfungen dieses Typs zu etablieren, müsste eine Erweiterung der bestehenden Datenvalidierung der drei CERTAIN-Register vorgenommen werden. Nach Umsetzung dieser neuen Plausibilitätsprüfungen auf Basis von JSR-303 scheint eine automatisierte Berechnung des QI unter Nutzung des Hibernate Validators möglich.

Der QI bildet eine Unterkategorie von TMF-1003 „Widerspruchsfreiheit“. Die Berechnung unterscheidet sich jedoch im Hinblick auf den Fokus der Berechnung von TMF-1003 auf Datenwerte und Plausibilitätsprüfungen in diesem QI. Auch dieser QI kann jedoch als Anteil der Datenwerte, die eine Plausibilitätsprüfung verletzen, die in einem möglichen Fehler oder Widerspruch resultiert, an der Gesamtzahl der überprüften Datenwerte berechnet werden. Dies verbessert die Vergleichbarkeit der beiden QI.

Aussagekraft des QI

Da es sich um mögliche Widersprüche handelt, welche einer abschließenden, manuellen Überprüfung bedürfen, ist die Aussagekraft des QI zur Datenqualität der drei CERTAIN-Register als gut zu erachten. Neben dem Wert der Berechnung des QI eignet sich der QI zudem, um die Häufigkeit von möglichen Widersprüchen und Warnungen zwischen Einrichtungen zu untersuchen, wodurch sich systematische Probleme erkennen lassen können. Plausibilitätsregeln für mögliche Widersprüche können, neben dem Hinweis für die Eingabekräfte, auch als Ansatzpunkt für eine intensivere Überprüfung von einzelnen Datenelementen oder, bei einer Häufung von Warnungen in einem Datensatz zur Überprüfung der gesamten Beobachtungseinheit durch den DQM genutzt werden. Eine hohe Anzahl von durch Plausibilitätsprüfungen festgestellten Widersprüchen kann ein Hinweis auf fehlerhafte Plausibilitätsregeln sein, wenn diese die Wirklichkeit nicht korrekt abbilden.

Unter diesen Gesichtspunkten findet sich die abschließende Bewertung des QI in Tabelle 12.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	+	+	automatische Validierung

Tabelle 12: Bewertung TMF-1005 - Mögliche Widersprüche/Warnungen

TMF-1006 - Werteverteilung

Dieser QI repräsentiert den Anteil der auffälligen Werteverteilungen von numerischen Datenelementen an der Gesamtzahl der überprüften Datenelemente von einzelnen Beobachtungseinheiten oder Datenbeständen.

Anwendbarkeit

Gemäß TMF-Leitlinie sollte der QI nur auf numerische Datenelemente angewendet werden. Je nach Ausrichtung des Registers kann die Breite der Anwendbarkeit somit variieren. Im Falle der klinischen Register CERTAIN und CERTAIN-LI, mit einer hohen Anzahl an numerischen Datenelementen in den Bereichen Labor und Medikation, lässt sich der QI vielfach anwenden. Bei numerischen Datenelementen, die an Einheiten gebunden sind, muss

zwingend eine Normalisierung der Datenwerte auf eine Einheit erfolgen. Der epidemiologisch ausgerichtete Datensatz des KRhOKo-Registers hingegen besteht weitestgehend aus qualitativen Datenelementen und nur eine geringe Anzahl numerischer Datenelemente im Bereich Labor. Das KRhOKo-Register bietet daher geringe Anwendungsmöglichkeiten.

Implementierbarkeit

Die Berechnung der Kenngrößen eines Datenelements ist für alle zu untersuchenden Datenelemente gleich, entsprechend ist die Wiederverwendbarkeit der zu entwickelnden Auswertungsmethoden als sehr gut einzuschätzen. Die Komplexität der Berechnung der Kenngrößen ist dabei als gering einzuschätzen und kann durch das Einbinden geeigneter Programm-bibliotheken weiter vereinfacht werden. Die zu entwickelnden Methoden zur Aggregation der Daten sollten sich bei einer hinreichend generischen Umsetzung für alle Datenelemente wiederverwenden lassen. In den drei CERTAIN-Registern kann dies durch die Programmier-technik Reflexion erreicht werden, die Teil der Programmiersprache Java ist, in der die drei Register geschrieben sind. Es gibt derzeit in den drei Registern nur begrenzte Möglichkeiten, numerische Datenwerte auf eine Einheit zu normalisieren. Zudem kann dieser Vorgang nicht für alle Datenelemente durchgeführt werden, sondern ist nur für einzelne Datenelemente möglich. Dies erschwert die automatisierte Berechnung des QI, die prinzipiell sehr gut umsetzbar erscheint.

Aussagekraft des QI

Die Empfehlung der Leitlinie, die Referenzwerte zu Minima, Maxima, Quantilen, Mittelwerten, Medianen und Standardabweichungen aus dem existierenden Datenbestand zu berechnen, ist unter bestimmten Aspekten problematisch. Dies liegt zum einen darin begründet, dass während der ersten Überprüfung der Daten des Registers zunächst keine vorherigen Daten vorhanden sind, anhand derer die Referenzwerte für den Vergleich berechnet werden könnten. Zum anderen, wenn bereits Daten vorhanden sein sollten und die Berechnung der Referenzwerte nachträglich ausgeführt wird, stellt sich die Frage nach der Vertrauenswürdigkeit der Daten zur Berechnung der Referenzwerte, da diese Daten zwar durch den Datenqualitätssicherungsprozess des Registers überprüft wurden, dieser aber nicht unfehlbar ist und somit auch möglicherweise fehlerhafte Werte in das Register eingeflossen sein könnten. So könnten im Weiteren nicht plausible Daten in die Berechnung der Referenzwerte einbezogen werden und somit die neu berechneten Kenngrößen plausibel erscheinen, auch wenn diese weitere fehlerhafte Daten enthalten.

Die berechneten Werteverteilungen werden zudem durch die einbezogenen Erhebungszeitpunkte beeinflusst, da beispielsweise Kreatininwerte direkt nach der Transplantation höher ausfallen als zu späteren Erhebungszeitpunkten, wenn das Transplantat seine Funktion aufgenommen hat. Dies sollte bei der Berechnung der Verteilungen berücksichtigt werden. Gegebenenfalls kann es erforderlich sein, die Werteverteilungen pro Erhebungszeitpunkt und Datenelement zu berechnen.

Eine Alternative zur Berechnung der Referenzwerte aus dem Datenbestand kann daher, vor allem bei Laborparametern, die Übernahme von Referenzwerten aus der Literatur und die weitere Berechnung des QI anhand dieser Referenzwerten sein. Neben den in der TMF-Leitlinie genannten Einflussfaktoren können Werteverteilungen auch durch Plausibilitätsprüfungen im Rahmen der Eingabe der zugrundeliegenden numerischen Datenelemente beeinflusst werden. Im Hinblick auf die QI TMF-1004 und TMF-1005 können beispielsweise Plausibilitätsprüfungen definiert werden, die physiologisch unmögliche oder unplausible Datenwerte erkennen und diese für eingehendere Untersuchungen markieren.

Die Untersuchung von Werteverteilungen bietet einen guten Überblick über die Datenwerte innerhalb einer Beobachtungseinheit, eines Datenbestands einer Einrichtung oder den Gesamtdatenbestand eines Registers. Der Vergleich der Verteilungen der Datenwerte zwischen den Einrichtungen kann nicht nur der Generierung von Übersichten sondern auch zur Prüfung gegen Datenfälschung dienen. Eine Umsetzung des QI für die beiden Transplantationsregister ist daher anzuraten um Auffälligkeiten festzustellen. Für KRhOKo kann eine Umsetzung aufgrund der geringen Anzahl der numerischen Datenelemente zurückgestellt werden.

Tabelle 13 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
++	+	+	Report

Tabelle 13: Bewertung TMF-1006 - Werteverteilung

TMF-1007 - Bevorzugung bestimmter Endziffern

Der QI „Bevorzugung bestimmter Endziffern“ ist eine Unterkategorie des vorherigen QI und betrachtet die Endziffern numerischer Datenelemente unter der Annahme, dass die Ziffern 0-9 als Endziffern gleich verteilt sein sollten. Häufungen bestimmter Endziffern werden als Zeichen schlechter Datenqualität erachtet und können ein Hinweis auf Auffälligkeiten bei verwendeten Untersuchungsverfahren und Analysemethoden sowie Datenfälschung sein.

Anwendbarkeit

Leitlinie und Quelle des QI geben zu bedenken, dass manche Datenelemente nur mit bestimmten Endziffern angegeben werden und dies kein Hinweis auf Datenfälschung sei (Nonnemacher et al. 2014, S. 26) (O’Kelly 2004). Bei Dosisangaben von Medikamenten ist beispielsweise nicht zu erwarten, dass die Endziffern gleich verteilt sind, eine Berechnung des QI ist in diesen Fällen nicht sinnvoll. Die Anwendung des QI sollte sich daher zunächst auf Labor- und Vitalparameter konzentrieren, wobei auch hier für jedes einzubeziehende Datenelement festzustellen ist, ob unterschiedliche Endziffern auftreten können. Aufgrund der klinischen Ausrichtung der beiden Transplantationsregister ist dieser QI in diesen Registern

im Bereich der Laborparameter gut anwendbar. Da in KRhOKo nur wenige numerische Datenelemente existieren, ist die Anwendbarkeit hier gering.

Implementierbarkeit

Gleich dem vorherigen QI ist für die Berechnung die Einbeziehung aller Datenwerte eines Datenelements erforderlich. Sowohl Umfang wie auch Komplexität der Berechnung sind daher als nicht sehr hoch einzuschätzen. Eine automatisierte Umsetzung mittels generischer Berechnungsmethoden scheint daher sehr gut möglich.

Aussagekraft des QI

Die Umsetzung dieses QI ist vor dem Hintergrund der Erkennung von Datenfälschungen sinnvoll, eine Durchführung der Berechnung des Wertes des QI unter diesem Gesichtspunkt stellt die Glaubwürdigkeit der Daten der Einrichtung in Frage und sollte daher zu einer ganzheitlichen Überprüfung der Daten der Einrichtung führen. Dem Wert des QI kommt daher mehr der Charakter einer Unterstützungsrolle zu, um weitere Indizien für Datenfälschung zu sammeln, da anhand des Ergebnisses des QI allein Aussagen in dieser Richtung nur schwer zu treffen sind. Nützlich kann der Wert des QI sein, um Bias bei numerischen Datenelementen zu erkennen und eingehendere Überprüfungen der Datenqualität anzustoßen. Abschließende Aussagen lassen sich in beiden Anwendungsszenarien allein anhand dieses QI nicht erwarten. Der QI bietet daher eine eingeschränkte Aussagekraft zur Datenqualität für die CERTAIN-Register.

Tabelle 14 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+	+ +	o	Report

Tabelle 14: Bewertung TMF-1007 - Bevorzugung bestimmter Endziffern

TMF-1009 - Werteverteilung der durch Untersucher erfassten Parameter

Eine Unterkategorie des QI TMF-1006 - Werteverteilung bildet der QI „Werteverteilung der durch Untersucher erfassten Parameter“, der die Werteverteilungen von durch Untersucher erfassten Parametern auf Auffälligkeiten untersucht. Ergebnis der Berechnung des QI ist Anteil der Untersucher mit auffälligen Werteverteilungen an der Gesamtzahl der überprüften Untersucher.

Da in keinem der drei CERTAIN-Register dokumentiert wird, wer eine Untersuchung durchgeführt hat und somit auch nicht feststellbar ist, welcher Untersucher welche Parameter erfasst hat, ist der QI nicht anwendbar. Deshalb wird auf eine Bewertung des QI in Tabelle 15 verzichtet.

Aus allgemeinerem Blickwinkel betrachtet konnte dem QI anhand der Ausführungen der TMF-Leitlinie keine bedeutende Aussagekraft zugemessen werden. Ebenfalls wurde von

Seiten des medizinischen Personals der Registerzentralen der Erkenntnisgewinn durch die Berechnung des QI hinterfragt. Eine Hürde zur Anwendung des QI stellen die für die Berechnung benötigten, sehr detaillierten Informationen dar. Unter diesen Gesichtspunkten sollten zunächst andere QI zur Verbesserung der Datenqualität in Betracht gezogen werden.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 15: Bewertung TMF-1009 - Werteverteilung der durch Untersucher erfassten Parameter

TMF-1010 - Werteverteilung der durch Geräte erfassten Parameter

Gleich dem vorherigen QI bildet dieser QI eine Unterkategorie von TMF-1006 „Werteverteilung“ und gleicht dem vorherigen QI, legt aber seinen Fokus statt auf Untersucher auf die verwendeten Geräte. Ergebnis der Berechnung ist der Anteil der Geräte mit auffälliger Werteverteilung an der Gesamtzahl der einbezogenen Geräte.

Informationen zu den genutzten Geräten liegen in keinem der CERTAIN-Register vor, eine Anwendung des QI ist daher nicht möglich. Auf eine abschließende Bewertung des QI in Tabelle 16 wird daher wie bei dem vorherigen QI verzichtet.

In Gesprächen mit dem medizinischen Personal der Registerzentralen konnte keine Punkte erarbeitet werden, die eine zwingende Umsetzung dieses QI rechtfertigen würden. Die Ausführungen der TMF-Leitlinie boten keine Ansatzpunkte um dem QI eine gesteigerte Aussagekraft zu zumessen. Eine Hürde zur Anwendung des QI stellen die für die Berechnung erforderlichen, sehr detaillierten Informationen dar. Im Hinblick auf die Aussagekraft zur Datenqualität der anderen QI der TMF-Leitlinie erscheint es zielführender den QI zurückzustellen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 16: Bewertung TMF-1010 - Werteverteilung der durch Geräte erfassten Parameter

TMF-1011 - Werteverteilung von Befunden

Der Anteil der Befunder mit auffälligen Werteverteilungen an der Gesamtzahl der untersuchten Befunder wird durch diesen QI erfasst, welcher gleich den beiden vorherigen QI eine Unterkategorie von TMF-1006 „Werteverteilung“ bildet.

Eine Anwendung des QI ist für die CERTAIN-Register nicht möglich, da keine Informationen über die befundende Person erfasst werden. Wie bei den beiden vorherigen QI wird daher gleichermaßen auf eine abschließende Bewertung, Tabelle 17, verzichtet.

Auf Basis der Ausführungen der TMF-Leitlinie konnten in Gesprächen mit dem medizinischen Personal der Registerzentralen keine bedeutenden Punkte identifiziert werden, aufgrund derer der QI mit hoher Priorität umgesetzt werden sollte. Dem QI wurde ein informativer Charakter zugestanden, der jedoch hinter der Aussagekraft anderer QI zurückbleibt.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 17: Bewertung TMF-1011 - Werteverteilung von Befunden

TMF-1052 - Werteverteilung von Parametern zwischen Zentren

Der QI „Werteverteilung von Parametern zwischen Zentren“ soll zur Erkennung von Zentren mit auffälligen Werteverteilungen genutzt werden, wobei der gesamte Datenbestand den Rahmen der Betrachtung bildet. Dieser QI bildet eine Unterkategorie von TMF-1006 „Werteverteilung“.

Anwendbarkeit

Wie bei dem übergeordneten QI TMF-1006 hängt die Breite der Anwendung von der Anzahl der numerischen Datenelemente des Datensatzes des jeweiligen Forschungsvorhabens ab. Daher ist der QI daher für CERTAIN und CERTAIN-LI aufgrund der klinischen Ausrichtung der Register sehr gut anwendbar, da viele numerische Datenelemente existieren. Geringere Anwendungsmöglichkeiten bieten sich im KRhOKo-Register, da nur wenige numerische Datenelemente Teil des Datensatzes sind.

Implementierbarkeit

Die Berechnung des QI bezieht nur numerische Datenelemente ein, bietet daher entsprechend sehr gute Wiederverwendbarkeit und kann automatisiert erfolgen. Die zu entwickelnden, generischen Auswertungsmethoden sollten sich für jedes numerische Datenelement wiederverwenden lassen. Der Aufwand zur Berechnung der Verteilungsparameter und die einhergehende Prüfung auf Auffälligkeit bieten geringe Komplexität. Durch die Verwendung der Programmier technik Reflexion der Programmiersprache Java kann die Berechnung des QI generisch in den drei CERTAIN-Registern implementiert werden.

Für die Anwendung des QI auf Datenwert/Einheit-Tupel, beispielsweise Labor- oder Medikationsangaben, müssen alle Datenwerte auf eine Einheit umgerechnet werden. In den drei CERTAIN-Registern existieren hierfür nur vereinzelte Methoden zur Umrechnung, die sich nur auf spezielle Datenelemente anwenden lassen. Für die Berechnung des QI ist daher die Entwicklung einer automatisierten, generischen Einheitenumrechnung zwingend notwendig.

Aussagekraft des QI

Die Aussagekraft des Wertes des QI ist, aufgrund weniger Schwächen, als gut zu erachten. Dies liegt zum einen an der Festlegung des Kriteriums für Auffälligkeit, da für diesen Vorgang kein standardisiertes Vorgehen durch die TMF-Leitlinie vorgegeben wird. Zur Identifikation von Auffälligkeiten sollten statistische Verfahren genutzt werden, beispielsweise durch Vergleich der Werteverteilung eines Zentrums mit der Werteverteilung aller Zentren im Rahmen von statistischen Tests. Zum anderen ist die Aussagekraft des Anteils der Einrichtungen mit auffälligen Werteverteilungen nur begrenzt interessant, da zwar erkennbar wird, welche Parameter bei vielen Zentren mit Problemen behaftet sind, Aussagen über die Datenqualität jeder einzelnen Einrichtung aber eine differenziertere Betrachtung der Datenqualität ermöglichen.

Eine Berechnung des Anteils der auffälligen Werteverteilungen einer Einrichtung an der Gesamtzahl der Werteverteilungen dieser Einrichtung bietet eine deutlich bessere Aussagekraft zur Datenqualität der Einrichtung als die ganzheitliche Betrachtung nur eines Datenelements über alle Einrichtungen, wie in der TMF-Leitlinie beschrieben. Diese Art der Berechnung ermöglicht zudem die bessere Vergleichbarkeit zwischen Einrichtungen und eine bessere Vergleichbarkeit des Verlaufs des Wertes des QI für jede Einrichtung. Auch für diese Form der Berechnung des QI sollten statistische Verfahren und Tests zur Identifikation von auffälligen Verteilungen genutzt werden.

Problematisch für die Berechnung des QI kann die Verzerrung der Vergleichsverteilung durch Datenfehler der anderen Einrichtungen sein.

Tabelle 18 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
++	+	+	Report

Tabelle 18: Bewertung TMF-1052 - Werteverteilung von Parametern zwischen Zentren

TMF-1008 - Untersuchungen am Wochenende

Dieser QI beschreibt den Anteil der Untersuchungen einer Beobachtungseinheit oder eines Datenbestandes, die an Wochenenden oder Feiertagen stattfanden. Untersuchungen an gesetzlichen Feiertagen sind, gemäß Leitlinie, dabei ortsabhängig gesondert zu prüfen, ebenso wie Untersuchungen am Wochenende oder an Feiertagen in größeren Einrichtungen, beispielsweise in der Notaufnahme von Krankenhäusern.

Anwendbarkeit

Der QI lässt sich grundsätzlich auf alle Register anwenden, die Erhebungsdaten von Untersuchungen dokumentieren und bietet damit eine sehr breite Anwendbarkeit. Vor diesem Hintergrund ist der QI auf die drei CERTAIN-Register anwendbar.

Für CERTAIN und CERTAIN-LI besteht hier eine Ungenauigkeit, da die Visiten dieser beiden Register nicht nur ein Datum sondern einen Zeitraum um ein vorausgerechnetes Visitedatum abdecken, in dem die Visite tatsächlich stattgefunden hat. Die dokumentierten Untersuchungen müssen nicht zwangsläufig am vorausgerechneten Erhebungsdatum der Visite stattgefunden haben, sondern können auch davor oder danach stattgefunden haben. Der genaue Tag ist mitunter nicht feststellbar. Für CERTAIN und CERTAIN-LI ist der QI daher nur auf Untersuchungen anwendbar, für welche explizit ein Erhebungsdatum dokumentiert wird. Dies ist beispielsweise bei Biopsien oder Abstoßungen des Transplantats der Fall.

Im KRhOKO-Register wird das Erhebungsdatum des Fragebogens dokumentiert, jedoch muss auch dieses nicht mit dem Datum der durchgeführten Untersuchungen übereinstimmen. Für die durchgeführten Untersuchungen werden im Fragebogen selbst entweder keine Datumsangaben erhoben oder nur im Format Monat/Jahr.

Wie in der TMF-Leitlinie beschrieben, können automatische Plausibilitätsprüfungen entwickelt werden, die im Rahmen der Datenerfassung Anwendung finden. Bei Verhinderung von falschen Datumsangaben während der Datenerfassung kann hierdurch die Anzahl der zu betrachtenden Datenelemente reduziert werden.

Implementierbarkeit

Im Falle eines nationalen Registers ist die Überprüfung der Feiertage im Rahmen von Plausibilitätsprüfungen aufgrund der vom Erhebungsort abhängigen gesetzlichen Feiertage bereits mit erhöhtem Aufwand verbunden, da sich die Feiertage zwischen den Bundesländern unterscheiden können. Für die Berechnung müsste dies entsprechend berücksichtigt werden. Eine, wie in der TMF-Leitlinie angedachte, automatische Plausibilitätsprüfung muss daher in der Lage sein, Feiertage anhand von Regeln vorauszuberechnen und unter Berücksichtigung des Erhebungsortes eine Aussage zu treffen. Bei internationalen Registern werden Aussagen für dieses Szenario durch die steigende Anzahl an unterschiedlichen Feiertagen komplizierter.

Zudem können Einrichtungen teilnehmen, die sich in den Abläufen und organisatorischen Vorgängen unterscheiden, beispielsweise durch gezielte Einbestellung zur Untersuchung am Wochenende. Während bei einem niedergelassenen Arzt eine durchgeführte Untersuchung am Wochenende oder Feiertagen unwahrscheinlich erscheinen mag, kann dies bei einem Maximalversorger die Regel sein. Entsprechend wäre es notwendig, zusätzlich zur Berechnung der Feiertage, ebenso eine Liste der Art der Einrichtung zu führen, um die Plausibilitätsprüfung gezielter anwenden zu können.

Die Anwendung des QI ist daher für CERTAIN und CERTAIN-LI aufgrund des internationalen Rahmens mit sehr hohem Implementierungsaufwand verbunden. Für KRhOKO müssten aufgrund der nationalen Ausrichtung nur deutsche Feiertage einbezogen werden. Aufgrund der unterschiedlichen Größe der teilnehmenden Einrichtungen und den in der

jeweiligen Einrichtungen üblichen Arbeitszeiten und Einbestellungstage wird für alle drei Register die Situation und damit die Berechnung des QI komplizierter.

Die ungenauen Datumsangaben im KRhOKo-Register erschweren eine Berechnung, da nur der Erhebungszeitpunkt des Fragebogens ein genaues Datum darstellt. Für die beiden CERTAIN-Register wäre eine Berechnung aufgrund der Datumsangaben zu einzelnen Untersuchungen möglich. Diese Datumsangaben müssten jedoch zunächst eigens identifiziert werden, wodurch die Berechnung nicht generisch umsetzbar wäre und somit der Aufwand der Berechnung weiter anstiege. Eine automatische Umsetzung des QI wäre unter sehr hohem Aufwand möglich.

Aussagekraft des QI

Durch diese Heterogenität der Rahmenbedingungen zur Anwendung dieses QI stellt sich die Frage nach dem zu erwartenden Nutzen, der den hohen Implementierungsaufwand rechtfertigen muss. Eine Datumsangabe, die auf ein Wochenende oder einen Feiertag fällt, begründet in sich nicht zwingend schlechte Datenqualität, da ein solches Erhebungsdatum alleine nicht die Plausibilität der Daten der Untersuchung in Frage stellt. Die TMF-Leitlinie gibt zu bedenken, dass solche Datumsangaben Hinweise auf Datenfehler oder Datenfälschung darstellen können. Aufgrund der Heterogenität der teilnehmenden Einrichtungen der CERTAIN-Register ist dies praktisch nur schwer zu überprüfen.

Durch den geringen zu erwartenden Nutzen im Hinblick auf den Implementierungsaufwand zur Umsetzung des QI ist von einer Realisierung des QI für die drei CERTAIN-Register abzuraten, Tabelle 19 fasst die abschließende Bewertung zusammen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
-	--	--	automatische Validierung

Tabelle 19: Bewertung TMF-1008 - Untersuchungen am Wochenende

TMF-1012 - Fehlende Module

Module lassen sich unter verschiedenen Gesichtspunkten definieren, beispielsweise als logisch zusammenhängende Einheiten von Datenelementen. Dieser QI betrachtet die Anzahl der fehlenden Module und damit die Vollständigkeit der Module einzelner Beobachtungseinheiten oder eines Datenbestands an der Gesamtzahl der überprüften Module.

Anwendbarkeit

Der QI „Fehlende Module“ bietet eine sehr gute Anwendbarkeit, da sich Datenelemente in den Datensätzen der CERTAIN-Register entsprechend dieses Kriteriums unterteilen lassen. Die Register werden in der Objekt-orientierten Programmiersprache Java entwickelt, Java-

Objekte des Datensatzes können als Module betrachtet werden, wodurch der QI sich sehr gut auf die drei Register anwenden lässt.

Die Anzahl der zu betrachtenden Module kann durch automatische Vollständigkeitsprüfungen zur Eingabezeit reduziert werden, indem das Vorhandensein der Module durch technische Maßnahmen erzwungen wird. Dies ist für Pflichtmodule in allen drei CERTAIN-Registern bereits der Fall, diese Module müssen somit für die Berechnung des QI nicht berücksichtigt werden.

Implementierbarkeit

Programmatische Prüfungen auf das Vorhandensein von Java-Objekten sind von geringer Komplexität und einfach umzusetzen oder bereits Teil existierender Programmbibliotheken. In den drei CERTAIN-Registern können Datenelemente und Module mit der JSR-303 Annotation *@NotNull* versehen werden, die bei einer Prüfung der Validität des beinhaltenden Objekts zu einer Fehlermeldung führt, sollte das annotierte Objekt und damit das Modul fehlen. Wie bereits beschrieben, wird das Vorhandensein mandatorischer Module zur Eingabezeit erzwungen, wodurch sie für die Berechnung nicht berücksichtigt werden müssen.

Modul-Objekte erben in den CERTAIN-Registern stets von der Klasse *RegistryObject*. Für die Berechnung des QI relevante Module können daher durch generische Methoden identifiziert werden, die den Datensatz der drei Register nach Objekten vom Typ *RegistryObject* durchsuchen, die nicht mit *@NotNull* annotiert sind. Die Berechnung ist daher automatisierbar und für alle drei Register gleichermaßen wiederverwendbar.

Aussagekraft des QI

Erst wenn Daten in Modulen überhaupt vorhanden sind, kann ihre Qualität festgestellt werden. Die Feststellung der Vollständigkeit der Module ist daher ein elementarer Punkt der Prüfung der Datenqualität, vor allem unter dem Gesichtspunkt der weiteren Berechnung der anderen QI, welche das Vorhandensein von Datenwerten erfordern. Daher lassen sich anhand dieses QI grundlegende Beobachtungen zur Datenqualität machen.

Durch die Erzwingung des Vorhandenseins von Pflichtmodulen in den CERTAIN-Registern und der dadurch bedingten Reduktion der Berechnung auf optionale Module sinkt die Aussagekraft des QI. Optionale Module können aus einer Vielzahl von Gründen fehlen, beispielsweise einer Verweigerung des Moduls, Vergessen beim Ausfüllen oder Probleme bei der Datenerhebung. Entsprechend wird der Wert des QI niedriger ausfallen, als bei einer gemeinsamen Betrachtung unter Einbeziehung von Pflichtmodulen.

Falls die Berechnung dennoch für alle Module durchgeführt werden soll, empfiehlt sich die getrennte Betrachtung der Module gleich den QI „TMF-1014 Fehlende Werte bei mandatorischen Datenelementen“ und „TMF-1015 Fehlende Werte bei optionalen Datenelementen“, um die Aussagekraft zu verbessern. Diese getrennte Betrachtung von mandatorischen und optionalen Modulen ist analog zur Empfehlung der Literaturquellen zum QI TMF-1014

hinsichtlich der Trennung von mandatorischen und optionalen Datenelementen (Nonnemacher et al. 2014, S. 36).

Wenn die systematische Absenz einzelner Module bei Datensätzen einer Einrichtung festgestellt wird, sollte dies Ansatzpunkt für eine Untersuchung der Ursachen der Absenz sein, speziell im Hinblick auf die in der TMF-Leitlinie genannten Einflussfaktoren Studiendesign und Belastung sowie Motivation der Teilnehmer. Neben den genannten Einflussfaktoren mag auch die Verständlichkeit der Formulare der Datenerhebung eine Rolle spielen.

Dieser QI sollte daher Beachtung bei der Betrachtung der Datenqualität der CERTAIN-Register finden, das abschließende Fazit findet sich in Tabelle 20.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	+ +	+	automatische Validierung

Tabelle 20: Bewertung TMF-1012 - Fehlende Module

TMF-1013 - Fehlende Werte bei Datenelementen

Ähnlich dem vorherigen QI, zur Betrachtung von fehlenden Modulen, berechnet dieser QI den Anteil von fehlenden Datenwerten bei Datenelementen an allen überprüften Datenwerten. Es findet keine Unterscheidung zwischen mandatorischen und optionalen Datenelementen statt. Berechnungsgrundlage können hierbei einzelne Datensätze, Beobachtungseinheiten oder ein Datenbestand sein.

Anwendbarkeit

Datenelemente bilden die einfachste Struktur eines Datensatzes, entsprechend lässt sich der QI in jedem Datensatz vielfach anwenden. In den drei CERTAIN-Registern werden Angaben zu mandatorischen Datenelementen bereits während der Datenerfassung erzwungen. Datenwerte dieser Datenelemente können daher nicht fehlen und müssen somit nicht für die Berechnung berücksichtigt werden.

Die Anwendbarkeit in den drei CERTAIN-Registern ist dennoch auf Basis der optionalen Datenelemente als sehr gut einzuschätzen.

Implementierbarkeit

Die Komplexität der Berechnung und die Überprüfung, ob ein Datenelement einen Datenwert enthält, sind mit geringem Implementierungsaufwand verbunden. Mandatorische Datenelemente sind in den Registern mit den JSR-303 Annotationen *@NotNull* oder *@NotEmpty* annotiert. Durch filtern aller Datenelemente hinsichtlich dieser Annotation können die verbleibenden, optionalen Datenelemente identifiziert und in die Berechnung einbezogen werden. In den CERTAIN-Registern wird der Aufwand durch die Schachtelung von

Folgefragen erhöht, da die geschachtelten Datenelemente nicht zwangsläufig für die Eingabebeurteilung eingeblendet sind und somit gegebenenfalls nicht ausgefüllt werden können oder müssen. Diese Sonderfälle müssen entsprechend berücksichtigt werden. Die Berechnung ist daher nur schwer automatisierbar. Sollte eine Automatisierung der Berechnung möglich sein, ist diese für alle drei Register gleichermaßen wiederverwendbar.

Aussagekraft des QI

Die Aussagekraft des QI ist durch die Vermischung von mandatorischen und optionalen Datenelementen in der Berechnung des QI als mäßig einzuschätzen. Durch die gemeinsame Betrachtung ist der Wert des QI voraussichtlich schlecht, da optionale Datenelemente nicht zwingend ausgefüllt werden müssen, und aus unterschiedlichsten Gründen fehlen können. Daher kann der QI unverhältnismäßig schlechte Werte annehmen, obwohl die Vollständigkeit der mandatorischen Datenelemente sehr hoch ist. Besonderes Augenmerk sollte bei der Berechnung des QI daher auf dem Verhältnis von mandatorischen zu optionalen Datenelementen liegen, da dieses die Aussagekraft des Wertes des QI stark beeinflussen kann.

Es bietet sich daher an, die Berechnung dieses QI gegenüber den beiden QI „Fehlende Werte bei mandatorischen Datenelementen“ und „Fehlende Werte bei optionalen Datenelementen“ zurückzustellen, da diese beiden QI jeweils eine differenzierte Aussage zur Datenqualität erwarten lassen und dieses Vorgehen auch auf Basis der Literaturquellen des QI in der TMF-Leitlinie angemerkt wird (ebd., S. 36).

Dieses Vorgehen ist im Falle der drei CERTAIN-Register vorzuziehen, da mandatorische Datenelemente ohnehin erzwungen werden und dieser QI bei Nichtberücksichtigung dieser Datenelemente somit dem QI „Fehlende Werte bei optionalen Datenelementen“ entspricht. Die zusammenfassende Bewertung des QI befindet sich in Tabelle 21.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	-	o	automatische Validierung

Tabelle 21: Bewertung TMF-1013 - Fehlende Werte bei Datenelementen

TMF-1014 - Fehlende Werte bei mandatorischen Datenelementen

Dieser QI entspricht dem vorherigen QI mit der Beschränkung der Berechnung auf mandatorische Datenelemente. Der Wert des QI beschreibt den Anteil fehlender Datenwerte an der Gesamtzahl der überprüften mandatorischen Datenwerte.

Anwendbarkeit

Die Anwendung des QI ist auf mandatorische Datenelemente beschränkt wodurch die allgemeine Anwendbarkeit vom jeweiligen Vorhaben abhängt. In den Transplantationsregistern

sind mandatorische Datenelemente aufgrund des zwingend erforderlichen minimalen Datensatzes weit verbreitet. Die Anwendbarkeit des QI ist in diesen beiden Registern daher sehr gut. Im KRhOKo-Register sind nur wenige Datenelemente mandatorisch, die Anwendbarkeit ist noch als gut zu erachten.

Während die TMF-Leitlinie für den vorherigen QI als Einflussfaktor die Erzwingung von Angaben mittels eines EDC-Systems erwähnt, bleibt die Erwähnung dieser Möglichkeit der Verbesserung des Wertes des QI hier aus, auch wenn dies weiterhin eine Verbesserung der Datenqualität bewirken könnte. Diese Erzwingung von Angaben findet in den drei CERTAIN-Registern bereits zur Eingabezeit statt. Deshalb ist stets eine komplette Vollständigkeit von mandatorischen Datenelementen und somit kein fehlender Datenwert zu erwarten.

Allein die mandatorischen Datenelemente, die nachträglich in den Datensatz eingebracht wurden und somit in bereits bestehenden Datensätzen nicht ausgefüllt werden konnten, bilden die Ausnahme.

Implementierbarkeit

Die Erfassung des Vorhandenseins einzelner Datenelemente ist technisch einfach umsetzbar. Die Auswahl der mandatorischen Datenelemente muss gegebenenfalls manuell erfolgen, sollten diese nicht entsprechend markiert sein. Mandatorische Datenelemente sind in den Registern in der Regel mit den JSR-303 Annotationen *@NotNull* oder *@NotEmpty* annotiert. Diese Datenelemente können daher einfach aus dem Datensatz extrahiert und für die Berechnung des QI ausgelesen werden. Zusätzlich gibt es Folgefragen, die bei bestimmten vorhergehenden Antworten eingeblendet werden. Diese Fragen können wiederum mandatorisch sein. Anhand der genannten Annotationen allein können daher nicht alle Pflichtfelder extrahiert werden. Diese Sonderfälle müssen entsprechend berücksichtigt werden. Die Berechnung ist daher nur schwer automatisierbar. Sollte eine Automatisierung der Berechnung möglich sein, ist diese für alle drei Register gleichermaßen wiederverwendbar.

Aussagekraft des QI

Während der übergeordnete QI TMF-1013 „Fehlende Werte bei Datenelementen“ auch optionale Datenelemente einbezieht, werden im Rahmen dieses QI nur mandatorische Datenelemente betrachtet. Durch die Fokussierung der Berechnung auf Pflichtfelder, die von größerer Bedeutung für das Forschungsvorhaben sind als optionale Datenelemente, steigt die Aussagekraft des Wertes des QI. Die Aussagekraft des Wertes des QI zur Datenqualität ist als sehr gut einzuschätzen.

Wenn mandatorische Datenelemente fehlen, kann es sinnvoll sein, die Meldung oder den Datensatz zurückzuweisen, bis die Daten korrigiert wurden, siehe QI TMF-1040 „Zurückgewiesene Meldungen“. Dieses Vorgehen würde den Wert des QI positiv beeinflussen.

Da die Vollständigkeit der Daten des Registers eines der Hauptmerkmale guter Datenqualität ist, ist eine Umsetzung dieses QI für die drei CERTAIN-Register grundsätzlich

zu empfehlen. Durch die automatische Datenvalidierung der drei Register zur Eingabezeit werden Angaben für mandatorische Datenelemente bereits erzwungen, eine Berechnung des QI kann daher zurückgestellt werden, da mandatorische Datenelemente grundsätzlich als vollständig zu erachten sind.

Die Berechnung des QI kann für die drei CERTAIN-Register im Hinblick auf Lücken in Altdaten von Relevanz sein.

Tabelle 22 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	-	+ +	automatische Validierung

Tabelle 22: Bewertung TMF-1014 - Fehlende Werte bei mandatorischen Datenelementen

TMF-1015 - Fehlende Werte bei optionalen Datenelementen

Dieser QI entspricht dem vorangegangenen QI „Fehlende Werte bei Datenelementen“ mit der Beschränkung der Betrachtung auf optionale Datenelemente. Der Wert des QI entspricht dem Anteil fehlender Datenwerte bei optionalen Datenelementen an der Gesamtzahl der Datenwerte der überprüften, optionalen Datenelemente.

Anwendbarkeit

In den beiden Transplantationsregistern sind eine Vielzahl von Datenelementen Teil des optionalen, erweiterten Datensatzes und somit Teil der Berechnung dieses QI. Im KRhOKo-Register ist der Großteil der Datenelemente optional, der QI bietet daher in allen drei Registern eine sehr gute Anwendbarkeit.

Implementierbarkeit

Optionale Datenelemente sind in den drei CERTAIN-Registern nicht explizit als solche markiert. Für die Berechnung ist es daher zunächst erforderlich, alle optionalen Datenelemente zu identifizieren. Mandatorische Datenelemente sind in den Registern in der Regel mit den JSR-303 Annotationen *@NotNull* oder *@NotEmpty* annotiert und können somit aus der Gesamtmenge der Datenelemente herausgefiltert werden. Die verbleibenden Datenelemente sind entweder optional oder Folgefragen zu anderen Datenelementen, die nur bei bestimmten Konstellationen eingeblendet werden. Diese Folgefragen können jedoch wiederum optional oder mandatorisch sein. Zur Entscheidung, ob ein Datenelement optional oder mandatorisch ist, muss daher der Kontext des Datenelementes berücksichtigt werden, wodurch die automatische Berechnung stark erschwert wird. Sollte eine Automatisierung der Berechnung möglich sein ist diese für alle drei Register gleichermaßen wiederverwendbar.

Aussagekraft des QI

Optionale Datenelemente nehmen einen geringeren Stellenwert in Forschungsvorhaben ein und können aufgrund einer Vielzahl von Gründen nicht in das Register übernommen worden sein, beispielsweise Verweigerung durch den Patienten oder Verzicht auf Dateneingabe durch das Eingabepersonal wegen Zeitmangels.

Im Patientenbogen des KRhOKO-Registers sind nahezu alle Datenelemente optional. Der QI ist daher besonders für diesen Fragebogen relevant um die Akzeptanz der Fragen durch die Patienten einschätzen zu können. Die beiden Transplantationsregister bestehen zum Großteil aus Pflichtfeldern, optionale Datenelemente beinhalten nur Informationen von geringerem Interesse und werden zudem aufgrund des Umfangs des Datensatzes mehrheitlich nicht durch die Eingabekräfte ausgefüllt. In der TMF-Leitlinie wird die Beschränkung auf mandatorische Datenelemente als Empfehlung von Gaus et al. wiedergegeben (Gaus 2003; Nonnemacher et al. 2014, S. 38).

Die Aussagekraft des QI ist daher im Vergleich mit dem verwandten QI „Fehlende Werte bei mandatorischen Datenelementen“ als geringer einzuschätzen.

Tabelle 23 fasst die abschließende Bewertung des QI zusammen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	-	o	automatische Validierung

Tabelle 23: Bewertung TMF-1015 - Fehlende Werte bei optionalen Datenelementen

TMF-1016 - Datenelemente mit Wert unbekannt o.ä.

Der QI „Datenelemente mit Wert unbekannt o.ä.“ erlaubt Aussagen über den Anteil von Datenwerten wie „unbekannt o.ä.“ bei mandatorischen Datenelementen an der Gesamtzahl der untersuchten Datenwerte.

Anwendbarkeit

In den drei CERTAIN-Registern ist die Anwendbarkeit dieses QI sehr unterschiedlich zu beurteilen. Datenwerte „unbekannt o.ä.“ wurden bereits während der Erstellung der Datensätze der drei Register als Zeichen schlechter Datenqualität erkannt. Bei der Zusammenstellung der Datensätze wurde „unbekannt o.ä.“ daher nur bei einigen Datenelementen zugelassen. Die Anwendbarkeit des QI war daher für alle drei Register initial mäßig.

Durch die Durchführung einer neuen Studie im CERTAIN-LI-Register wurde der Datensatz dieses Registers überarbeitet und auf Wunsch der Projektpartner wurde für nahezu alle Datenelemente „unbekannt o.ä.“ als zulässige Merkmalsausprägung ergänzt. Für das CERTAIN-LI-Register ist die Anwendbarkeit des QI daher als sehr gut zu beurteilen.

Die Identifikation der Datenelemente zur Berechnung des QI kann zur Überprüfung der Datenelemente hinsichtlich der Notwendigkeit des Vorhandenseins der Eingabemöglichkeit von „unbekannt o.ä.“ genutzt werden. Sollte der Datenwert nicht erforderlich sein, würde das Streichen dieser Auswahlmöglichkeit automatisch die Datenqualität verbessern und den Berechnungsaufwand durch die Reduktion der zu betrachtenden Datenelemente verringern.

Implementierbarkeit

Für die Berechnung des QI sind zunächst alle Datenelemente zu identifizieren, welche den Datenwert „unbekannt o.ä.“ annehmen können. Neben einfachen Textfeldern kann „unbekannt o.ä.“ auch in Form von Tupeln in den drei CERTAIN-Registern auftreten, zum Beispiel als Kombination aus Zahl/Text, was die Komplexität der Berechnung steigert. In den drei CERTAIN-Registern ist nicht anhand des Datentyps eines Datenelementes oder Tupels entscheidbar, welche Datenwerte darin gespeichert werden können. Die Auswahl der Datenelemente für die Berechnung muss daher zwingend manuell erfolgen, was den Implementierungsaufwand deutlich steigert. Unter diesen Gesichtspunkten ist die Implementierbarkeit für CERTAIN und KRhOKo aufgrund der geringen Anzahl der Datenelemente noch als neutral einzuschätzen, für CERTAIN-LI ist der Aufwand aufgrund der großen Anzahl der Datenelemente als hoch zu erachten. Die eigentliche Berechnung nach Auswahl der Datenelemente und der Berücksichtigung der jeweiligen Datentypen kann automatisiert erfolgen.

Aussagekraft des QI

Die Aussagekraft des QI zur Datenqualität in den drei CERTAIN-Registern ist als gut zu beurteilen, da ungenaue Datenwerte als Zeichen schlechter Datenqualität zu erachten sind. Die Berechnung dieses QI ist grundsätzlich sinnvoll und kann Hinweise auf problematische Eingaben liefern, besonders wenn Beobachtungseinheiten einer Einrichtung einen sehr hohen Anteil aufweisen, da hier möglicherweise Datenfälschung vorliegen kann. Speziell wenn es sich um Schlüsselwerte handelt, die im Rahmen der Routinebehandlung oder des Studienplans normalerweise immer erfasst werden, sollte dies Grund für eine eingehendere Untersuchung der betroffenen Beobachtungseinheit sein.

Unter diesen Gesichtspunkten findet sich die abschließende Bewertung des QI in Tabelle 24.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+	-	+	Report

Tabelle 24: Bewertung TMF-1016 - Datenelemente mit Wert unbekannt o.ä.

TMF-1017 - Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten

Der Wert dieses QI gibt Auskunft über den Anteil der Datenelemente eines Datenbestands, für die Einträge in allen Beobachtungseinheiten eines Datenbestands vorhanden sind, an der Gesamtzahl der überprüften Datenelemente. Die Berechnung sollte dabei für mandatorische als auch optionale Datenelemente erfolgen und vereint daher Aspekte verschiedener QI, unter anderem „Fehlende Werte bei mandatorischen Datenelementen“ und „Fehlende Werte bei optionalen Datenelementen“

Anwendbarkeit

Durch die Betrachtung aller Datenelemente des Datenbestands ist der QI sehr breit anwendbar. Datenelemente, für die Angaben zur Eingabezeit durch technische Maßnahmen erzwungen werden, müssen nicht berücksichtigt werden.

Implementierbarkeit

Im Zuge der Berechnung müssen alle Datenwerte aller Datenelemente eines Datenbestands auf ihr Vorhandensein überprüft werden. Die Überprüfung auf das Vorhandensein der Datenwerte eines jeden Datenelementes ist für alle Datenelemente als gleich zu erachten und die Berechnung selbst somit stark generalisierbar. Der allgemeine Implementierungsaufwand ist daher als sehr niedrig einzuschätzen. In den CERTAIN-Registern wird der Aufwand durch die Schachtelung von Folgefragen jedoch stark erhöht, da diese Datenelemente nicht zwangsläufig ausgefüllt werden müssen und somit für jedes Datenelement zu prüfen ist, ob dieses auszufüllen gewesen wäre. Diese Sonderfälle müssen entsprechend berücksichtigt werden. Sollte eine automatisierte Berechnung möglich sein kann diese für alle drei Register wiederverwendet werden.

Aussagekraft des QI

Bei Registern mit spezifischer Fragestellung und entsprechend eng gefasstem Studienplan sowie Datensatz kann die Anwendung des QI sinnvoll sein, um systematische Probleme der Datenerhebung erkennen zu können. Bei Registern, die als Hypothesengenerator fungieren und somit einen umfassenderen, zuweilen wachsenden Datensatz mit entsprechenden Lücken in Altdaten besitzen, kann dem QI eine geringere Priorität zugewiesen werden. In diesem Szenario spielt zudem auch die eigentliche Größe des Datensatzes eine Rolle. Wenn beispielsweise alle Eventualitäten in den Bereichen Labor und Medikation zur Eingabe abgedeckt werden sollen, sind entsprechende Lücken und damit niedrigere Werte des QI zu erwarten. Nicht jede Einrichtung führt zwingend alle, im Datensatz vorhandenen, Untersuchungen durch, wodurch der Wert des QI weiter sinkt, sollten keine Datenwerte wie „nicht durchgeführt“ möglich sein.

Im Falle einer ganzheitlichen Berechnung über mandatorische und optionale Datenelemente stellt sich die Frage nach der Aussagekraft des zu erwartenden Ergebnisses, da in dieser

Berechnung Datenelemente mit unterschiedlich zu erwartender Vollständigkeit gemeinsam betrachtet werden. Der Wert der optionalen Datenelemente dürfte erfahrungsgemäß niedriger sein als der Wert der mandatorischen Datenelemente und somit den berechneten Gesamtwert des QI negativ beeinflussen. Es empfiehlt sich daher, bei der Berechnung des QI zwischen mandatorischen und optionalen Datenelementen zu trennen.

Bei mandatorischen Datenelementen kann die Berechnung vernachlässigt werden, wenn fehlende Datenwerte zur Zurückweisung der zu beanstandenden Beobachtungseinheit führen oder das Ausfüllen der Datenelemente zur Eingabezeit erzwungen wird. Bei optionalen Datenelementen kann die Berechnung des QI, aufgrund der niedrigeren Relevanz der Datenelemente selbst, zurückgestellt werden.

Gemäß der Berechnung des QI führt bereits ein fehlender Datenwert eines Datenelementes in nur einer Beobachtungseinheit dazu, dass dieses Datenelement negativ in die Berechnung einfließt. Somit sind bei hoher Anzahl von eingehenden Beobachtungseinheiten und der damit steigenden Wahrscheinlichkeit, dass einzelne Datenwerte fehlen, zunehmend geringere Werte des QI und damit einhergehend die Bewertung des gesamten Datenbestands als von schlechter Datenqualität zu erwarten. Die Aussagekraft des berechneten Wertes des QI ist aufgrund der dargelegten Probleme als mäßig einzuschätzen.

Tabelle 11 fasst die abschließende Bewertung des QI zusammen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	-	-	Report

Tabelle 25: Bewertung TMF-1017 - Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten

TMF-1018 - Ausreißer bei stetigen Datenelementen

Als Ausreißer bei stetigen Datenelementen werden Datenwerte bezeichnet, die einen bestimmten Wertebereich verlassen und somit fragwürdig sind oder unplausibel erscheinen. Dieser QI beschreibt den Anteil dieser Datenwerte an der Gesamtzahl der überprüften Datenwerte. Ausreißer können dabei für einzelne Datensätze, Beobachtungseinheiten oder Datenbestände berechnet werden.

Anwendbarkeit

In den beiden Transplantationsregistern ist dieser QI aufgrund der klinischen Ausrichtung der Register mit Fokus auf klinische Parameter, wie Labor- und Vitalparameter, sehr gut anwendbar. Im KRhOKo-Register ist die Breite der Anwendung geringer, da nur eine geringe Anzahl an numerischen Datenelementen Teil der Erhebung ist. Plausibilitätsprüfungen zur Eingabezeit in Form von Prüfungen von Wertebereichen können die Anzahl der zu

betrachtenden Datenelemente reduzieren, da Ausreißer von vornherein verhindert werden können.

Implementierbarkeit

Die TMF-Leitlinie nennt die Berechnung des 1,5-fachen Interquartilsabstand als Intervall als Beispiel zur Erkennung von Ausreißern. Als Alternative aus der Literatur bietet sich zudem die *median absolute deviation (MAD)* zur Erfassung von Ausreißern an (Leys et al. 2013). Beide Techniken sind robust gegenüber Ausreißern, wobei MAD robuster sei (ebd.). Der Aufwand der Berechnung der Ausreißer lässt sich daher gegebenenfalls durch entsprechende Statistik-Programmbibliotheken reduzieren.

Bei der Auswahl der verwendeten Analysemethoden zur Erkennung von Ausreißern muss zwingend sichergestellt werden, dass die Analysemethoden auf die jeweilige Verteilung der Datenwerte des Datenelements anwendbar sind. Bei unsymmetrischen Verteilungen sind gegebenenfalls Transformationen der Verteilung notwendig, um Analysen wie MAD durchzuführen. Die Berechnung der Ausreißer kann daher zusätzliche Analysen erfordern, komplex und möglicherweise nicht automatisierbar sein. Die Aggregation der Datenwerte zur Berechnung der Verteilung ist mit geringen Aufwand verbunden und steht somit im Kontrast zur Problematik der Ausreißerererkennung.

Aussagekraft des QI

Gemäß TMF-Leitlinie sei zu bedenken, dass nicht jeder Ausreißer auch ein Datenfehler sei, sondern dass der Zustand des Patienten und der Schweregrad der Erkrankung Laborwerte beeinflussen können.

Erkannte Ausreißer sollten daher im Rahmen der QI „Sichere Widersprüche/Fehler“ sowie „Mögliche Fehler/Widersprüche“ betrachtet werden. Je weiter ein Datenwert außerhalb des zu erwartenden Intervalls liegt, um so geringer seine Plausibilität und um so größer die Gewissheit, dass ein möglicher oder sicherer Fehler vorliegt.

Neben der Variante, Ausreißer durch statistische Verfahren zu identifizieren, können die Intervalle für jedes Datenelement auch durch das medizinische Personal anhand klinischer Erfahrung bestimmt werden. Vorteil dieses Ansatzes ist, dass statistische Ausreißer im Rahmen der medizinisch definierten Intervalle nicht zwangsweise unplausibel sein müssen. So können diese Intervalle etwa den Bereich des physiologisch Möglichen abdecken und somit zu weniger falsch positiven Ausreißern führen.

Die Ergebnisse der Berechnung des QI anhand des Interquartilabstands oder MAD sind daher nicht ohne weitergehende Überprüfung der erkannten Ausreißer zu verwenden. Die Aussagekraft des Wertes des QI allein infolgedessen kein zwingendes Merkmal schlechter Datenqualität.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 26.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	o	+	automatische Validierung

Tabelle 26: Bewertung TMF-1018 - Ausreißer bei stetigen Datenelementen

TMF-1019 - Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten

Dieser QI beschreibt den Anteil an Datenwerten, welche die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten an der Gesamtzahl der untersuchten Datenwerte.

Da in keinem der drei CERTAIN-Register explizit dokumentiert wird, welche Verfahren zur Erfassung der Messwerte verwendet wurden, ist der QI nicht anwendbar. Deshalb wird auf eine Bewertung des QI in Tabelle 27 verzichtet.

Aus Sicht des medizinischen Personals der Registerzentralen blieb der Erkenntnisgewinn bei Berechnung dieses QI hinter dem Großteil der anderen QI der TMF-Leitlinie zurück. Auf Basis der Ausführungen der TMF-Leitlinie zu diesem QI konnte ebenfalls keine zwingende Notwendigkeit der Berechnung dieses QI gegenüber anderen QI erkannt werden, zumal die TMF-Leitlinie die angedachte Datennutzung als maßgeblich für die Interpretation des Wertes des QI anführt. Somit ist die Erforderlichkeit der Berechnung dieses QI mit dem jeweiligen Vorhaben verbunden.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 27: Bewertung TMF-1019 - Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten

TMF-1020 - Werte aus Standards

Der Anteil der Datenwerte mit Bezeichnungen aus kontrollierten Vokabularen an allen überprüften Datenwerten wird durch diesen QI berechnet. Gemäß TMF-Leitlinie ist die Berechnung nur für Freitext-Datenelemente oder Datenelemente sinnvoll, die während der Erfassung ergänzt werden können.

Anwendbarkeit

Die TMF-Leitlinie nennt zwei Szenarien in denen die Berechnung des QI sinnvoll ist: für Freitextfelder sowie um Freitext ergänzbare Datenelemente mit bestehenden Auswahllisten. Grundsätzlich sollte während der Definition des Datensatzes des Registers darauf geachtet werden, Auswahllisten für möglichst viele Datenelemente nur mit Merkmalen aus kontrollierten Vokabularen zu befüllen, um die Notwendigkeit von Freitext zu reduzieren oder

gänzlich auf diesen verzichten zu können. Die Begrenzung der Eingabemöglichkeiten gewährleistet die Vergleichbarkeit der Datenwerte und sichert die Auswertbarkeit des Datenelements. Zudem sinkt die Anzahl der im Rahmen der Berechnung des QI zu überprüfenden Datenelemente, wenn von vornherein sichergestellt werden kann, dass in einem Datenelement niemals Datenwerte vorhanden sein könnten, die nicht aus kontrollierten Vokabularen stammen.

Bei ergänzbaren Datenelementen mit kontrollierten Vokabularen für die Auswahlliste stellt sich zum einen die Frage nach der Sinnhaftigkeit der Ergänzung des Datenfeldes, da das gewählte Vokabular vollständig sein sollte. Zum anderen die Frage nach der Qualität des verwendeten Vokabulars, wenn davon ausgegangen werden muss, dass dieses nicht vollständig ist und daher ergänzbar sein muss. Die Betrachtung der Datenelemente im Rahmen der Berechnung dieses QI kann daher auch Ansatzpunkt sein, die Vollständigkeit der Auswahllisten der Datenelemente zu verbessern.

Für Freitextfelder ist generell die Berechnung dieses QI in Frage zu stellen, da für die Berechnung zwingend ein kontrolliertes Vergleichsvokabular vorhanden sein muss, dieses aber nicht als Auswahlliste angeboten wird.

Ergänzbare Datenelemente oder Freitext-Datenelemente existieren in den drei CERTAIN-Registern in überschaubarer Zahl. Die Anwendbarkeit des QI ist daher nicht sehr weitreichend. In den beiden Transplantationsregistern basieren nur wenige Auswahllisten auf offiziellen Standards. Im KRhOKo-Register basieren Teile der Eingabemasken auf standardisierten Instrumenten. Jedoch basiert auch hier der Großteil der Datenelemente nicht auf Standards.

Implementierbarkeit

In den drei CERTAIN-Registern werden bei Datenelementen mit vordefinierten Auswahllisten Datenwerte, die nicht der vordefinierten Auswahlliste entstammen, gesondert durch einen Textbaustein markiert, beispielsweise *Other#Freitext*. Auf Basis dieses Formates können Datenwerte, welche nicht aus Auswahllisten stammen, einfach extrahiert werden. Alle verbleibenden Datenwerte stammen aus kontrollierten Vokabularen und gehen in den Zähler der Berechnung ein. Die eigentliche Berechnung ist von geringer Komplexität und kann automatisiert erfolgen. Lediglich die Auswahl der Datenelemente, die durch den Benutzer ergänzbar sind, muss manuell erfolgen, da sich diese Datenelemente programmatisch nicht von Datenelementen mit nicht-ergänzbaaren Auswahllisten unterscheiden sind.

Aussagekraft des QI

Die Auswertung dieses QI kann auf Probleme bei der Datenerfassung aufmerksam machen, beispielsweise wenn ein hoher Anteil an Freitextangaben bei Datenelementen mit Auswahllisten festgestellt wird. Eine der möglichen Ursachen dieses Umstandes kann das Fehlen von Merkmalen in der Auswahlliste des Datenelements sein. Die Auswahlliste ist dann im Hinblick auf die Realität der Datenerfassung unzureichend.

Um die Aussagekraft des QI zu verbessern sollten die Arbeiten von Standardisierungsinitiativen zur Definition der verwendeten Vokabulare genutzt werden, welche den aktuellen wissenschaftlichen Stand auf dem jeweiligen Arbeitsgebiet widerspiegeln. Die Berechnung des QI kann gleichermaßen Anlass bieten, Datenelemente zu identifizieren, die zwar keinen Freitext zulassen, deren Auswahllisten allerdings auch nicht auf standardisierten Vokabularen beruhen und hier nachzubessern.

Bei Freitextfeldern ohne Vergleichsvokabular wird der berechnete Wert immer null sein. Dies steht in diesem Fall jedoch nicht für schlechte Datenqualität, sondern eine schlechte Definition des Datensatzes, da jeder Datenwert des Datenelements von vornherein als von schlechter Datenqualität erachtet werden muss. Von der Berechnung des QI für Datenelemente ohne Auswahllisten ist daher abzuraten, da diese Berechnung keinerlei Aussagekraft über die Datenwerte selbst besitzt.

Tabelle 28 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
o	o	+	Registerplanung

Tabelle 28: Bewertung TMF-1020 - Werte aus Standards

TMF-1021 - Unerlaubte Werte bei qualitativen Datenelementen

Die Berechnung dieses QI beschreibt den Anteil der unerlaubten Datenwerte eines qualitativen Datenelements an der Gesamtzahl der überprüften Datenwerte der Datenelemente eines Datensatzes, einer Beobachtungseinheit oder eines Datenbestands.

Anwendbarkeit

Gemäß den Anmerkungen der TMF-Leitlinie ist die Berechnung des QI nur sinnvoll, wenn die Angabe von unerlaubten Werten durch technische Maßnahmen nicht ausgeschlossen werden kann. Die Nutzung von Auswahllisten für Datenelemente während der Erhebung kann die Anzahl der zu untersuchenden Datenelemente daher reduzieren, da nur ergänzbare Datenelemente oder Freitextfelder in die Berechnung einbezogen werden müssen. Bei einer technischen Beschränkung der Eingabemöglichkeiten ist die ganzheitliche Definition der Auswahlmöglichkeiten sehr wichtig, um das Eingabepersonal nicht zu ungenauen oder falschen Angaben zu zwingen.

In allen drei CERTAIN-Registern werden technische Maßnahmen zur Verhinderung von unerlaubten Datenwerten bereits beim Großteil der Datenelemente eingesetzt. Dies resultiert in einer geringen Anzahl an zu betrachtenden Datenelementen, somit besitzt der QI nur eine geringe Breite der Anwendbarkeit. In KRhOKo wird diese weiter reduziert, da in diesem Register der Großteil der Eingaben über die Auswahl aus Checkboxen erfolgt.

Implementierbarkeit

Für die Erkennung der unerlaubten Datenwerte müssen gemäß TMF-Leitlinie zunächst die erlaubten Datenwerte für das Datenelement festgelegt werden. Im Hinblick auf die beiden möglichen Typen der Datenelemente ist die Überprüfung von Datenelementen mit Auswahllisten einfacher, da die Liste der erlaubten Datenwerte mit der Auswahlliste des Datenelementes übereinstimmen sollte. Sollten im Rahmen der Berechnung weitere Datenwerte als erlaubte Datenwerte in die Berechnung eingehen, sind diese der Auswahlliste des Datenelements hinzuzufügen, um die Benutzereingaben zu standardisieren und die künftige Berechnung zu vereinfachen.

Gleichermaßen lässt sich dieser Vorgang auch auf Freitextfelder anwenden. Die Liste der festgelegten, erlaubten Datenwerte wird dem Datenelement als Auswahlliste beigeordnet um künftige Eingaben zu vereinfachen. Diese Auswahlliste bildet anschließend die Basis der künftigen Berechnung.

Die Erkennung von unerlaubten Datenwerten anhand der Listen ist einfach umzusetzen. Die Berechnung selbst ist damit von geringer Komplexität. Problematisch für die CERTAIN-Register ist an dieser Stelle die Zusammenstellung der erlaubten Datenwerte, da die Auswahllisten der erlaubten Werte nicht im Datenmodell sondern nur in der grafischen Oberfläche hinterlegt sind und den Datenelementen aufwändig zugeordnet werden müssten. Dies erschwert die Berechnung des QI erheblich.

Aussagekraft des QI

Wie in der TMF-Leitlinie angemerkt, wird hierbei nur die formale Korrektheit im Hinblick auf die Zugehörigkeit der Datenwerte zur Menge der möglichen Merkmalsausprägungen des Datenelements betrachtet. Aussagen über die inhaltliche Korrektheit des Datenwertes selbst sind keine zu treffen.

Bei einer Häufung von unerlaubten Datenwerten bei einem Datenelement kann dies ein Ansatzpunkt für die Überprüfung der erlaubten Datenwerte dieses Datenelements sein. Die definierten Merkmalsausprägungen könnten unvollständig sein und die festgestellten unerlaubten Datenwerte gegebenenfalls eine sinnvolle Ergänzung der Merkmalsausprägungen darstellen.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 29.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
- -	-	+	Registerplanung

Tabelle 29: Bewertung TMF-1021 - Unerlaubte Werte bei qualitativen Datenelementen

TMF-1022 - Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings

Bei qualitativen Datenelementen kann es möglich sein, das Fehlen des Datenwertes durch eine Kodierung anzugeben, dieser QI misst den Anteil der unerlaubten Kodierungen an der Anzahl der überprüften Datenwerte. Die Berechnung für einzelne Datensätze, Beobachtungseinheiten oder Datenbestände sollte gemäß TMF-Leitlinie nur erfolgen, wenn Missings auftreten und kodiert werden können.

Anwendbarkeit

Die Verhinderung von unerlaubten Datenwerten zur Eingabezeit durch Auswahllisten oder automatische Kodierung kann die Anzahl der zu betrachtenden Datenelemente reduzieren. In den drei CERTAIN-Registern können unerlaubte Datenwerte zur Kodierung von Missings bei Datenelementen nur auftreten, wenn „Missing“ Teil der Auswahlliste des Datenelementes ist, jedoch Freitext genutzt wird um dies zu dokumentieren. Dieses Szenario tritt in allen drei Registern in sehr geringer Zahl auf. Alle anderen Angaben von „Missing“ erfolgen durch Auswahlmöglichkeiten in der Oberfläche, welche „Missing“ automatisch kodieren. Die Anwendbarkeit des QI ist daher gering.

Implementierbarkeit

Für die Berechnung des QI im Rahmen der drei CERTAIN-Register müssen relevante Datenelemente manuell identifiziert werden, da sich technisch nicht feststellen lässt, welche Auswahl- und Eingabemöglichkeiten ein Datenelement bietet. Ebenso muss manuell festgelegt werden, welche der Freitext-Datenwerte für „Missing“ stehen. Die Suche nach unerlaubten Werten kann anschließend automatisiert erfolgen. Die Berechnung kann daher nur in Teilen automatisiert erfolgen und ist deshalb mit hohem Aufwand verbunden.

Aussagekraft des QI

Der QI gibt Auskunft über die Auswert- und Vergleichbarkeit der Daten. Wenn zu viele unerlaubte Werte zur Kodierung genutzt werden, müssen die Daten gegebenenfalls manuell bereinigt werden. Entsprechend ist eine Berechnung als sinnvoll zu erachten, sodenn sich hierdurch problematische Datenelemente erkennen lassen. Die Zahl der zu untersuchenden Datenelemente lässt sich sinnvoll einschränken, indem die Eingabemasken entsprechend restriktiv gestaltet werden, so dass keine unerlaubten Werte zur Kodierung von Missings zugelassen werden. Als problematisch erkannte Datenelemente sollten gleichermaßen nachgebessert werden. Hierdurch wird eine direkte Verbesserung der Datenqualität erreicht, die Aussagekraft des QI steigt durch die Reduktion der zu berücksichtigenden Datenelemente. Bei ganzheitlicher Umsetzung von technischen Maßnahmen zur Verhinderung von unerlaubten Datenwerten kann die Berechnung des QI vernachlässigt werden, da keine Fehler zu erwarten sind.

Tabelle 30 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
- -	-	+	Registerplanung

Tabelle 30: Bewertung TMF-1022 - Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings

TMF-1023 - Unerlaubte Werte zur Kodierung von fehlenden Modulen

Gleich TMF-1022 betrachtet dieser QI unerlaubte Datenwerte zur Kodierung, im Gegensatz zum vorherigen QI jedoch im Hinblick auf Module statt auf Datenelemente. Entsprechend kann der QI ebenfalls für einzelne Datensätze, Beobachtungseinheiten oder den Datenbestand berechnet werden, wobei der berechnete Wert des QI den Anteil von unerlaubten Datenwerten an der Gesamtzahl der überprüften Datenwerte wiedergibt.

Anwendbarkeit

Die TMF-Leitlinie gibt beim vorherigen QI zu bedenken, dass eine Anwendung des QI nur Sinn macht, wenn das Auftreten von fehlenden Datenwerten technisch möglich und eine Kodierung vorgesehen sei (Nonnemacher et al. 2014, S. 48). Für diesen QI bleibt diese Anmerkung aus, auch wenn sich die beiden QI nur hinsichtlich des Betrachtungsrahmens von Datenelementen zu Modulen unterscheiden.

Datenwerte zur Signalisierung von fehlenden Modulen werden in den drei CERTAIN-Registern automatisch kodiert, das Auftreten von unerlaubten Datenwerten zur Kodierung von fehlenden Modulen ist daher nicht möglich. Einzig eine Überprüfung des Datenbestands hinsichtlich der Kodierungen von fehlenden Modulen bei Altdaten kann sinnvoll sein.

Implementierbarkeit

Die Kodierung des Status der Module wird in den CERTAIN-Registern in einem speziellen Datenelement des Moduls gespeichert und ist somit einfach auffindbar. Zwar unterstützen viele Module ein Datenelement, um den Status des Moduls festzuhalten, nicht alle Module können jedoch über dieses Datenelement als fehlend markiert werden. Die möglichen Datenwerte des Status sind nicht standardisiert, somit muss für jedes Modul einzeln betrachtet werden, ob die Kodierung des Fehlens dieses Moduls vorgesehen ist.

Durch die Notwendigkeit der manuellen Auswahl der zu berücksichtigenden Module steigt der Aufwand zur Berechnung des QI. Nach Auswahl der zu berücksichtigenden Module kann die Erfassung der unerlaubten Datenwerte und die Berechnung automatisiert erfolgen.

Aussagekraft des QI

Gleich dem vorherigen QI gibt dieser QI Auskunft über die Auswert- und Vergleichbarkeit der Daten. Neben der Verbesserung des Wertes des QI durch die in der TMF-Leitlinie angeratene automatische Belegung von fehlenden Modulen durch entsprechende Codes

kann die Aussagekraft des QI verbessert werden indem nur Module berücksichtigt werden, welche fehlen können und deren fehlen kodiert werden kann.

Die Auswahl der Module für die Berechnung kann zu einer Verbesserung der Datenqualität durch eine Ergänzung der Module um eine automatische Kodierung fehlender Module genutzt werden. Bei konsequenter Umsetzung der technischen Maßnahmen zur Kodierung kann die Berechnung des QI selbst vernachlässigt werden.

Tabelle 31 fasst die abschließende Bewertung des QI zusammen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
- -	o	+	Registerplanung

Tabelle 31: Bewertung TMF-1023 - Unerlaubte Werte zur Kodierung von fehlenden Modulen

TMF-1024 - Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen

Dieser QI ist mit dem vorherigen QI verwandt, legt seinen Fokus statt auf Module jedoch auf Datenelemente. Ergebnis der Berechnung des QI ist der Anteil der unerlaubten Datenwerte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen an der Gesamtzahl der überprüften Datenwerte.

In keinem der CERTAIN-Register ist es möglich, Unter- oder Überschreitungen von Messbarkeitsgrenzen durch unerlaubte Datenwerte zu dokumentieren, eine Anwendung des QI ist daher nicht möglich. Auf eine abschließende Bewertung des QI in Tabelle 32 wird daher wie bei dem vorherigen QI verzichtet.

Allgemein betrachtet ist diesem QI dieselbe Bedeutung zu zumessen wie den anderen QI, anhand deren Aussagen zum Auftreten von unerlaubten Werten gemacht werden können. Der QI sollte daher mit ähnlicher Priorität betrachtet werden wie beispielsweise die QI TMF-1022 - „Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings“ oder TMF-1023 „Unerlaubte Werte zur Kodierung von fehlenden Modulen“.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie

Tabelle 32: Bewertung TMF-1024 - Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen

TMF-1025 - Datenelemente mit unspezifischen Werten

Der Anteil der Datenelemente mit unspezifischer Angabe an der Gesamtzahl der überprüften Datenelemente wird durch den Wert dieses QI beschrieben. Anhand dieses Wertes

lassen sich Aussagen zur Informationsdichte treffen. Betrachtet werden für diese Berechnung einzelne Beobachtungseinheiten oder ein Datenbestand.

Anwendbarkeit

In den drei CERTAIN-Registern existieren sehr wenige Datenelementen mit Auswahllisten, die unspezifische Datenwerte enthalten. Dies betrifft in allen drei Registern die Auswahllisten von Diagnosen. Darüber hinaus besteht bei Datenelementen die Freitext zulassen die Möglichkeit unspezifische Datenwerte anzugeben. In allen drei Registern besteht daher eine Anwendbarkeit dieses QI. Wie am Beispiel der TMF-Leitlinie im Hinblick auf ICD-10 Codes deutlich wird können unspezifische Datenwerte nicht komplett ausgeschlossen werden. Gegebenenfalls können unspezifische Datenwerte mit Warnungen belegt werden.

Implementierbarkeit

Durch die für jedes Datenelement erforderliche manuelle Auswahl der Datenwerte die als unspezifisch zu erachten sind, ist ein hoher Aufwand bei der Berechnung des QI zu erwarten. Dies ist besonders im Hinblick auf Datenelemente mit Freitext der Fall, da für jeden Datenwert entschieden werden muss, ob dieser unspezifisch ist. Die Berechnung ist daher kaum zu automatisieren.

Aussagekraft des QI

Die Aussagekraft des QI ist als gut zu erachten, da unspezifische Datenwerte auf eine ungenaue Datenerhebung schließen lassen und sowohl Auswertbarkeit als auch Vergleichbarkeit der Daten negativ beeinflussen können. Die Bedeutung des Auftretens von unspezifischen Datenwerte hängt dabei auch von der Wichtigkeit des zugehörigen Datenelementes für die Forschungsfragen ab. Eine unspezifische Hauptdiagnose wiegt im Falle der drei CERTAIN-Register beispielsweise schwerer als eine unspezifische Angabe zum Bildungsniveau des Patienten. Unspezifische Datenwerte bei mandatorischen Datenelementen ist eine größere Bedeutung zu zumessen als unspezifischen Datenwerte in optionalen Datenelementen. Die zusammenfassende Bewertung des QI befindet sich in Tabelle 33.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
o	- -	+	Report

Tabelle 33: Bewertung TMF-1025 - Datenelemente mit unspezifischen Werten

TMF-1026 - Beobachtungseinheiten mit unbekanntem Primärtumor

Dieser QI beschreibt den Anteil der Beobachtungseinheiten mit unbekanntem Primärtumor einer einzelnen Beobachtungseinheit oder eines Datenbestandes an allen überprüften Beobachtungseinheiten.

Anwendbarkeit

In den beiden Transplantationsregistern gibt es im Rahmen von Angaben zu Komplikationen die Möglichkeit, Tumorerkrankungen zu dokumentieren. Hierbei wird jedoch nicht erhoben, ob es sich dabei um den Primärtumor handelt. In KRhOKo werden keine Tumorerkrankungen erfasst.

Der QI ist daher für die drei Register nicht anwendbar. Einzig bei einer Änderung des Betrachtungsrahmens hin zu unbekanntem Primärerkrankungen wäre der QI anwendbar. Diese Information ist jedoch wiederum in allen drei Registern ein Pflichtfeld, so dass keine Beobachtungseinheiten mit fehlender Primärerkrankung auftreten können. Im CERTAIN-Register existieren in der Auswahlliste des Datenelementes „Hauptdiagnosegruppe“ drei Einträge, deren Subgruppe als „unbekannt“ angegeben werden kann. Diese Angaben sind entsprechend zu berücksichtigen.

Ausnahme bilden hier Altdatensätze im CERTAIN-Register, zu deren Erstellungszeitpunkt die primäre Krankheitsgruppe noch kein Pflichtfeld war. Eine Analyse der CERTAIN-Registerpopulation zeigte, dass nur eine Beobachtungseinheit des gesamten Datenbestands keine eingetragene Primärerkrankung besaß. Nachträglich zum Datenmodell hinzugefügt wurde die primäre Krankheitsdiagnose, die daher nicht in allen Beobachtungseinheiten ausgefüllt ist und somit bessere Anwendbarkeit bietet.

Implementierbarkeit

Die Primärerkrankung wird in allen drei Registern in einem Datenelement vermerkt und kann entsprechend einfach abgerufen und überprüft werden. Die Sonderfälle der Hauptdiagnosegruppe des CERTAIN-Registers können hierbei aufgrund der geringeren Anzahl dieser Datenwerte einfach berücksichtigt werden. Die Berechnung des QI ist daher automatisierbar und einfach umzusetzen.

Aussagekraft des QI

Die Abwesenheit von Informationen zur Primärerkrankung ist ein Datenfehler. Im Falle des KRhOKo-Registers gehen Patienten aufgrund ihrer Primärerkrankung in das Register ein. Eine unbekannt Primärerkrankung ist daher ein deutliches Zeichen schlechter Datenqualität. In die beiden Transplantationsregister gehen Patienten aufgrund der durchgeführten Therapie in Form der Organtransplantation ein. Aus Sicht des medizinischen Personals des CERTAIN-Registers ist eine fehlende oder unbekannt Primärerkrankung dennoch ein deutliches Zeichen schlechter Datenqualität. Das Vorhandensein der Primärerkrankung ist wichtig für die korrekte Schätzung von Inzidenzen und Prävalenzen und die Aussagekraft der Auswertungen der Registerdaten. Die Berechnung des QI ist vor dem Hintergrund der Plausibilität der Daten als sinnvoll zu erachten da besonders bei krankheitsspezifischen Registern nicht ersichtlich sein mag, wieso die Beobachtungseinheit in das Register eingegangen ist.

Der QI besitzt daher eine sehr gute Aussagekraft zur Datenqualität. Die Berechnung ist aber aufgrund der existierenden technischen Maßnahmen zur Sicherstellung des Vorhandenseins der Primärerkrankung für KRhOKo und CERTAIN-LI nicht zwingend notwendig. Die Berechnung für das CERTAIN-Register hingegen ist aufgrund der Möglichkeit der Angabe einer unbekanntem Diagnosegruppe sehr zu empfehlen.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 34.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
-	+ +	+ +	automatische Validierung

Tabelle 34: Bewertung TMF-1026 - Beobachtungseinheiten mit unbekanntem Primärtumor

TMF-1027 - Nachweis bekannter Korrelationen

Der QI „Nachweis bekannter Korrelationen“ untersucht Datenelemente eines Datenbestands im Hinblick auf das Vorhandensein, der zwischen den Datenelementen erwarteten, statistischen Beziehung. Der Wert des QI setzt sich aus dem Anteil der erwarteten, jedoch nicht nachweisbaren Korrelationen an der Gesamtzahl der überprüften Korrelationen zusammen.

Anwendbarkeit

Im Hinblick auf Korrelationen in den Datensätzen der drei Register wurde das medizinische Personal der Register befragt. Es wurde in diesen Gesprächen eine geringe Anzahl an bekannten Korrelationen benannt. Die Anwendbarkeit wurde daher als ebenso groß beurteilt. Die Berechnung der Korrelationen kann automatisiert erfolgen und während der Eingabezeit durchgeführt werden. In Form von Plausibilitätsprüfungen können die fehlerhaften Datenwerte der korrelierten Datenelemente bemängelt und verhindert werden.

Implementierbarkeit

Auf der Basis der anzuwendenden statistischen Verfahren zum Nachweis der Korrelationen sind diese automatisiert berechenbar. Gleiches gilt somit für die Berechnung des QI. Für jede bekannte Korrelation muss eine eigene Berechnung umgesetzt werden. Da jedoch nur wenige Korrelationen identifiziert werden konnten, ist der Implementierungsaufwand als neutral einzuschätzen.

Aussagekraft des QI

Die Überprüfung hinsichtlich des Nachweises von bewiesenen, statistischen Korrelationen ist wichtig für die Einschätzung der Datenqualität, da eine nicht nachweisbare Korrelation Hinweis auf einen Datenfehler sein kann.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 35.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
-	o	+	automatische Validierung

Tabelle 35: Bewertung TMF-1027 - Nachweis bekannter Korrelationen

TMF-1050 - Umfang der Metadaten bei Untersuchungen

Dieser QI betrachtet die Prozessqualität der Übermittlung von Metadaten im Rahmen von technischen Untersuchungen. Ergebnis der Berechnung des QI ist der Anteil der vorhandenen Parameter an der Gesamtzahl der erwarteten Parameter.

In keinem der CERTAIN-Register werden Metadaten von technischen Untersuchungen erfasst, die Anwendung des QI ist daher nicht möglich. Auf eine abschließende Bewertung des QI in Tabelle 36 wird daher verzichtet.

In Diskussion mit dem medizinischen Personal der Registerzentralen wurde diesem QI keine erhöhte Bedeutung zugemessen. Auf der Basis der Ausführungen der TMF-Leitlinie konnte ebenfalls kein hoher Stellenwert des QI erkannt werden. Aufgrund des eng gefassten Anwendungsszenarios lassen andere QI mit breiterer Anwendbarkeit bessere Erkenntnisse zur Datenqualität erwarten.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 36: Bewertung TMF-1050 - Umfang der Metadaten bei Untersuchungen

6.3.2 Qualitätsindikatoren der Ebene Organisation

Im Folgenden werden die Eigenschaften der QI der Ebene Organisation für die drei CERTAIN-Register dargelegt.

TMF-1028 - Aktualität der gespeicherten Daten

Dieser QI beschreibt den Anteil der Datenwerte mit inakzeptablen Alter an der Gesamtzahl der betrachteten Datenwerte aus einzelnen Datensätzen, Beobachtungseinheiten oder einem Datenbestand. Gemäß TMF-Leitlinie ist die Berechnung nur für veränderliche Datenelemente sinnvoll.

Anwendbarkeit

Die Datensätze der drei CERTAIN-Register wurden mit dem Ziel definiert, Mehrfachhebungen von unveränderlichen Datenelementen zu vermeiden. Entsprechend sind nahezu alle Datenelemente der drei Register in die Berechnung des QI einzubeziehen.

Implementierbarkeit

Die TMF-Leitlinie beschreibt zwei Verfahren zur Berechnung des Wertes des QI. Zum einen durch Bildung der Differenz zwischen Eingang der Daten in das Register und dem Erhebungszeitpunkt und Abgleich dieser Differenz mit der akzeptablen Differenz. Der Wert des QI berechnet sich aus dem Anteil der Daten mit inakzeptabler Differenz an der Gesamtzahl der überprüften Datenwerte. In allen drei CERTAIN-Registern wird der Erhebungszeitpunkt des Datensatzes dokumentiert sowie dessen Eingang in das Register. Die Berechnung des QI ist in dieser Variante einfach automatisiert umzusetzen. Für die beiden Transplantationsregister besteht hier eine Unschärfe, da der Erhebungszeitpunkt vorausberechnet wird, die Visite aber in einem Zeitraum um diesen Erhebungszeitpunkt stattgefunden haben kann. Die Daten können daher tatsächlich älter oder jünger sein als anhand des Erhebungszeitpunktes ersichtlich.

Die zweite Variante legt das Augenmerk auf die Anzahl von eingegangenen Meldungen in einem Zeitintervall an der Anzahl der erwarteten Meldungen für dieses Zeitintervall. Letztere sollen aus Prävalenzen und Inzidenzen abgeleitet werden. Da diese Zahlen aus der Literatur entnommen werden müssen, ist in dieser Variante mit erhöhtem Aufwand zu rechnen, zumal die Zahlen sich für jedes Zeitintervall, beispielsweise jedes Jahr, unterscheiden mögen. Die TMF-Leitlinie verweist an dieser Stelle auf die Ähnlichkeit dieser Berechnungsformel zur Definition des QI „Vollständigkeit“.

Für das CERTAIN-Register kann diese Form der Berechnung durch den Datenaustausch mit Eurotransplant vereinfacht berechnet werden. CERTAIN erhält von Eurotransplant kontinuierlich Informationen über alle neuen Transplantationen von ET-Einrichtungen. Somit kann einfach und automatisiert für ET-Länder überprüft werden, ob alle Transplantationen einer Einrichtung in einem Zeitraum in das Register eingegangen sind.

Aussagekraft des QI

Neben den beiden genannten Berechnungsmöglichkeiten der TMF-Leitlinie ist auch die Aktualität des letzten Follow-ups von Beobachtungseinheiten ein wichtiger Betrachtungspunkt der Aktualität der Registerdaten und somit der Datenqualität, besonders bei iterativen Erhebungszeitpunkten. Ansatzpunkt hierfür ist die Frage nach dem Alter der gespeicherten Daten (Naumann und Rolker 2000; Wang et al. 1993). Beobachtungseinheiten, welche kein aktuelles Follow-up besitzen, können auf eine mangelnde Dateneingabe oder Betreuung durch die teilnehmende Einrichtung schließen lassen. Die Aktualität der Registerdaten ist daher auch als Gradmesser für die Aktivität der teilnehmenden Einrichtungen im Register zu verstehen und kann Anlass geben, Kontakt mit nachlässigen Einrichtungen aufzunehmen. Zudem ist mangelnde Aktualität schlecht um aktuelle Fragestellungen des medizinischen Fachgebiets des Registers beantworten zu können, besonders im Hinblick auf die Nutzung eines Registers als Hypothesengenerator oder, wie in der TMF-Leitlinie genannt,

im Rahmen von Feasibilitystudien. Die Aussagekraft von Auswertungen der Registerdaten hängt von der Aktualität der Daten ab, besonders hinsichtlich des klinischen Outcomes. Die Umsetzung des QI ist bei prospektiver Datenerhebung als sinnvoll zu erachten um Problem bei der Datenerhebung feststellen zu können. Bei retrospektiven Registern ist die Aussagekraft als nicht sehr hoch einzuschätzen, da die Daten möglicherweise erst Jahre nach dem Erhebungszeitpunkt in das Register eingehen oder schon erhoben wurden, als das Register noch nicht existierte. Die beschriebene alternative Betrachtung der Aktualität des letzten Follow-ups ist als grundsätzlich sehr sinnvoll zu erachten.

Tabelle 37 fasst die abschließende Bewertung des QI zusammen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	+ +	+ +	Report

Tabelle 37: Bewertung TMF-1028 - Aktualität der gespeicherten Daten

TMF-1029 - Dubletten im Datenbestand

„Dubletten im Datenbestand“ bezeichnen mindestens zwei Informationseinheiten, welche dasselbe Subjekt oder Objekt beschreiben, und dessen Identität nachgewiesen oder als hinreichend sicher angenommen werden kann. Der Wert des QI ist der Anteil der gefundenen Dubletten an der Anzahl der überprüften Sets der Datenelemente.

Anwendbarkeit

Dubletten sind prinzipiell in allen drei CERTAIN-Registern möglich, entsprechend ist der QI in allen drei Registern anwendbar. Eine Überprüfung auf das Vorhandensein eines Patienten im Register anhand seiner identifizierenden Daten oder eines Sets dieser Daten erfolgt bisher nicht beim Anlegen von neuen Patienten.

Um den Indikator anwenden zu können muss zwingend ein Algorithmus entwickelt werden, der es ermöglicht, Dubletten zu identifizieren. Dubletten könnten durch eine Überprüfung anhand des Algorithmus bereits während des Anlegens verhindert werden und somit die Berechnung des QI überflüssig machen.

Implementierbarkeit

Gemäß Berechnungsformel des QI werden Dubletten durch die Bildung von Werte-Sets aus den Schlüsselementen der Daten der Patienten und deren Vergleich gegen Sets des Datenbestands erkannt. Teil des Sets können sowohl identifizierende als auch medizinische Daten sein. In den drei CERTAIN-Registern ist die Kombination dieser Daten nicht möglich, da identifizierende und medizinische Daten stets strikt getrennt sind. Die Sets zur Erkennung von Dubletten können daher nur für den jeweiligen Datensatz aus medizinischen oder identifizierenden Daten erfolgen. Die Berechnung des QI für die identifizierenden Daten ist zudem aufgrund des Datenschutzkonzeptes nicht möglich, da die identifizierenden Daten

bei einem Treuhänder vorgehalten werden, für das Registerpersonal nicht zugänglich sind und somit nicht für die Berechnung verwendet werden können.

Für die beiden Transplantationsregister kann die Dublettenprüfung anhand des Transplantationsdatums erfolgen, da zwei Transplantationen an einem Tag in der selben Einrichtung unwahrscheinlich sind. Weiterer Ansatzpunkt für die Bildung der Sets sind Spender- und Empfängerdaten.

Die Prüfung auf Dubletten anhand von medizinischen Daten ist in KRhOKo deutlich schwieriger, da Datumsangaben zumeist nur aus Monat und Jahr bestehen und somit häufige Überschneidungen zu erwarten sind. Zudem werden in KRhOKo weit mehr Patienten erhoben, was Überschneidungen wahrscheinlicher macht.

Die eigentliche Berechnung inklusive der Bildung der Sets und deren Abgleich gegen den Datenbestand ist einfach und automatisiert umzusetzen.

Aussagekraft des QI

Dubletten im Datenbestand haben erheblichen Einfluss auf die Aussagekraft von Auswertungen sowie deren Durchführbarkeit und sind daher unbedingt zu vermeiden. Der Wert des QI ist daher von großer Bedeutung für die Datenqualität der drei CERTAIN-Register. Tabelle 38 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
++	+	++	Registerplanung

Tabelle 38: Bewertung TMF-1029 - Dubletten im Datenbestand

TMF-1030 - Rekrutierungsrate

Der QI „Rekrutierungsrate“ beschreibt den Anteil der rekrutierten Patienten in einem Zeitintervall an der Gesamtzahl der Patienten, welche in diesem Zeitintervall nach Prüfung der Ein- und Ausschlusskriterien in das Register hätten eingehen können.

Anwendbarkeit

Der QI ist für alle drei CERTAIN-Register anwendbar, da dokumentiert wird, wann ein Patient in das Register einging.

Implementierbarkeit

Aus den Daten der Register kann anhand eines Datenelementes einfach entnommen werden, wann ein Patient in das jeweilige Register eingegangen ist. Schwieriger gestaltet sich die Berechnung der Anzahl Patienten, welche in das Register hätten eingehen können.

Für die Berechnung des QI sollten zwei Fälle unterschieden werden. Wenn nicht bekannt ist, wie viele Patienten tatsächlich die Einschlusskriterien erfüllen, sollte die in der TMF-Leitlinie vorgeschlagene Berechnungsmöglichkeit mit der Betrachtung des vorherigen In-

tervalls unter Berücksichtigung einer zu wählenden Variabilität genutzt werden. Dies liegt darin begründet, dass keine verlässlichen, absoluten Zahlen für das zu prüfende Intervall vorhanden sind und die Berechnung sich somit an einem Schätzer orientieren sollte. Zudem muss nicht zwangsläufig die exakt gleiche Anzahl an Patienten in der Einrichtung zu rekrutieren gewesen sein, die Variabilität berücksichtigt dies. Variabilität kann ebenso sinnvoll sein, um externe Faktoren abzubilden, wie beispielsweise den Abfall der Transplantationszahlen im Zuge des Organtransplantationsskandals.

Für das CERTAIN-Register ist durch Auskunft von Eurotransplant für die Eurotransplant-Länder bekannt, wie viele Transplantationen an jeder Einrichtung in einem Zeitintervall erfolgten. Für Länder, in denen Transplantate nicht durch ET allokiert werden, muss diese Zahl aus der Fachliteratur entnommen werden. In diesem Fall bleibt nur die Betrachtung der offiziellen Transplantationszahlen in den jeweiligen Ländern. Bei einer Aufschlüsselung dieser Zahlen für jede Einrichtung kann die Rekrutierungsrate ebenfalls berechnet werden. Für die deutschen Einrichtungen des CERTAIN-LI-Registers können die exakten Transplantationszahlen jeder Einrichtung aus den Berichten der Deutschen Stiftung Organtransplantation entnommen und mit den rekrutierten Patienten der jeweiligen Einrichtung verrechnet werden. Bei den ausländischen Einrichtungen ist die Berechnung gegebenenfalls über die offiziellen Transplantationszahlen des jeweiligen Landes möglich.

Die Berechnung der Rekrutierungsrate ist für das KRhOKo-Register nur über die Nutzung von Schätzungen der Inzidenz möglich, welche selbst wiederum auf empirischen Studien basieren (Zink und Albrecht 2016).

Aussagekraft des QI

Da mit der Größe der Registerpopulation die Aussagekraft der wissenschaftlichen Auswertungen über die Registerdaten steigt, ist die Rekrutierungsrate eine wichtige Kenngröße für die Aussagekraft von Auswertungen anhand der Registerdaten und somit für die wissenschaftliche Relevanz des Registers. Die Berechnung des Wertes des QI ist daher sehr wichtig für die Attraktivität des Registers und sollte entsprechend hoch priorisiert werden. Die Rekrutierungsrate besitzt zudem, wie in der TMF-Leitlinie beschrieben, Aussagekraft zur Schwere der Einwilligung sowie der Teilnahmebereitschaft von Patienten und Einrichtungen, entscheidende Punkte für den Erfolg eines Registers.

Für die Einschätzung der Aussagekraft sollte die jeweils gewählte Form der Berechnung und deren Berechnungsgrundlage berücksichtigt werden. Sollten absolute Zahlen der zu rekrutierenden Patienten pro Einrichtung verfügbar sein ist die Aussagekraft des QI im Vergleich zur Berechnung der Rekrutierungsrate anhand der Inzidenz als sehr viel besser einzuschätzen.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 39.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
++	o	++	Report

Tabelle 39: Bewertung TMF-1030 - Rekrutierungsrate

TMF-1051 - DCO-Rate (Death Certificate Only)

Der QI „DCO-Rate“ beschreibt den Anteil der Krebserkrankungen, für die einzig eine Todesbescheinigung Hinweis auf eine Krebserkrankung darstellt, an der Gesamtzahl der erfassten Krebserkrankungen.

Eine Anwendung des QI ist für die CERTAIN-Register nicht möglich, da keine Informationen anhand von Todesbescheinigungen erfasst werden. Auf eine abschließende Bewertung, Tabelle 40, wird daher verzichtet.

Die Ausführungen der TMF-Leitlinie beschreiben diesen QI als wichtigsten Indikator für epidemiologische Krebsregister. Für Krebsregister ist die Berechnung des QI auf Basis dieser Ausführungen dringend zu empfehlen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 40: Bewertung TMF-1051 - DCO-Rate (Death Certificate Only)

TMF-1031 - Verweigerungsrate von Untersuchungen

Dieser QI beschreibt den Anteil der verweigten Untersuchungen an der Anzahl der überprüften Untersuchungen eines Datenbestands und kann nur bestimmt werden, wenn Verweigerungen von Untersuchungen dokumentiert werden können.

Eine Anwendung des QI ist für die CERTAIN-Register nicht möglich, da keine Informationen zur Verweigerung von Untersuchungen erfasst werden. Auf eine abschließende Bewertung, Tabelle 41, wird daher verzichtet.

In Gesprächen mit dem Personal der Registerzentrale des KRhOKo-Registers wurden Verweigerungsraten bei epidemiologischen Vorhaben als nur bedingt interessant erachtet. Größere Bedeutung wurde Verweigerungsraten bei kleineren Studienpopulationen zugemessen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 41: Bewertung TMF-1031 - Verweigerungsrate von Untersuchungen

TMF-1032 - Verweigerungsrate von Modulen

Der Anteil der verweigten Module an der Gesamtzahl der überprüften Module wird durch den Wert dieses QI beschrieben. Die Berechnung kann nur erfolgen wenn die Verweigerung dokumentiert wird.

Eine Anwendung des QI ist für die CERTAIN-Register ist daher nicht möglich, da keine Informationen zur Verweigerung von bestimmten Untersuchungen erfasst werden. Auf eine abschließende Bewertung, Tabelle 42, wird daher verzichtet.

Wie auch beim vorherigen QI wurden diesem QI von Seiten des Personals des KRhOKo-Registers mehr Bedeutung zugemessen je kleiner die Studienpopulation ist.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 42: Bewertung TMF-1032 - Verweigerungsrate von Modulen

TMF-1033 - Verweigerungsrate von einzelnen Datenelementen

Der Wert dieses QI beschreibt den Anteil der Datenwerte mit dokumentierter Verweigerung an der Gesamtzahl der überprüften Datenwerte. Gleich den beiden vorangegangenen QI kann die Berechnung nur erfolgen wenn die Verweigerung dokumentiert wird.

Eine Anwendung des QI ist für die CERTAIN-Register nicht möglich, da keine Informationen zur Verweigerung von bestimmten Datenelementen erfasst werden. Auf eine abschließende Bewertung, Tabelle 43, wird daher verzichtet.

Wie auch beim vorherigen QI wurden diesem QI von Seiten des Personals des KRhOKo-Registers mehr Bedeutung zugemessen je kleiner die Studienpopulation ist.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 43: Bewertung TMF-1033 - Verweigerungsrate von einzelnen Datenelementen

TMF-1034 - Vorzeitig aus dem Register ausscheidende Beobachtungseinheiten

Dieser QI berechnet das Verhältnis der aus dem Datenbestand des Registers vorzeitig ausscheidenden Patienten in einem gewählten Zeitraum im Hinblick auf die Anzahl der neu in das Register aufgenommenen Patienten in diesem Zeitraum.

Anwendbarkeit

In allen drei Registern gibt es die Möglichkeit, Patienten als lost-to-follow-up zu markieren sowie einen Grund diesbezüglich anzugeben. Der QI ist daher für alle drei CERTAIN-Register gleichermaßen anwendbar.

Implementierbarkeit

Wenn Patienten in den CERTAIN-Registern als lost-to-follow-up markiert werden, wird der genaue Zeitpunkt dieses Schrittes nicht erfasst. Die Berechnung der Anzahl der ausgeschiedenen Patienten in einem Zeitintervall muss daher über den Abgleich der lost-to-follow-up-Zahlen zu Beginn und zu Ende des gewählten Zeitintervalls erfolgen. Dies erschwert die Berechnung, da das Zeitintervall im Voraus festgelegt und die zu Beginn des Intervalls festgestellte Anzahl extra dokumentiert werden muss. Die Berechnung verliert hierdurch an Flexibilität. Der Eingangszeitpunkt eines Patienten in das jeweilige Register wird explizit dokumentiert. Die Anzahl der im betrachteten Zeitintervall in das Register eingegangenen Patienten ist daher einfach für die Berechnung des QI auslesbar.

Die Berechnung kann in Teilen automatisiert erfolgen, ist jedoch mit einer Vorlaufzeit und Vorarbeiten verbunden. Für eine vereinfachte Berechnung sollte künftig in den drei Registern dokumentiert werden, wann ein Patient als lost-to-follow-up markiert wurde.

Aussagekraft des QI

Die Aussagekraft des QI hängt von dem jeweiligen Vorgehen der Einrichtungen bezüglich lost-to-follow-up und der Wahl des Zeitintervalls ab. Werden Patienten kontinuierlich oder automatisch nach Verstreichen einer Frist seit der letzten Überprüfung auf lost-to-follow-up als lost-to-follow-up markiert sind die Zeitintervalle frei wählbar. Wenn hingegen beispielsweise nur einmal im Jahr alle Patienten überprüft und zu diesem Zeitpunkt als lost-to-follow-up markiert werden, muss dies entsprechend bei der Wahl des Zeitintervalls und des Berechnungszeitpunktes berücksichtigt werden, da sonst keine schlüssigen Aussagen möglich sind.

Für die Berechnung sollte zudem auch berücksichtigt werden, ob Patienten durch einen krankheitsbedingten Tod aus dem Register ausgeschieden sind oder durch unvorhergesehene Ereignisse, wie beispielsweise einen tödlichen Autounfall. Der Grund des Ablebens eines Patienten wird in beiden Transplantationsregistern dokumentiert.

Verbesserte Aussagekraft des QI kann auch durch eine Unterscheidung zwischen den Gründen des vorzeitigen Ausscheidens erreicht werden, in der TMF-Leitlinie beispielhaft Rückzug der Einwilligung und lost-to-follow-up. Beide Gründe haben verschiedene Ursachen, eine getrennte Berechnung kann differenziertere Aussagen ermöglichen. In den drei CERTAIN-Registern ist hier insbesondere die Transition in den Erwachsenenbereich zu berücksichtigen, die sich gänzlich vom Rückzug der Einwilligung und dem Verlust des Kontakts zum Patienten unterscheidet.

Alternativ zur Berechnungsformel der TMF-Leitlinie kann auch die Anzahl der vorzeitig im Zeitintervall ausgeschiedenen Patienten über der Summe aus Patienten zum Ende des Zeitintervalls und der vorzeitig ausgeschiedenen Patienten betrachtet werden. Dies bietet den Vorteil, dass die Rate sich an der Gesamtzahl der Patienten orientiert und somit in besserem Bezug zur Gesamtpopulation des Registers steht.

Durch die unterschiedlichen Aspekte, welche bei einer Betrachtung des QI sichtbar werden können, ist eine Umsetzung des QI als grundsätzlich sinnvoll zu erachten.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 44.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	o	+	Report

Tabelle 44: Bewertung TMF-1034 - Vorzeitig aus dem Register ausscheidende Beobachtungseinheiten

TMF-1036 - Synonyme

Synonyme bezeichnen im Zusammenhang mit Registern mindestens zwei Patientenidentifikatoren, welche auf den selben Patienten zeigen. Der Wert des QI beschreibt den Anteil der Synonyme an der Gesamtzahl der überprüften Patienten eines Datenbestands. Synonyme beeinflussen die Inzidenzen und Prävalenzen der Registerpopulation.

Anwendbarkeit

Bisher existieren in den drei CERTAIN-Registern keine expliziten Überprüfungen hinsichtlich der Existenz von Synonymen. Durch *primary key constraints* auf Datenbankebene für den Patientenidentifikator wird bereits verhindert, dass einem Patienten mehrere Patientenidentifikatoren zugewiesen werden können.

Darüber hinaus existieren in allen drei Registern Studienidentifikatoren, welche einem Patienten im Rahmen einer Studie einen eindeutigen Identifikator zuweisen. Für diese Studienidentifikatoren existieren derzeit keine technischen Maßnahmen auf Datenbankebene zur Verhinderung von Synonymen, lediglich in den Eingabemasken der drei Register wird die Mehrfachvergabe von Studienidentifikatoren für die selbe Studie für einen Patienten verhindert. Beim Anlegen der Studienidentifikatoren in der Applikationslogik wird ebenfalls nicht überprüft, ob dem Patienten bereits ein Studienidentifikator zu dieser Studie zugewiesen wurde.

Da Patienten in den CERTAIN-Registern keine anderen Identifikatoren zu gewiesen werden können ist die Anwendbarkeit begrenzt. Durch Überprüfungen und das Verhindern von Synonymen beim Anlegen der Studienidentifikatoren kann die Berechnung des QI zurückgestellt werden.

Implementierbarkeit

Synonyme Studienidentifikatoren sind einfach durch die Bildung eines Sets aus dem Patientenidentifikator und der Studie identifizierbar, da diese Wertkombination einzigartig sein muss. Die Berechnung ist daher einfach durchführbar, für alle drei Register wiederverwendbar und kann automatisiert erfolgen.

Aussagekraft des QI

Die Mehrfachzählung von Patienten beeinflusst die Aussagekraft von Auswertungen der Registerdaten, entsprechend wichtig ist es, die Existenz von Synonymen zu erkennen und zu verhindern.

Tabelle 45 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
o	+ +	+ +	Registerplanung

Tabelle 45: Bewertung TMF-1036 - Synonyme

TMF-1037 - Homonyme

Homonyme bezeichnen im Zusammenhang mit Registern Patientenidentifikatoren, welche fälschlich zwei Patienten zugeordnet sind. Der Wert des QI beschreibt den Anteil der Homonyme an der Gesamtzahl der überprüften Patienten eines Datenbestands. Homonyme beeinflussen die Inzidenzen und Prävalenzen der Registerpopulation.

Anwendbarkeit

In allen drei CERTAIN-Registern unterliegen die eingesetzten Patienten- und Studienidentifikatoren auf Datenbankebene *primary key constraints*. Dies stellt sicher, dass keine Identifikatoren doppelt an verschiedene Patienten vergeben werden können. Zudem werden die Patientenidentifikatoren in jedem Register durch eine fortlaufende Sequenz erzeugt. Die Studienidentifikatoren werden durch einen Zufallsgenerator erzeugt und vor dem Anlegen auf eine mögliche Doppelvergabe hin überprüft. Der QI besitzt für die CERTAIN-Register daher keine Relevanz, grundsätzlich sollten alle Identifikatoren durch eine zentrale Stelle vergeben werden.

Implementierbarkeit

Die Überprüfung auf Homonyme kann in den CERTAIN-Registern einfach durch das Laden aller Patienten oder Studien aus der Datenbank und anschließenden Vergleich der verwendeten Identifikatoren gegeneinander erfolgen. Der Aufwand zur automatischen Überprüfung auf Homonyme ist daher gering, die verwendeten Algorithmen können für alle drei Register wiederverwendet werden.

Aussagekraft des QI

Die fälschliche Zuordnung von Daten zu einem anderen Patienten kann, wie in der TMF-Leitlinie beschrieben, zu widersprüchlichen Daten führen, welche gesondert geprüft werden müssen und somit hohen Arbeitsaufwand für Einrichtung als auch Registerzentrale nach sich ziehen. Homonyme stellen die Integrität der Registerdaten in Frage, da sie ein Symptom mangelnder Absicherung der Korrektheit der Prozesse des Registers sind.

Die Überprüfung auf das Vorhandensein von Homonymen im Rahmen des QI ist bei hinreichend sorgfältiger Umsetzung von technischen Maßnahmen zur Verhinderung von Homonymen nicht erforderlich. Sollte dies nicht gewährleistet werden können, sind Überprüfungen des Datenbestands auf Homonyme zwingend erforderlich da die Integrität des Registers ansonsten in Frage zu ziehen ist. Der QI bietet entsprechend eine sehr gute Aussagekraft zur Prozessqualität.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 46.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
--	++	++	Registerplanung

Tabelle 46: Bewertung TMF-1037 - Homonyme

TMF-1038 - Einzelmeldungen pro Beobachtungseinheit

Der QI „Einzelmeldungen pro Beobachtungseinheit“ berechnet den Anteil der Tumorfälle mit nur einer Meldung an der Gesamtzahl der überprüften Tumorfälle einer einzelnen Beobachtungseinheit oder eines Datenbestands.

Eine Anwendung des QI ist für die CERTAIN-Register nicht möglich, da keine Tumorfälle erfasst werden. Die Betrachtung der Datenquellen pro Beobachtungseinheit findet bereits im Rahmen der Berechnung des QI „Datenquellen pro Beobachtungseinheit“ statt. Eine mögliche Änderung des Betrachtungsrahmens dieses QI von Einzelmeldungen hin zu Datenquellen, um eine Anwendung zu ermöglichen, ist daher nicht sinnvoll. Auf eine abschließende Bewertung, Tabelle 47, wird daher verzichtet.

Die TMF-Leitlinie betont die Wichtigkeit dieses QI, da das Vorliegen nur einer einzelnen Meldung Hinweis auf unvollständige oder fehlende Meldungen sein kann. Da es sich hierbei um einen QI handelt der nur auf Tumorregister anwendbar ist, konnten durch Befragung des medizinischen Personals der Registerzentralen keine weiteren Einblicke gewonnen werden.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 47: Bewertung TMF-1038 - Einzelmeldungen pro Beobachtungseinheit

TMF-1039 - Solitäre Meldungen von Pathologen

Der QI „Solitäre Meldungen von Pathologen“ berechnet den Anteil der Beobachtungseinheiten mit einer solitären Meldung eines Pathologen an der Gesamtzahl der überprüften Beobachtungseinheiten eines Datenbestands.

Eine Anwendung des QI ist für die CERTAIN-Register nicht möglich, da keine Meldungen von Pathologen erfasst werden. Auf eine abschließende Bewertung, Tabelle 48, wird daher verzichtet.

Dieser QI stellt eine Unterkategorie des vorherigen QI dar. Die TMF-Leitlinie betont in ihren Ausführungen, dass beim Eingang eines Patienten in ein Tumorregister nur auf Basis einer einzelnen Meldung eines Pathologen oft wichtige Informationen fehlen. Diese seien oft nicht Teil der Meldung, da sie dem Pathologen nicht bekannt sein. Da dieser QI eine Spezialisierung des vorherigen QI darstellt und die TMF-Leitlinie die Wichtigkeit des vorherigen QI betont, sollte diesem QI daher mehr Bedeutung zugemessen werden als dem vorherigen QI. Da es sich hierbei um einen QI handelt der nur auf Tumorregister anwendbar ist, konnten durch Befragung des medizinischen Personals der Registerzentralen keine weiteren Einblicke gewonnen werden.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 48: Bewertung TMF-1039 - Solitäre Meldungen von Pathologen

TMF-1040 - Zurückgewiesene Meldungen

Dieser QI misst den Anteil der zurückgewiesenen Meldungen an der Gesamtzahl der, von einer Einrichtung erhaltenen, Meldungen im Datenbestand. Meldungen sollten, gemäß Leitlinie, zurückgewiesen werden, wenn fehlerhafte Angaben vorliegen. Die meldende Einrichtung hat dann die Möglichkeit die Angaben zu korrigieren und erneut einzureichen.

Anwendbarkeit

Die drei CERTAIN-Register erhalten keine Meldungen zu Krebserkrankungen, um den QI anzuwenden kann der Betrachtungsrahmen jedoch auf Datensätze beziehungsweise Visiten geändert werden. In den drei CERTAIN-Registern können Visiten von Patienten aufgrund von Beanstandungen während des Datenqualitätsmanagementprozesses durch den Datenqualitätsmanager zurückgewiesen werden. Der QI kann unter diesem Betrachtungswinkel somit für alle drei Register berechnet werden.

Implementierbarkeit

Sollte eine Visite zurückgewiesen worden sein, wird dies im Bearbeitungsstatus der Visite dokumentiert. Durch Erfassung aller zurückgewiesenen Visiten kann der QI einfach automatisiert berechnet werden.

Problematisch an der Art der Berechnung ist, dass eine Visite zwischen den Berechnungszeitpunkten des QI mehrmals zurückgewiesen werden kann, die Visite aber nur einmal in die Berechnung des QI eingeht. Zudem wird nur jeweils die Anzahl der zurückgewiesenen Visiten zum jeweiligen Berechnungszeitpunkten betrachtet, Informationen über gehäufte Zurückweisungen werden nicht in die Berechnung miteinbezogen.

Die absolute Anzahl der vorgenommenen Zurückweisungen kann in den drei CERTAIN-Registern aus den Protokolldaten der Änderungen der Bearbeitungsstatus der Visiten entnommen werden. Diese Informationen dokumentieren auch mehrfache Zurückweisungen und liefern ein genaueres Bild der Prozessqualität.

Beide Berechnungsvarianten sind für alle drei Register gleich und können automatisiert durchgeführt werden.

Aussagekraft des QI

Die Berechnung dieses QI ist als sinnvoll zu erachten, da sie einen Überblick über anhaltende Probleme der Datenqualität bietet, wenn Visiten gehäuft zurückgewiesen werden. Mehrfach zurückgewiesene Visiten sollten sehr genau untersucht werden, um auszuschließen, dass die Daten mit dem Ziel korrigiert wurden, die Visite als augenscheinlich korrekt darzustellen.

Die Anwendung von Plausibilitätsprüfungen zur Eingabezeit, zu den QI TMF-1003, TMF-1004 und TMF-1014, kann genutzt werden, um zu verhindern dass fehlerhafte Visiten überhaupt durch den DQM überprüft und gegebenenfalls zurückgewiesen werden müssen. Dieses Vorgehen verbessert die Aussagekraft des QI da die Anzahl der zurückgewiesenen Visiten gesenkt wird und einer Zurückweisung ein höherer Stellenwert zukommt.

Teil der Dokumentation der Zurückweisung sollten neben dem Grund, in einer verbalen Beschreibung, auch Informationen zum zurückweisenden DQM sowie die betroffenen Datenelemente und ihre Datenwerte sein, damit bei einer erneuten Vorlage der Meldung nicht nochmals alle Datenelemente kontrolliert werden müssen. Durch die Dokumentation der Datenelemente kann zudem analysiert werden, ob bei der meldenden Einrichtung spezielle Datenelemente besonders häufig betroffen sind, und somit gezielt einer Prüfung unterzogen werden sollten.

Die erhobenen Daten zur Zurückweisung von Meldungen können auch unter dem Aspekt der Zurückweisungen pro Datenqualitätsmanager betrachtet werden, um zu überprüfen, ob verschiedene DQM Meldungen unterschiedlich häufig zurückweisen. Eine Vielzahl von zurückgewiesenen Meldungen sollte Anlass für eine Kontaktaufnahme mit der Einrichtung sein und kann Anlass für eine Nachschulung des Eingabepersonals bieten.

Die CERTAIN-Register bieten derzeit nur die Möglichkeit den Grund für Zurückweisungen als Kommentar zur Visite hinzuzufügen. Die Verarbeitung dieser unstrukturierten Daten ist daher nur schwer möglich, zumal die Kommentarfunktion auch für die Kommunika-

tion mit den Eingabekräften der Einrichtung genutzt wird und Kommentare daher nicht zwangsläufig mit der Zurückweisung in Verbindung stehen.

Da zurückgewiesene Visiten in den drei Registern keinen Eingang in die Auswertungen der Registerdaten finden, besteht aufgrund zurückgewiesener Visiten zunächst kein direkter Handlungsbedarf. Bei vermehrtem Auftreten von Zurückweisungen sollten die Visiten unter Berücksichtigung der ausgeführten Überlegungen analysiert werden.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 49.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
++	++	+	Report

Tabelle 49: Bewertung TMF-1040 - Zurückgewiesene Meldungen

TMF-1041 - Datenquellen pro Beobachtungseinheit

Dieser QI berechnet den Anteil der Beobachtungseinheiten mit nur einer Datenquelle an der Gesamtzahl der überprüften Beobachtungseinheiten.

Anwendbarkeit

Die TMF-Leitlinie beschränkt die Anwendbarkeit des QI in der Beschreibung auf Tumorfälle, der QI ist somit in seiner ursprünglichen Form nicht auf die drei CERTAIN-Register anwendbar. Anwendbarkeit kann durch Änderung des Betrachtungswinkels auf jegliche Erkrankung erreicht werden, wenn Informationen aus mehreren Datenquellen zum Eingang in das Register führen können.

Für das CERTAIN-Register besteht die Möglichkeit des Erhaltes von Informationen durch die transplantierende Einrichtung, von der Collaborative Transplant Study, sowie von Eurotransplant, in Eurotransplant-Ländern. In Deutschland existiert als weitere Datenquelle das Kuratorium für Dialyse und Nierentransplantation³ (KfH). Es besteht zum Zeitpunkt dieser Arbeit jedoch kein Zugang zu den Daten des KfH. Für teilnehmende Einrichtungen aus anderen Ländern existiert nur CTS als weitere bekannte Datenquelle. Aufgrund des Aufwandes des Datenaustausches mit CTS wurde jedoch nach dem Import der CTS-Daten für die Einrichtung in Heidelberg auf weitere Datenimporte für andere Einrichtungen in der derzeitigen Form des Imports verzichtet, so dass diese Datenquelle nicht mehr genutzt wird.

Der Datenaustausch mit CTS wurde aus diesem Grund nicht für CERTAIN-LI etabliert. CERTAIN-LI erhält ebenso keine Daten von Eurotransplant, die Daten des CERTAIN-LI-Registers entstammen daher immer nur aus einer Quelle, der jeweiligen Einrichtung selbst. Der QI ist somit derzeit nicht für das CERTAIN-LI-Register anwendbar.

³<https://www.kfh.de/>

Gleichermaßen existieren für KRhOKo keine weitere Datenquellen als die eingebende Einrichtung selbst, der QI ist daher auch für dieses Register nicht anwendbar.

Implementierbarkeit

Für Eurotransplant-Einrichtungen ist der QI anhand der von ET erhaltenen *Notifications* berechenbar. *Notifications* informieren das CERTAIN-Register über jede Transplantation einer teilnehmenden ET-Einrichtung. Bei Aktivierung einer *Notification* durch die Einrichtung werden die verfügbaren Daten von ET abgefragt und in den Datensatz integriert. Im selben Schritt wird in der *Notification* vermerkt, dass Daten zu dieser Transplantation angefordert wurden. Die Berechnung für Beobachtungseinheiten aus ET-Ländern ist daher einfach möglich.

Für die Heidelberger Einrichtung kann, über die in jeder Visite vermerkte Herkunft der Visite, festgestellt werden, ob Daten durch den CTS-Import in diese Visite übernommen wurden.

Anhand dieser Informationen kann die Berechnung zur Anzahl der Quellen pro Beobachtungseinheit einfach automatisiert werden.

Die Herkunft der Visite wird auch in den beiden anderen Registern dokumentiert, bei Gewinnung von Datenquellen für diese Register können die zu entwickelnden Berechnungsmethoden wiederverwendet werden.

Aussagekraft des QI

Die Überprüfung des Datenbestands im Hinblick auf diesen QI ist als sinnvoll zu erachten, da die Vollzähligkeit mit steigender Anzahl an Datenquellen ebenso steigen sollte (Hentschel und Katalinic 2008). Mehrere Datenquellen pro Beobachtungseinheit sind darüber hinaus ein effektives Mittel gegen Datenfälschung.

Tabelle 50 fasst die abschließende Bewertung des QI zusammen.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
-	+ +	+ +	Report

Tabelle 50: Bewertung TMF-1041 - Datenquellen pro Beobachtungseinheit

TMF-1042 - Beobachtungseinheiten mit Follow-up

Der Anteil der Beobachtungseinheiten mit Follow-up an der Gesamtzahl der überprüften Beobachtungseinheiten wird durch den Wert dieses QI beschrieben.

Anwendbarkeit

Da alle drei Register auf einer Dokumentation auf der Basis von Visiten aufbauen, ist der QI für alle drei Register berechenbar.

Implementierbarkeit

Die Prüfung, ob ein Patient Follow-ups besitzt, ist technisch sehr einfach umzusetzen und für alle drei Register wiederverwendbar. Die Berechnung kann somit mit sehr geringem Aufwand automatisiert erfolgen.

Aussagekraft des QI

Die Aussagekraft des QI ist als gut einzuschätzen, da sich anhand des berechneten Wertes des QI Aussagen zur Nachhaltigkeit der Dateneingabe durch Einrichtungen machen lassen. Für das medizinische Personal des CERTAIN-Registers ist der Wert des QI darüber hinaus ein Faktor für die Bewertung der Attraktivität eines Registers sowie der Wertigkeit der Aussagen die Anhand der Daten des Registers getroffen werden. Der QI kann zudem Hinweise auf Probleme bei der Datenerhebung der Einrichtungen oder auf fehlende Dokumentation bieten.

Tabelle 51 stellt die zusammenfassende Bewertung des QI dar.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	+ +	+	Report

Tabelle 51: Bewertung TMF-1042 - Beobachtungseinheiten mit Follow-up

6.3.3 Qualitätsindikatoren der Ebene Richtigkeit

Im Folgenden werden die Eigenschaften der QI der Ebene Richtigkeit für die drei CERTAIN-Register dargelegt.

TMF-1043 - Genauigkeit der Registerdaten

Dieser QI beschreibt den Anteil der vorhandenen und als korrekt bewerteten Datenwerte im Register an der Gesamtzahl der überprüften Datenwerte. Die Berechnung des QI erfolgt für Datenelemente und kann einzelne Beobachtungseinheiten oder den Datenbestand umfassen.

Anwendbarkeit

Grundsätzlich ist der QI auf alle Datenwerte der drei Register anwendbar.

Implementierbarkeit

Die TMF-Leitlinie fordert, die Korrektheit der Datenwerte anhand einer unabhängigen Datenquelle zu verifizieren, welche „nicht für die Erhebung der Registerdaten benutzt wurde“ (Nonnemacher et al. 2014, S. 77). Die Erschließung dieser Datenquellen ist aufgrund politischer und technischer Hürden mit sehr hohem Aufwand verbunden und zuweilen aus diesen Gründen nicht möglich. Zudem werden verfügbare Datenquellen meist auch für die Datenerhebung genutzt.

Für KRhOKo existieren keine weiteren bekannten Datenquellen außer den Primärsystemen der Einrichtungen, eine für die Berechnung benötigte unabhängige Datenquelle existiert somit nicht.

Das CERTAIN-Register nutzt bereits alle zugänglichen Quellen für die Datenerhebung. Ausnahme bildet hier CTS, welches als Datenquelle zwar zur Verfügung steht, jedoch jeweils nur einen geringen Teil des CERTAIN- und CERTAIN-LI-Datensatzes abdeckt. Das CERTAIN-Register bietet eine Funktion zum Import von CTS-Daten in das Register. Aufgrund des mit dem Import verbundenen Aufwandes wurde beschlossen die Importfunktion in der bestehenden Form nicht weiter zu verwenden, so dass die CTS-Daten zur Bestimmung der Genauigkeit genutzt werden könnten. Die CTS-Daten besitzen jedoch einen anderen Fokus als der CERTAIN/-LI-Datensatz, sind mitunter grobgranularer und lassen sich nicht zwingend auf den CERTAIN/-LI-Datensatz abbilden. Grundsätzlich könnte nur ein kleiner Ausschnitt der Registerdaten der beiden Transplantationsregister überprüft werden. Die Möglichkeit des Datenaustauschs zwischen CTS und CERTAIN-LI existiert derzeit nicht. Die Importfunktion aus CERTAIN müsste in CERTAIN-LI übernommen und für den Datensatz angepasst werden. Weitere für CERTAIN/-LI unzugängliche Datenquellen werden auch nicht für die Berechnung dieses QI erschlossen werden können.

Ein technischer Abgleich zwischen unabhängiger Datenquelle und Registerdaten ist aufgrund der Anzahl der Datenelemente der drei Register nur schwer realisierbar. Zugleich ist die in der TMF-Leitlinie genannte „Abstraktion der Wahrheit“ aus der Datenquelle erforderlich, dies kann eine Harmonisierung der Datenmodelle beinhalten und entsprechend aufwändig sein. Somit ist die Implementierbarkeit des QI als sehr schlecht zu erachten.

Aussagekraft des QI

Der QI bietet sehr gute Aussagekraft zur Datenqualität, unter der Annahme, dass die für den Vergleich gewählte Datenquelle auch tatsächlich korrekte Daten enthält. Gleich dem zu überprüfenden Datenbestand kann auch der Datenbestand der gewählten Quelle Fehler enthalten, entsprechend ist es zwingend notwendig, sich vor dem Abgleich der Daten im Rahmen der Berechnung dieses QI von der Datenqualität der gewählten Datenquelle zu überzeugen.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 52.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
-	--	++	Datenverifizierung

Tabelle 52: Bewertung TMF-1043 - Genauigkeit der Registerdaten

TMF-1044 - Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente

Der Anteil der Datenwerte, welche nicht mit den Originaldaten übereinstimmen, an der Gesamtzahl der überprüften Datenwerte wird durch den Wert dieses QI wiedergegeben. Originaldaten können auf Papier oder in Form von elektronischen Dokumenten existieren.

Anwendbarkeit

Dieser QI ist auf alle drei Register und deren Datenelemente anwendbar, wobei für das KRhOKO-Register Einschränkungen existieren. Im KRhOKo-Register können Patienten auch direkt ihre Fragebogen online über das Register-System ausfüllen, in diesem Fall entsprechen die Registerdaten den Originaldaten und der QI ist somit nicht anwendbar. Alle anderen medizinischen Daten der drei Register werden nicht direkt über die Register erhoben sondern entstammen anderen Datenquellen.

Implementierbarkeit

Die Originaldaten entstammen im Falle der drei CERTAIN-Register aus den Primärsystemen von Krankenhäusern oder Arztpraxen verteilt in Deutschland und, im Falle von CERTAIN und CERTAIN-LI, europa- beziehungsweise weltweit. Der Zugriff auf diese Systeme ist aufgrund der Entfernung, des damit verbundenen Reiseaufwands und der großen Zahl der zu besuchenden Einrichtungen als unwahrscheinlich zu erachten. Technischer Zugriff ist aufgrund der Heterogenität der Systeme und den damit zu etablierenden Einzelösungen ebenso unwahrscheinlich. Die Überprüfung der Datenwerte ist somit nur schwer möglich, eine technische Umsetzung einer automatisierten Berechnung ebenso schwierig. Der Abgleich müsste daher für alle Datenelemente manuell erfolgen. Dies geht mit sehr hohem Arbeitsaufwand einher.

Aussagekraft des QI

Zwar werden weder der Wahrheitsgehalt noch die Richtigkeit oder Plausibilität der Datenwerte überprüft, der QI bietet dennoch sehr gute Aussagekraft zur Datenqualität da jede Differenz zwischen Original- und Registerdaten einen Datenfehler darstellt.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 53.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+ +	- -	+ +	Datenverifizierung

Tabelle 53: Bewertung TMF-1044 - Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente

TMF-1045 - Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten

Dieser QI berechnet den Anteil der Beobachtungseinheiten, welche in mindestens einem Datenwert nicht mit den Originaldaten übereinstimmen, an der Gesamtzahl der überprüften Beobachtungseinheiten.

Anwendbarkeit

Gleichermaßen wie der vorherige QI ist dieser QI auf alle drei Register anwendbar. Die Ausnahme bilden wiederum Fragebogen, welche von den Patienten direkt in das KRhOKo-Register eingegeben wurden, da es sich hierbei um die Originaldaten handelt und somit kein Vergleich möglich ist.

Implementierbarkeit

Durch die Änderung des Betrachtungsrahmens auf Beobachtungseinheiten ergeben sich im Hinblick auf die Implementierbarkeit des QI keine Änderung im Vergleich zum vorherigen QI. Der QI ist gleich seinem Vorgänger aufgrund von technischen Beschränkungen und Medienbrüchen kaum automatisiert umsetzbar.

Aussagekraft des QI

Da bereits ein Datenfehler zu einem negativen Eingang in die Berechnung führt ist der Wert des QI mitunter überzeichnet, die Datenqualität wird sehr viel schlechter dargestellt als sie tatsächlich ist. Besonders Register mit einem großen Datensatz, wie die beiden Transplantationsregister, sind bei dieser Art der Berechnung prädestiniert für einen schlechten berechneten Wert des QI. Prins et al. nennen für den QI „Genauigkeit der Registerdaten“ eine Übereinstimmung von 95 % als gut (Prins et al. 2000). Nach dieser Definition wären auch Beobachtungseinheiten mit nur einem Fehler noch von guter Datenqualität.

Eine alternative Berechnung wäre die Festlegung eines Schwellenwertes für die Anzahl von Fehlern, welche die Anzahl der Datenelemente der Beobachtungseinheit einbezieht. Beispielsweise könnte die Beobachtungseinheit in den Zähler der Berechnung eingehen wenn $\frac{\text{Anzahl fehlerhafter Datenwerte}}{\text{Anzahl überprüfter Datenwerte}} * 100 > 5$.

Zusätzlich kann die Berechnung der durchschnittlichen Fehleranzahl pro Beobachtungseinheit zusätzliche Informationen zur Datenqualität bereitstellen.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 54.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
++	--	+	Datenverifizierung

Tabelle 54: Bewertung TMF-1045 - Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten

TMF-1046 - Vollständigkeit der Registerdaten

Der Wert dieses QI beschreibt den Anteil der im Register vorhandenen Datenwerte an der Gesamtzahl der Datenwerte, welche in das Register eingegangen sind oder hätten eingehen können.

Anwendbarkeit

Die Berechnung des QI erfordert zusätzlich zu den Registerdaten eine weitere Datenquelle anhand derer die Vollständigkeit überprüft werden kann. Für die beiden Transplantationsregister kann hierfür das CTS-Register herangezogen werden, welches gleichermaßen Daten zu Transplantationen erfasst. Andere Datenquellen stehen zum Zeitpunkt dieser Arbeit nicht zur Verfügung oder werden, beispielsweise im Falle von ET für CERTAIN, zum Import von Daten in das Register genutzt.

Für das KRhOKo-Register existieren außer der Originalquelle keine weiteren Datenquellen, der QI ist daher nicht anwendbar.

Implementierbarkeit

Aufgrund der bereits zum QI TMF-1043 ausgeführten Problematik hinsichtlich des CTS-Datensatzes stellt das CTS-Register keine ideale Datenquelle zur Prüfung der Vollständigkeit auf Ebene von Datenelementen dar. Die CTS-Daten können zur Berechnung des QI im Hinblick auf das Eingehen der Fälle genutzt werden, da die Fälle in beiden Registern vorhanden sein sollten. Je nach Aktualität der Daten im jeweiligen Register kann es jedoch vorkommen, dass der Fall in einem der Register noch nicht vorhanden ist. Dies sollte bei der Berechnung berücksichtigt werden.

Für das CERTAIN-LI-Register existiert Funktion zum Import von CTS-Daten nicht und wäre aus dem CERTAIN-Register zu übernehmen sowie für den CERTAIN-LI-Datensatz anzupassen. Die Berechnung wäre mit sehr hohem Aufwand verbunden.

Aussagekraft des QI

Die Vollständigkeit der Registerdaten ist sehr wichtig für die Nutzung der Daten für aussagekräftige Auswertungen, entsprechend wichtig ist die Überprüfung des Datenbestandes auf diese Eigenschaft anhand des QI. Aufgrund der geringen Menge an abgleichbaren Datenelementen im Hinblick auf den jeweiligen gesamten Datensatz der beiden Transplantationsregister ist die Aussagekraft der Berechnung jedoch vermindert.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 55.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
-	--	+	Datenverifizierung

Tabelle 55: Bewertung TMF-1046 - Vollständigkeit der Registerdaten

TMF-1047 - Übereinstimmung mit Verfahrensregeln

Register zu spezifischen Fragestellungen folgen einem eigens definierten Studienplan, in dem die wissenschaftliche Fragestellung des Vorhabens festgehalten wird. Dieser QI beschreibt die Anzahl der Abweichungen vom Studienplan in einem Datenbestand über der Anzahl der überprüften Teilnehmer.

Anwendbarkeit

Für die drei Register existieren jeweils Studienpläne, welche den Basisdatensatz des Registers umfassen.

Kriterien für die Aufnahme in die beiden Transplantationsregister sind eine Transplantation nach 1995 sowie ein Alter bei Transplantation von unter 18 Jahren. Technisch überprüft wird lediglich ob die Transplantation nach 1990 stattfand.

Die Datensätze der beiden Transplantationsregister umfassen sowohl die Definition der Datenelemente als auch die jeweiligen Erhebungszeitpunkte. Die Erhebungszeitpunkte werden automatisch vorausberechnet, die jeweilige Visite kann in einem Zeitraum um diesen Zeitpunkt stattfinden. Gleichzeitig bedingt der Basisdatensatz keine Untersuchungen, welche explizit nur für die Datenerhebung des Registers durchgeführt werden müssten. Somit können im Basisdatensatz der beiden Transplantationsregister keine Abweichungen vom Studienplan abseits der beiden genannten auftreten.

Diese beiden Register bieten zudem die Möglichkeit, Studien einzubringen, welche einen zusätzlichen Studiendatensatz retro- oder prospektiv zum Basisdatensatz erheben. Im Falle von retrospektiven Studien werden, gleichermaßen wie beim Basisdatensatz, keine Datenelemente erhoben, welche nicht aus der bereits existierenden Dokumentation der Einrichtung extrahiert werden können. Wesentliche Datenelemente des jeweiligen Studienplans sind sowohl bei retro- als auch prospektiven Studien Pflichtfelder und können somit nicht leer gelassen werden. Sollten zu einem Patienten nicht alle benötigten Informationen vorhanden sein, würde dieser die Einschlusskriterien des Studienplans verletzen und somit nicht in die Auswertung aufgenommen werden.

Für diese zusätzlichen Studien müssen die Patienten durch die jeweilige Einrichtung als Teilnehmer markiert werden. Hierbei findet keine Überprüfung statt, ob der Patient die Kriterien der Studie auch tatsächlich erfüllt. Die Kriterien werden erst manuell überprüft, wenn Datenqualitätsmanager die Studiendaten im Rahmen des Datenqualitätsmanagementprozesses sichten. Patienten können daher zunächst auch mit Abweichungen vom Studienplan Teil der Studie sein.

Der Basisdatensatz des KRhOKo-Registers besteht nur aus Datenelementen, welche sich aus der existierenden Dokumentation der Einrichtung ergeben. Für bestimmte Diagnosen gibt es zusätzlich die Möglichkeit weitere Datenelemente auszufüllen. Diese Datenelemente sind optional und orientieren sich an Untersuchungen, welche im Falle der gestellten Diagnose normalerweise durchgeführt werden sollten. Der Studienplan des KRhOKo-Registers

ist zudem sehr weit ausgelegt um alle Fälle zu erfassen, Einschlusskriterien sind ein Geburtsdatum frühestens im Jahre 1978 und eine rheumatologische Hauptdiagnose. Das Geburtsdatum wird bereits beim Anlegen des Patienten überprüft und gegebenenfalls der Eingang in das Register verhindert. Gleichmaßen ist es nicht möglich, Visiten ohne eine rheumatologische Hauptdiagnose in Auswertungen einzubringen. Aufgrund dessen sind keine Abweichungen zum Studienplan zu berechnen.

Implementierbarkeit

Die Überprüfung der Einschlusskriterien muss für jeden Studienplan eigens programmiert werden, da sich die Datensätze und Einschlusskriterien der Studienpläne unterscheiden. Je nach Umfang des Studienplan kann der Aufwand zur technischen Umsetzung der Berechnung daher schwanken. Aufgrund der Unterschiede der Studienpläne wird die Erfassung von Abweichungen nicht verallgemeinerbar sein, da unterschiedliche Datenelemente zu berücksichtigen sind. Eine technische Umsetzung der Berechnung ist automatisiert möglich, bedingt aber für jede Studie neue Programmteile und somit erhöhten Implementierungsaufwand.

Aussagekraft des QI

Je nach Schwere der festgestellten Abweichungen vom Studienplan können diese zum Ausschluss des Patienten von der Studie führen, da die Vergleichbarkeit der Daten gegebenenfalls nicht mehr gewährleistet werden kann. Der QI ist somit wesentlich für die Durchführung von Studien in den beiden Transplantationsregistern. Die TMF-Leitlinie bietet eine alternative Formel zur Berechnung des QI an. Diese Berechnung betrachtet den Anteil an Patienten mit Abweichungen an allen überprüften Patienten. Hierdurch scheint eine bessere Aussagekraft gegeben zu sein da ersichtlich wird, wie viele Patienten tatsächlich Abweichungen aufweisen und somit deutlich wird, wie groß der Anteil der Patienten ist die nicht Teil der Studie sein sollten.

Die zusammenfassende Bewertung des QI befindet sich in Tabelle 56.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
+	-	+ +	Report

Tabelle 56: Bewertung TMF-1047 - Übereinstimmung mit Verfahrensregeln

TMF-1048 - Repräsentativität der Registerdaten

Der QI beschreibt die Repräsentativität der Stichprobe der Registerpopulation im Hinblick auf die zu untersuchende Gesamtpopulation anhand von Einflussgrößen wie bekannten Inzidenzen oder Prävalenzen. Der QI ist dabei nur relevant, wenn eine Subgruppe der Gesamtpopulation untersucht werden soll, und nicht die Gesamtpopulation selbst.

Da alle drei CERTAIN-Register zum Ziel haben, jeweils die gesamte Zielpopulation zu erfassen, ist der QI nicht relevant. Auf eine abschließende Bewertung, Tabelle 57, wird daher verzichtet.

Die TMF-Leitlinie betont, dass der QI nur anzuwenden sei, wenn lediglich eine repräsentativer Anteil der Zielpopulation und nicht die gesamte Zielpopulation erfasst werden soll. Dem QI wurde nach Gesprächen mit dem medizinischen Personal der Registerzentralen eine große Bedeutung zugemessen, da die Repräsentativität ein Faktor für die Verallgemeinerbarkeit der Aussagen ist, die anhand der Registerdaten getroffen werden.

Anwendbarkeit	Implementierbarkeit	Aussagekraft des QI	QI-Kategorie
			Report

Tabelle 57: Bewertung TMF-1048 - Repräsentativität der Registerdaten

6.4 Zusammenfassung der Analyse der QI

Auf der Basis der Analyse der QI für die drei CERTAIN-Register wurden mehrere QI aus der Menge der für die drei CERTAIN-Register relevanten QI gestrichen. Die für die Berechnung dieser QI notwendigen Daten wurden nicht erhoben und somit können diese QI nicht berechnet werden. Zunächst wurden einige vorhabensspezifischen QI aufgrund mangelnder Anwendbarkeit gestrichen. Dies betraf die QI „DCO-Rate (Death Certificate Only)“, „Einzelmeldungen pro Beobachtungseinheit“ sowie „Solitäre Meldungen von Pathologen“, da diese QI nur auf Tumorregister anwendbar sind. Alle QI zum Themenkomplex „Verweigerungen“ wurden gestrichen, da Verweigerungen nicht dokumentiert werden und die QI somit nicht berechnet werden können. Somit wurden die QI zu Verweigerungsraten von einzelnen Datenelementen, Modulen und Untersuchungen gestrichen (TMF-1033, TMF-1032, TMF-1031). Ebenso wurden vier QI zum Themenkomplex „Werteverteilungen“ gestrichen, da die benötigten Daten nicht erhoben werden (TMF-1009, TMF-1010, TMF-1011, TMF-1019). Gestrichen wurden zudem die QI „Repräsentativität der Registerdaten“, da die zu rekrutierende Population der Zielpopulation entspricht, sowie „Umfang der Metadaten bei Untersuchungen“, da keine Metadaten zu Untersuchungen erfasst werden. Ebenso wurde auf die QI der QI-Kategorie „Datenverifizierung“ verzichtet, da hierfür in den drei Registern keine Ressourcen zur Verfügung stehen. Im Rahmen einer Studie des CERTAIN-Registers werden zwar Besuche bei Einrichtungen durchgeführt, die Überprüfung der Originaldaten beschränkt sich jedoch auf für diese Studie relevante Daten und erreicht nicht den Umfang, der für die Berechnung der QI notwendig wäre.

Im KRhOKo-Register entfallen zudem die QI „Endlosüberlebende“ und „Datenquellen pro Beobachtungseinheit“, da das Ableben der Patienten nicht erfasst wird und keine weiteren Datenquellen zur Verfügung stehen. Ebenso stehen im CERTAIN-LI-Register keine weite-

ren Datenquellen zur Verfügung so dass der QI „Datenquellen pro Beobachtungseinheit“ auch dort keine Beachtung findet.

Auf der Basis der Analyse der QI wurden diesen Prioritäten innerhalb der QI-Kategorien zugewiesen. Die Prioritäten von 1-3 basieren auf den vergebenen Bewertungen hinsichtlich Anwendbarkeit, Implementierbarkeit und Aussagekraft des jeweiligen QI. Bei der Priorisierung der QI wurde zudem auf eine schlüssige Priorisierung innerhalb der QI-Kategorie geachtet und die Prioritäten der QI einer QI-Kategorie gegeneinander abgewogen. Aufgrund der Ausrichtung des Datensatzes des KRhOKo-Registers auf qualitative Datenelemente wurden die Prioritäten einzelner QI herabgesetzt. Betroffen hiervon waren QI deren Berechnung nur für numerische Datenelemente sinnvoll ist, da hiervon im KRhOKo-Datensatz nur wenige Datenelemente existieren (TMF-1006, TMF-1052).

Die Analyse der QI für die drei CERTAIN-Register wird zusammengefasst in Anhang A Zusammenfassung QI-Analyse der CERTAIN-Register dargestellt.

6.5 Gewichte der QI für die CERTAIN-Register

In den drei CERTAIN-Registern wird gänzlich auf die Berechnung der QI der Kategorie „Datenverifizierung“ aufgrund des damit verbundenen Aufwandes verzichtet. Eine Gewichtung der QI dieser Kategorie ist daher nicht erforderlich. Da in den drei Registern nur die Kategorie „Report“ für die Berechnung des Gesamtscores zu berücksichtigen ist, können die bereits vergebenen Prioritäten direkt als Gewichte zur Berechnung des Gesamtscores genutzt werden. Weitere Schritte zur Feststellung der Gewichte sind daher nicht erforderlich.

6.6 Vorlage für elektronische Register

Auf der Basis der durchgeführten QI-Analyse und der anschließenden Priorisierung für die CERTAIN-Register wurde eine allgemeine QI-Empfehlung für elektronische Register entwickelt und als tabellarische Vorlage vorgelegt, siehe Anhang B QI-Empfehlung für elektronische Register. In dieser Vorlage werden Prioritäten für alle QI der Leitlinie vorgeschlagen. Den Nutzern der Vorlage steht es hierbei frei, die Anwendbarkeit, Implementierbarkeit und Aussagekraft der einzelnen QI für ihre Register festzulegen, diese in die Vorlage einzutragen und auf Basis dieses Zwischenergebnisses die QI selbst zu priorisieren. Auf der Basis der Prioritäten können dann die registerspezifischen Gewichte für die Berechnung des Gesamtscores festgelegt werden.

Das in Kapitel 5 Vorgehen zur Verbesserung der Datenqualität in Registern in Abbildung 11 dargelegte allgemeine Vorgehen wurde daher erweitert, um Anwendern aufzuzeigen, wie die in diesem Kapitel erfolgte Analyse der QI und die erstellte Vorlage in das allgemei-

ne Vorgehen mit einbezogen werden können. Das erweiterte allgemeine Vorgehen wird in Abbildung 12 dargestellt.

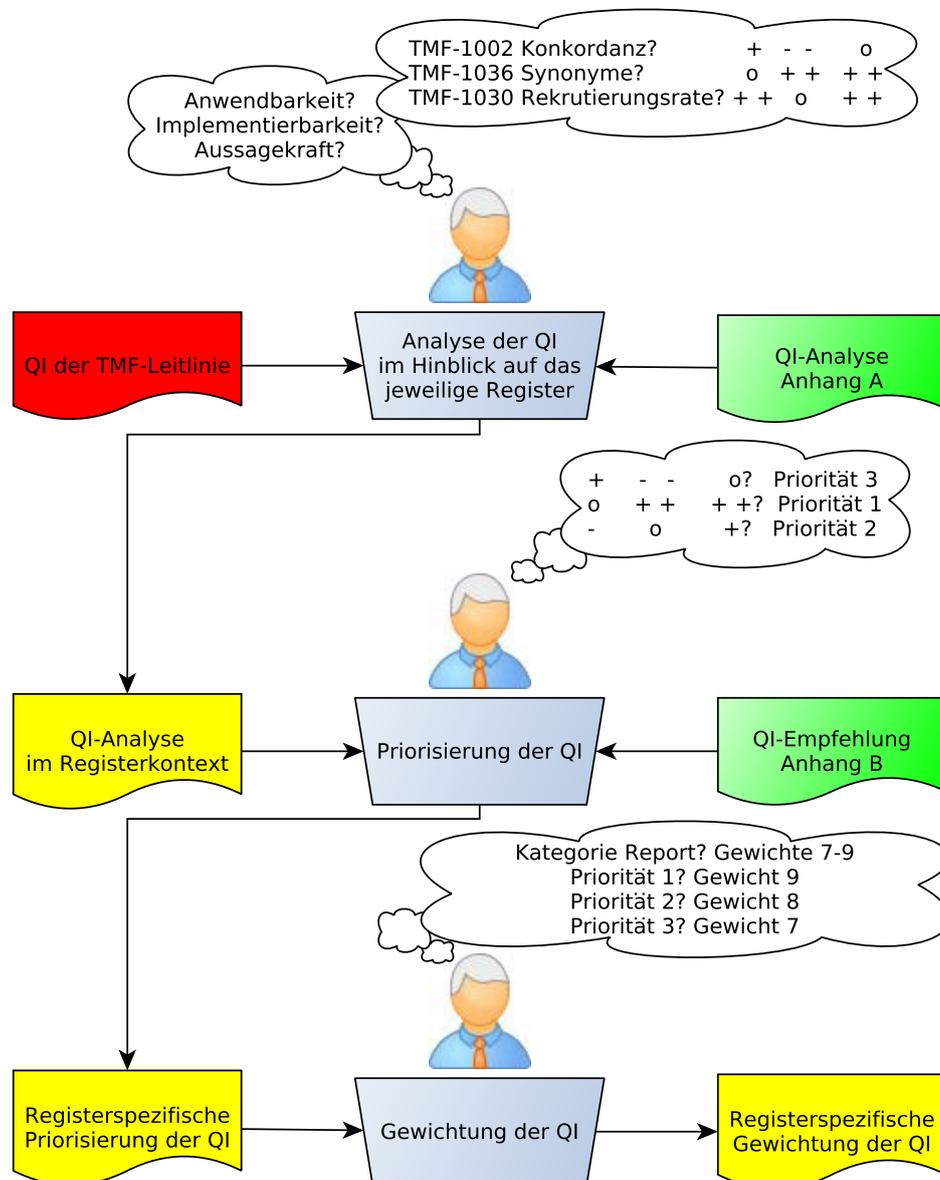


Abbildung 12: Erweitertes allgemeines Vorgehen

7 Implementierung

Die entwickelten, registerspezifischen Prioritäten der QI wurden als Basis genutzt, um die QI zu identifizieren, die in allen drei CERTAIN-Registern einen ähnlich hohen Stellenwert zur Sicherung der Datenqualität einnehmen. Anschließend wurden diese QI entweder ganzheitlich in neuen Software-Komponenten in Form von Java-Modulen umgesetzt oder in Java-Modulen, die einen wichtigen Aspekt der Datenqualität der drei Register und einen Teilaspekt eines der identifizierten QI abdeckten. Ausgehend von den im Kapitel 6.2 Technische Anforderungen festgestellten technischen Anforderungen wurden die neuen Java-Module auf Basis von Java Reflection und des Hibernate Validators entwickelt. Die Neuentwicklungen wurden mit dem Ziel verfolgt, unabhängig vom jeweiligen Datenmodell der Register, die gleichen Java-Module verwenden zu können. Die bisherige Validierung sollte dabei entweder erhalten bleiben oder durch verbesserte Funktionen ersetzt werden.

7.1 Auswahl der umzusetzenden Validierungsregeln

Die bisherige Validierung der CERTAIN-Registerplattform deckte bereits einige der Aspekte der QI ab, die der Kategorie „Automatische Validierung“ der QI-Empfehlung zugeordnet wurden. Fehlende Werte bei mandatorischen Datenwerten sowie Modulen wurden bereits durch die Nutzung der Annotationen *@NotNull* und *@NotEmpty* des Hibernate Validators geprüft. Für die Prüfung auf fehlende Datenwerte bei Datenelementen, die nur in Folge von bestimmten Angaben angezeigt werden, werden *ScriptAsserts* des Hibernate Validators genutzt. *ScriptAsserts* werden zudem zur Prüfung der Widerspruchsfreiheit sowie zur Feststellung sicherer Fehler zur Eingabezeit genutzt. *ScriptAsserts* bedingen hierbei jedoch für die Prüfung jeder Konstellation von Datenelementen eine eigene, aufwändige Programmierung und eignen sich zumeist nur für die Prüfung von Sachverhalten innerhalb einer Klasse. Zur Prüfung komplexerer Sachverhalte auf sichere Fehler oder Widersprüche existieren daher eine Reihe von eigens entwickelten *ConstraintValidator*, die einen spezifischen Aspekt des Datensatzes überprüfen. Das Vorhandensein von Primärerkrankungen im Hinblick auf den QI „Beobachtungseinheiten mit unbekanntem Primärtumor“ wird bereits ebenfalls zur Eingabezeit durch die Markierung als Pflichtfeld überprüft.

Im Hinblick auf die zur automatischen Validierung identifizierten QI verblieben daher verschiedene Ansatzpunkte zur Implementierung neuer Validierungsregeln. Zum einen die Verbesserung der Erkennung von sicheren Widersprüchen und Fehlern durch neue *ConstraintValidator*, die bisher ungeprüfte Sachverhalte überprüfen und einen größeren Geltungsbereich umfassen sollten als die bisherigen *ScriptAsserts* und *ConstraintValidator*. Zum anderen die Umsetzung der QI „Ausreißer bei stetigen Datenelementen“ und „Übereinstimmung mit Vorwert“ als Teil der automatischen Validierung. Die Möglichkeit der Feststellung von fehlenden Werten bei optionalen Datenelementen wurde ebenfalls als sinnvoll

erachtet. Als wichtig wurde auch die Erweiterung der Möglichkeiten des Feedbacks an die Benutzer erachtet um Warnungen im Sinne der QI „Mögliche Widersprüche/Warnungen“ zu ermöglichen.

7.2 Etablierung von Warnungen

Für die Nutzung von Warnungen im Hinblick auf den QI „Mögliche Widersprüche/Warnungen“ wurden die Validierung und Verarbeitung der Validierungsnachrichten der Registerplattform um die Möglichkeit der Anzeige von Warnungen erweitert. Warnungen werden nach den vorgenommenen Erweiterungen durch die Angabe des Strings „#WARN“ innerhalb der zurückgegebenen Validierungsnachricht signalisiert. Die festgestellten Verstöße gegen die Validierungsregeln der durchgeführten Validierung werden durch die Register-Software im Hinblick auf diesen String gefiltert. Fehler und Warnungen werden der Eingabekraft in unterschiedlichen Kategorien in der Weboberfläche angezeigt. Warnungen können dabei stets durch die Eingabekräfte ignoriert werden. Den Eingabekräften ist es daher möglich, die mit Warnungen behaftete Visite in den Zustand „Freigegeben“ zu versetzen, um diese durch den DQM prüfen und akzeptieren zu lassen. Durch die Einbettung des Strings „#WARN“ in die Fehlernachricht können auch die *ScriptAsserts* des Hibernate Validators genutzt werden um mögliche Widersprüche festzustellen, da die Fehlernachricht der *ScriptAsserts* durch den Programmierer festgelegt werden kann. Warnungen konnten durch die Einbettung in die Validierungsnachricht auch in die bestehenden Validierungsregeln genutzt werden.

7.3 Nutzung von Annotationen und Java Reflection

Die zur Abdeckung der QI der Kategorie „Automatische Validierung“ neu entwickelten *ConstraintValidator* bauen nahezu alle auf demselben entwickelten Grundkonzept auf. Dieses Konzept sieht die Markierung von zu validierenden Objekten durch eine *Class-level*- oder *Field-level*-Annotation vor. Eine solche Annotation bewirkt im Falle der *Class-level*-Annotation die Validierung jedes Objektes, das eine Instanz dieser Klasse darstellt. *Field-level*-Annotationen bewirken die Validierung des Objektes, das innerhalb des annotierten Feldes enthalten ist. Wenn der Hibernate Validator eine der beiden Annotationen während der Validierung eines Objektes antrifft, wird der mit der Annotation verbundene *ConstraintValidator* angesprochen. Der *ConstraintValidator* führt die in seiner Programmierung festgelegten Plausibilitätsprüfungen der zu validierenden Objekte durch.

Die neuen Validatoren sollten nicht an ein spezifisches Datenmodell gebunden sein, sondern die zur Validierung benötigten Informationen unabhängig von Klassenstrukturen aus den zu validierenden Objekten entnehmen können. Daher wurde ein Annotations-basierter Ansatz verfolgt. Im Rahmen dieses Ansatzes werden zur Validierung des jeweiligen Ob-

jekt benötigte Daten mit Annotation auf *Field-level* gekennzeichnet. Jeder entwickelte *ConstraintValidator* besitzt hierzu Annotationen, die für ihn spezifisch sind. Der *ConstraintValidator* durchsucht während der Validierung das zu validierende Objekt nach diesen spezifischen Annotationen und extrahiert aus den annotierten Feldern die benötigten oder zu validierenden Daten. Auf Basis dieser Daten wird das zu validierende Objekt gegen die in der *Class-level*- oder *Field-level*-Annotation enthaltenen Rahmenbedingungen validiert. Zur Realisierung dieses Ansatzes wurden 39 Hilfsmethoden entwickelt, um das gewünschte Verhalten der entwickelten *ConstraintValidator* durch Reflection realisieren zu können. Unter anderem durchsuchen diese Methoden rekursiv die Klassenstruktur der zu validierenden Objekte und beziehen auch die Klassen der deklarierten Felder des zu validierenden Objektes in die Suche mit ein. Durch die entwickelten Methoden wurde es auch möglich, ein Datenelement auf Basis seiner Position in der Klassenstruktur des zu validierenden Objektes aus allen anderen Objekten desselben Typs zu extrahieren. Sollte ein mit einer *Field-level*-Annotation annotiertes Feld gefunden werden, wird die Position dieses Feldes innerhalb der baumartigen Klassenstruktur des zu validierenden Objektes durch die durchlaufenen Knoten des Baumes identifiziert. Der vorherige Datenwert kann dann einfach, durch das Durchlaufen des festgestellten Pfades in der vorherigen Visite, gefunden, extrahiert und in die Validierung einbezogen werden. Aus Gründen der Performanz durchsuchen die entwickelten Hilfsmethoden stets nur Klassen die Teil der Klassenstruktur der Registerplattform sind. Dies wird über die Package-Struktur der jeweiligen Klasse überprüft. Hierdurch werden nur die Klassen auf die *ConstraintValidator*-spezifischen *Field-level*-Annotationen hin überprüft, welche diese auch tatsächlich enthalten können. Vorgänger-Visiten können in den neu entwickelten *ConstraintValidator* aus der Datenbank geladen werden. In den bisherigen *ConstraintValidator* stand diese Funktionalität aus technischen und Gründen der Performanz nicht zu Verfügung. Die Etablierung dieser Funktionalität ermöglicht die Durchführung von visitenübergreifenden Plausibilitätsprüfungen.

7.4 Verhinderung von Synonymen

In den drei Registern ist es aufgrund der verwendeten Datenbankschemata nicht möglich, Synonyme für Patientenidentifikatoren anzulegen. Dies steht im Kontrast zu den Studienidentifikatoren, für die Synonyme bisher nicht verhindert wurden. Patienten konnten in einer Studie mehrere Studienidentifikatoren zugewiesen werden. Zur Verhinderung von Synonymen für Studienidentifikatoren wurde daher in allen drei Registern auf Datenbankebene ein *unique constraint* für das Werte-Set aus Patientenidentifikator und Studie eingefügt, das die Vergabe von mehr als einem Studienidentifikator an einen Patienten innerhalb einer Studie verhindert. Synonyme konnten hierdurch gänzlich ausgeschlossen werden.

7.5 Prüfung auf Übereinstimmung mit Vorwert für variable Datenwerte

Bei der Durchsicht der Daten des CERTAIN-Registers konnten immer wieder teils stark sinkende Körpergrößen der Patienten festgestellt werden. Zur besseren Absicherung dieser Datenwerte sollten diese künftig hinsichtlich ihrer Übereinstimmung mit dem Vorwert überprüft werden.

Zur Umsetzung des QI „Übereinstimmung mit Vorwert“ in Form von automatischen Validierungsregeln wurde der *ContinuousValuesValidator* implementiert, Abbildung 13. Dieser *ConstraintValidator* überprüft bei wiederkehrenden, veränderlichen Datenelementen unplausible oder fehlerhafte Abweichungen zum Vorwert. Der *ContinuousValuesValidator* baut auf zwei hierfür etablierten Annotationen auf. Zum einen ist es *@ValidContinuousValues*, eine *Class-level-Annotation*, mit der eine Klasse annotiert werden muss, die durch den *ContinuousValuesValidator* validiert werden soll. Zum anderen ist es die *Field-level-Annotation* *@ContinuousValue*, mit der die zu validierenden Felder innerhalb der zu validierenden Klasse annotiert werden müssen. Die Annotation *@ValidContinuousValues* ist aus technischen Gründen erforderlich, da nur bei der Validierung der gesamten Visite die Datenbank-Id der Visite ausgelesen werden kann. Diese Id wird zwingend benötigt um den zugehörigen Patienten und anhand des Patienten seine vorherigen Visiten und Datenwerte identifizieren zu können.

Wenn der Hibernate Validator die Annotation *@ValidContinuousValues* prozessiert, wird die Validierung der zu validierenden Visite durch den *ContinuousValuesValidator* ausgelöst. Der *ContinuousValuesValidator* durchsucht per Reflection und Rekursion die Klassenstruktur der Klasse der zu validierenden Visite nach allen Feldern, die mit *@ContinuousValue* annotiert sind. Aus jedem gefundenen Feld wird das jeweilige Objekt extrahiert und auf Basis der in der jeweiligen *@ContinuousValue*-Annotation festgelegten maximalen Abweichungen mit dem vorherigen Datenwert verglichen. Der vorherige Datenwert wird, wie im Kapitel 7.3 Nutzung von Annotationen und Java Reflection beschrieben, per Reflection aus der Vorgänger-Visite extrahiert.

@ContinuousValue bietet die Möglichkeit maximal erlaubte Abweichungen sowohl als Fehler als auch als Warnung an den Benutzer zurückzugeben. Werte, die um mehr als die *hard deviation* vom Vorwert abweichen, werden als Fehler gewertet. Werte, die um mehr als die *soft deviation* vom Vorwert abweichen, werden als Warnungen deklariert. Es ist möglich sowohl *hard* als auch *soft* in einer *@ContinuousValue*-Annotation zeitgleich zu verwenden und zu validieren. Sollten die harten Validierungsgrenzen verletzt werden, wird nur ein Fehler angezeigt, auch wenn zusätzlich die weichen Validierungsgrenzen verletzt werden. Es kann auch nur eine Unter- oder Obergrenze definiert werden. Der *ContinuousValuesValidator* unterstützt darüber hinaus die Möglichkeit maximale Abweichungen sowohl prozentual als auch absolut festzulegen.

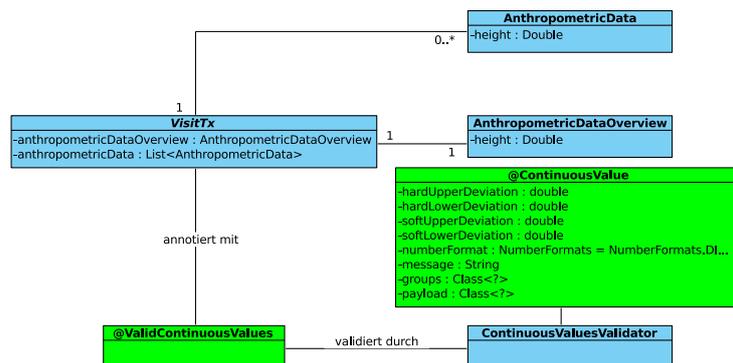


Abbildung 13: Ausschnitt Klassendiagramm ContinuousValuesValidator

Im Beispiel in Abbildung 13 wurde die Körpergröße sowohl im minimalen als auch im erweiterten Datensatz zur Validierung annotiert. Die Körpergröße darf prozentual nicht um mehr als 5 % abnehmen. Eine Warnung wird bei einer Abnahme um mehr als 3 % ausgegeben.

Durch die Annotierung der Klasse *VisitTx* werden auch alle von dieser Klasse erbenenden Klassen durch den *ContinuousValuesValidator* validiert. Für die Register CERTAIN und CERTAIN-LI sind dies beispielsweise die Klassen *Visit*, *FollowUp*, *Initial*, *DayOfDischarge* und *ExtendedFollowUp*, siehe Abbildung 4 im Kapitel 2.4.2 Datenmodell.

Stand November 2017 sind drei Felder mit *@ContinuousValue* annotiert, alle drei beinhalten die Körpergröße des Patienten und sind in den beiden Transplantationsregistern zu finden. Es findet sich jeweils ein Datenelement im minimalen Datensatz in der fortlaufenden Dokumentation des erweiterten Datensatzes sowie in einem Studienmodul. 34127 Objekte fallen in den Geltungsbereich des *ContinuousValuesValidator*.

7.6 Prüfung auf Übereinstimmung mit Vorwert für invariable Datenwerte

Während der vorherige *ContinuousValuesValidator* den Aspekt der variablen Datenwerte des QI „Übereinstimmung mit Vorwert“ abdeckt, prüft der *InvariableValueValidator* Abweichungen bei Datenelementen, bei denen keine Abweichungen erwartet werden. Dieser *ConstraintValidator* baut auf dem selben Konzept wie der vorherige *ConstraintValidator* auf. Es existiert eine *Class-level-Annotation*, welche die Klasse als Anwendungsbereich des *InvariableValueValidator* deklariert. Der *InvariableValueValidator* wiederum durchsucht die annotierte Klasse nach Feldern, die mit *@Invariable* annotiert sind. Auf Basis von Reflexion wird das jeweilige Feld in der vorherigen Visite gesucht und mit dem aktuellen Datenwert abgeglichen. Sollten sich die beiden Datenwerte unterscheiden, wird dies als Fehler zurückgegeben.

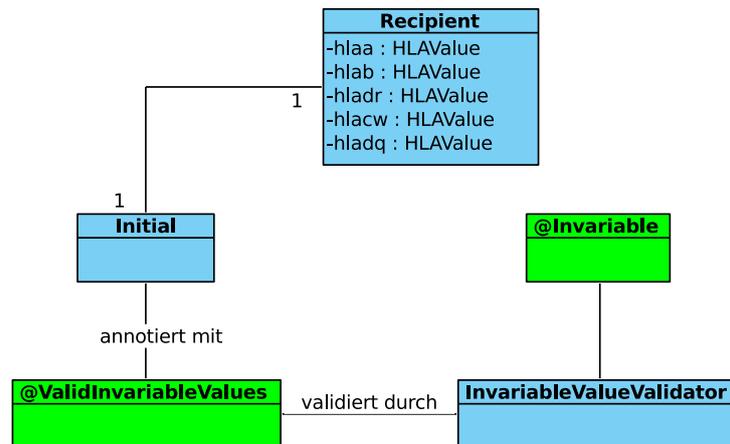


Abbildung 14: Ausschnitt Klassendiagramm InvariableValueValidator

Im Beispiel in Abbildung 14 werden die HLA-Daten des Organempfängers in der initialen Visite der Transplantation validiert. Diese Daten sollten sich im Falle einer erneuten Transplantation zwischen den Transplantationen nicht ändern. Sollte dies dennoch der Fall sein, liegt ein Datenfehler vor. Stand November 2017 findet dieser Validator bei den HLA-Werten des Organempfängers in den beiden Transplantationsregistern Anwendung. Die Überprüfung durch den *InvariableValueValidator* betrifft 10770 Objekte.

7.7 Reihenfolge der fortlaufenden Medikationsangaben

Im Zuge von Auswertungen der medizinischen Daten des CERTAIN-Registers zeigten sich immer wieder Widersprüche in den fortlaufenden Medikationsangaben des extendeden Datensatzes. Die Einträge der fortlaufenden Medikationsangaben verfügen über ein Datenelement vom Typ *MedicationEntry*, welches die Merkmalsausprägungen „Start“, „Dosisänderung“, „Stop“ und „Einzeldosis“ annehmen kann. Da bisher keine Validierung stattfand, war es beispielsweise möglich, Einträge der Abfolge „Stop“ → „Dosisänderung“ → „Stop“ zu dokumentieren. Zur Verbesserung der Auswertbarkeit sollten daher Plausibilitätsprüfungen für die Abfolgen der fortlaufenden Medikationsgaben etabliert werden.

Hierfür wurde der *EntryOrderValidator* entwickelt, der in den Geltungsbereich des QI „Sichere Widersprüche/Fehler“ fällt, da Abweichungen von den festgelegten Abfolgen durch die Registerzentralen als Fehler bewertet werden. Ziel dieses Validators ist die Sicherstellung einer plausiblen Abfolge der Medikationseinträge.

Um dieses Ziel zu erreichen war es zunächst notwendig, die als plausibel erachteten Abfolgen von Medikationseinträgen festzulegen, siehe Abbildung 15. Diese Abfolgen wurden in Gesprächen mit dem medizinischen Personal der Registerzentralen erarbeitet und beschlossen.

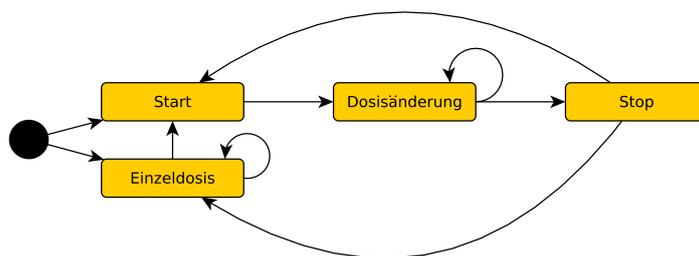


Abbildung 15: Erlaubte Abfolge Medikationseinträge

Wie die beiden vorherigen *ConstraintValidator* baut auch der *EntryOrderValidator* auf einer *Class-level*-Annotation auf, der Annotation *@ValidEntryOrder*. Der *EntryOrderValidator* durchsucht per Reflection alle Felder der annotierten Klasse auf Objekte, die vom Interface *OrderedEntry* erben. Die entwickelte Suche berücksichtigt auch alle Listen, die in der annotierten Klasse enthalten sind.

Im Beispiel in Abbildung 16 werden alle Objekte vom Typ *Drug* validiert, da diese von *OrderedEntry* erben. Das in *VisitTx* enthaltene Feld *medication* enthält hierbei alle zu dieser Visite zugehörigen Medikationseinträge. Medikationsgaben können über den zeitlichen Rahmen einer Visite hin andauern, entsprechend ist es notwendig, alle Medikationsangaben aus allen Visiten eines Patienten zur Validierung heranzuziehen. Der *EntryOrderValidator* bietet hierfür die benötigten Funktionalitäten zur Extraktion der Medikationsangaben aus den vorherigen Visiten mittels Reflection. Anschließend müssen die Medikationseinträge dem jeweiligen Medikament zugeordnet werden. Hierfür wird auf die Methode *getTradename* des Interfaces *OrderedEntry* zurückgegriffen. Durch Sortierung der Medikamente nach Handelsnamen kann der Verlauf der Medikationsgaben zu jedem Medikament nachvollzogen werden. Eine Sortierung nach Datum des jeweiligen Eintrages sichert den zeitlichen Bezug. Für jedes Medikament existiert nach Abschluss der Sortierung eine Liste, die alle Medikationsgaben zu diesem Medikament enthält, sortiert nach Datum. Diese Listen werden anschließend gegen die erlaubten Abfolgen der Medikationseinträge geprüft. Da die verabreichenden Regime mehrere Visiten umspannen können, werden in der aktuellen Visite auch Fehler in den Abfolgen der Angaben vorheriger Visiten angezeigt. Dies wurde angestrebt, da die Medikationsgaben nur bei ganzheitlicher Korrektheit auswertbar sind.

Stand November 2017 werden alle Einträge der fortlaufenden Dokumentationen der Medikationsgaben der drei Register durch den *EntryOrderValidator* validiert. Dies umfasst insgesamt 90258 Medikationseinträge.

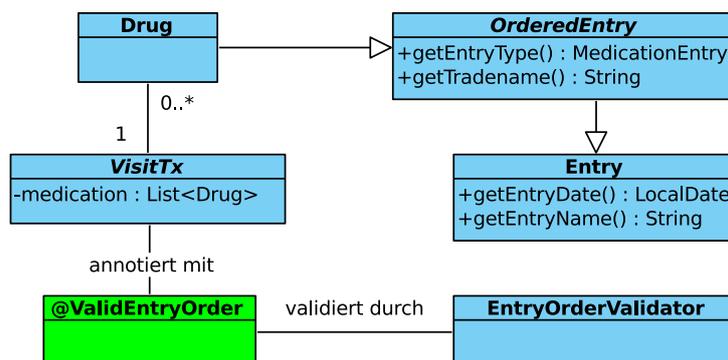


Abbildung 16: Ausschnitt Klassendiagramm EntryOrderValidator

7.8 Häufigkeit der fortlaufenden Medikationsangaben

Zur weiteren Verbesserung der Datenqualität der fortlaufenden Dokumentation von Medikationsangaben wurde zudem beschlossen, nur noch einen Eintrag pro Medikament pro Tag zu erlauben. Dies wurde als notwendig erachtet, da Medikationseinträge nur den Tag der Gabe erfassen, nicht jedoch den genauen Zeitpunkt der Gabe an diesem Tag. Hierdurch konnte bei mehreren Einträgen pro Tag nicht eindeutig festgestellt werden, in welcher Reihenfolge diese verabreicht wurden, wodurch die Daten nicht problemlos auswertbar waren. Mehrfache Gaben an einem Tag sollten künftig zu einem Eintrag mit der Gesamtdosis aller Gaben zusammengefasst werden.

Daher wurde der *SingleOrderedEntryPerDayValidator* etabliert, der einen Fehler ausgibt, sollte mehr als eine Medikationsgabe zu einem Medikament pro Tag dokumentiert sein, siehe Abbildung 17. Die Ausnahme bildet hier die Abfolge *Dosisänderung* → *Stop*, die weiterhin erlaubt wurde um Absetzungen bei Unverträglichkeit des Medikamentes aufgrund der Dosiserhöhung dokumentieren zu können. Der *SingleOrderedEntryPerDayValidator* nutzt den selben Ansatz wie der vorherige *ConstraintValidator* und basiert auf einer *Class-level*-Annotation sowie dem Interface *OrderedEntry*. Die Extraktion und Sortierung der Medikationsangaben erfolgt wie beim *EntryOrderValidator*. Gleich dem vorherigen *ConstraintValidator* werden auch durch diesen *ConstraintValidator* Fehler in den Medikationsangaben früherer Visiten festgestellt und der Eingabekraft angezeigt.

Dieser Validator fällt in den Geltungsbereich des QI „Sichere Widersprüche/Fehler“ Stand November 2017 werden alle Einträge der fortlaufenden Dokumentationen der Medikationsangaben der drei Register durch den *EntryOrderValidator* validiert. Dies umfasst insgesamt 90258 Medikationseinträge.

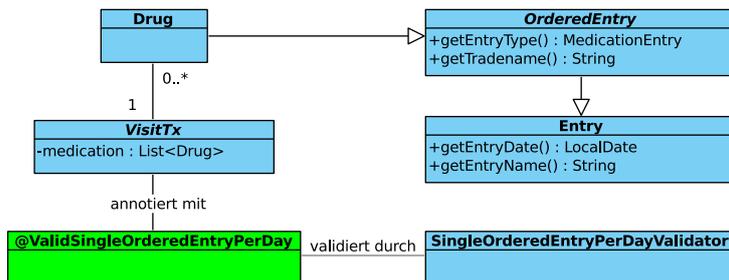


Abbildung 17: Ausschnitt Klassendiagramm SingleOrderedEntryPerDayValidator

7.9 Prüfung auf fehlende Datenwerte bei optionalen Datenelementen

Durch Anpassungen an der Behandlung der Validierungsnachrichten wurde es möglich, fehlende Werte bei optionalen Datenelementen zu bemängeln und diese Mängel in Form von Warnungen an den Benutzer zu melden. Zur Überprüfung des Vorhandenseins von optionalen Datenwerten wurden die Validatoren *NotEmptyOptionalValidator* und *NotNullOptionalValidator* entwickelt, siehe Abbildung 18.



Abbildung 18: Ausschnitt Klassendiagramm NotEmptyOptionalValidator und NotNullOptionalValidator

Der *NotNullOptionalValidator* prüft, ob ein mittels *@NotNullOptional* annotiertes Feld den Wert „null“ enthält. Ist dies der Fall, wird eine Warnung an den Benutzer zurückgegeben. Der *NotEmptyOptionalValidator* prüft gleich dem Validator *NotNullOptionalValidator*, ob ein mittels *@NotEmptyOptional* annotiertes Feld den Wert „null“ enthält. Sollte das Feld nicht „null“ sein und ein Objekt vom Typ *String*, *Collection*, *AbstractMap* oder *Array* enthalten, wird zudem überprüft, ob der String keine Zeichen enthält oder das Objekt im Falle der drei anderen Typen keine weiteren Objekte enthält. Sollte einer dieser Fälle eintreten, wird dem Benutzer eine Warnung angezeigt.

Die beiden Validatoren fallen in den Gültigkeitsbereich des QI „Fehlende Werte bei optionalen Datenelementen“. Der *NotEmptyOptionalValidator* wird zur Validierung von Körpergröße und Gewicht im KRhOKo-Register verwendet. Der *NotNullOptionalValidator* zur Validierung des Kreatinin-Wertes in einem Studienobjekt im CERTAIN-Register. Dies umfasst insgesamt 2538 Objekte.

7.10 Validierung von Wertebereichen

Im Hinblick auf die umfassenden Labor- und Medikationsangaben der beiden Transplantationsregister kam es immer wieder zu Rückfragen hinsichtlich physiologisch unmöglicher oder unplausibler numerischer Datenwerte. Weiterhin konnten für den Großteil der zu den Datenwerten zugehörigen Einheiten beliebige Einheiten angegeben werden. Hintergrund der Zulassung von beliebigen Einheitenangaben war es, die Einrichtungen nicht bei der Eingabe der Daten durch zu strikte Beschränkungen zu behindern. Beide Fehlerquellen sollten durch neue *ConstraintValidator* reduziert werden. Das Registerpersonal des CERTAIN-Registers erstellte zu diesem Zweck einen Validierungsplan für die Labor- und Medikationsangaben, der die Basis der neuen Validierung darstellen sollte. In diesem Validierungsplan wurden für die betroffenen Datenelemente die Validierungsgrenzen in Form von Ober- und Untergrenzen verbunden mit der Einheit der Grenzen festgelegt. Festgelegt wurden harte Ober- und Untergrenzen zur Verhinderung von physiologisch unmöglichen Werten, die mit dem Ableben des Patienten einhergehen würden, sowie weiche Ober- und Untergrenzen für Werte, die außerhalb des Normbereichs des Datenelementes liegen und unplausibel erscheinen. Eine Überschreitung der harten Validierungsgrenzen resultiert in einer Fehlermeldung. Sollten die weichen Validierungsgrenzen überschritten werden, wird eine Warnung ausgegeben.

Zur Prüfung dieser Grenzen wurde der *RangeValidatorBase* entwickelt, der in den beiden *ConstraintValidator* *EntryRangeValidator* und *RangeValidator* spezialisiert wurde, siehe Abbildung 19.

Der *RangeValidator* nutzt die *Field-level-Annotation* *@Range* und kann entweder zur Validierung von komplexen Objekten oder zur Validierung von Datentypen benutzt werden, die von der Java-Klasse *java.lang.Number* erben. Diese Klasse stellt die Basis-Klasse für alle Java-Klassen dar, die Zahlen darstellen, beispielsweise *java.lang.Integer* oder *java.lang.Double*.

@Range bietet die Möglichkeit harte Validierungsgrenzen festzulegen, die in einem Fehler resultieren, sowie weiche Validierungsgrenzen, die in eine Warnung münden. In beiden Fällen ist es möglich Ober- und Untergrenzen zu validieren. Es ist ebenso möglich, nur Ober- oder Untergrenzen zu validieren oder nur harte oder weiche Validierungsgrenzen zu prüfen. Sollten die harten Validierungsgrenzen verletzt werden, wird nur ein Fehler angezeigt, auch wenn zusätzlich die weichen Validierungsgrenzen verletzt werden. Weiche Validierungsgrenzen müssen dabei stets innerhalb der harten Validierungsgrenzen liegen, wenn diese angegeben sind.

Sollte nicht nur ein Feld vom Typ *Number* annotiert sein sondern ein komplexes Objekt, wird dieser Fall durch den *RangeValidator* gesondert behandelt. In diesem Fall durchsucht der *RangeValidator* das zu validierende Objekt per Reflection nach den Annotationen *@Value*, *@Unit* und *@Status*. *@Value* annotiert das Feld, das den zu validierenden Zahlenwert

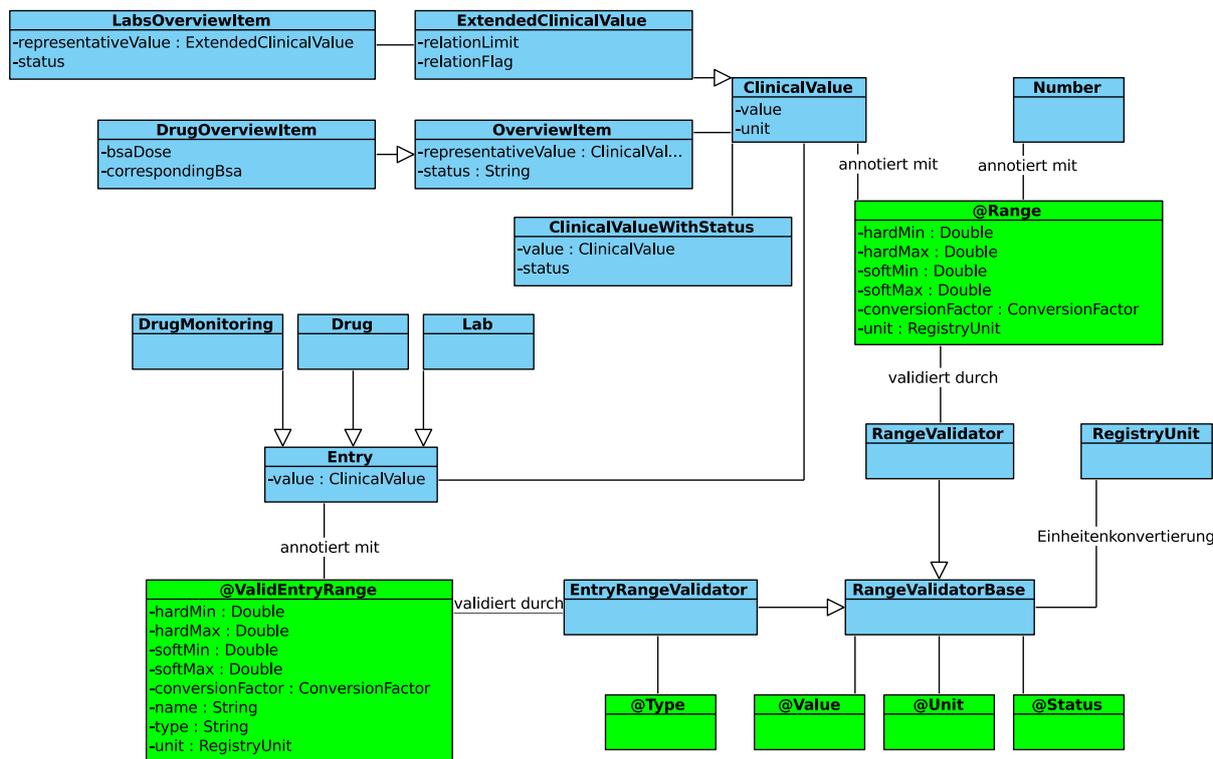


Abbildung 19: Ausschnitt Klassendiagramm RangeValidator und EntryRangeValidator

enthält. *@Unit* annotiert das Feld, das die Einheit des Zahlenwertes enthält. *@Status* annotiert das Feld, das Informationen über die Erfordernis der Validierung des Objektes enthält. Beispielsweise wird mit einem Status von „No“ signalisiert, dass das Objekt keine Datenwerte für *@Value* und *@Status* enthält und somit nicht validiert werden muss. *@Status* wird nur berücksichtigt, wenn die Annotation an einem Feld gefunden werden konnte. Sollte kein solches Feld existieren, wird die Kombination aus *@Value* und *@Unit* stets validiert.

Tupel aus Zahlenwert und Einheit werden in den drei CERTAIN-Registern in Objekten vom Typ der Klasse *ClinicalValue* gespeichert. Dieser Objekt-Typ wird in diversen Klassen verwendet, die medizinisch-relevante Zahlenwerte speichern. In der Klasse *ClinicalValue* sind die Felder *value* mit *@Value* und *unit* mit *@Unit* annotiert. Durch die Nutzung von Reflection können daher auch alle Klassen validiert werden, die einen *ClinicalValue* beinhalten. Gleiches gilt für künftige Klassen, welche nicht auf *ClinicalValue* aufbauen. Auch diese Klassen können bei Annotation eines Tupels aus Zahlenwert und Einheit mit *@Value* und *@Unit* validiert werden.

Zur Validierung von Wertebereichen in der fortlaufenden Dokumentation musste eine zusätzliche Annotation entwickelt werden, da die Klassen der fortlaufenden Annotation nur mittels einer *Class-level*-Annotation zu validieren waren. Die *Class-level*-Annotation *@ValidEntryRange* entspricht in ihrer Funktion der Annotation *@Range* und bietet die selben

Parameter zur Konfiguration der Validierung. Zur Verarbeitung der fortlaufenden Dokumentation wurde jedoch die Annotation `@Type` zu `@ValidEntryRange` hinzugefügt, um zusammengehörende fortlaufende Einträge besser für die Validierung identifizieren zu können. Dies liegt darin begründet, dass der Typ des Eintrages sich zwischen den zu validierenden Klassen *DrugMonitoring*, *Drug* und *Lab* des Beispiels in Abbildung 19 unterscheidet und daher ebenfalls per Reflection abgefragt werden muss.

In allen drei Registern ist es für Wert-Einheiten-Tupel möglich, dynamisch Einheiten einzugeben, die aus vorgegebenen Einheiten ausgewählt werden können. Oftmals ist jedoch auch Freitext für die Angabe der Einheit zugelassen und die Auswahlliste der Einheiten nicht auf zu diesem Datenelement passende Einheiten beschränkt. Die Validierung von Wert-Einheiten-Tupeln erfordert daher eine dynamische Einheitenumrechnung auf Basis des zu validierenden Objektes sowie der Einheit auf dessen Basis der Wertebereich der Validierungsgrenzen festgelegt ist. Hierfür wurde die Klasse *RegistryUnit* entwickelt, die eine Sammlung zugelassener Einheiten darstellt und zwischen diesen Einheiten Umrechnungen von Werten durch die Nutzung der Java-Bibliothek JScience⁴ vornehmen kann. Sollte ein entsprechender Konvertierungsfaktor in der `@Range`- oder `@ValidEntryRange`-Annotation angegeben sein, ist zudem die Umrechnung in die Stoffmengenkonzentration (bspw. „mmol/l“) und die Validierung auf Basis dieser Einheit möglich. Die Konvertierungsfaktoren wurden aus der Literatur entnommen und durch das medizinische Personal der CERTAIN-Registerzentrale bestätigt. In der Klasse *RegistryUnit* werden nur SI-Einheiten zur Umrechnung angeboten. Sollten durch `@EntryRangeValidator` oder `@RangeValidator` zu validierende Objekte eine Einheit enthalten, die nicht mit der Validierungseinheit kompatibel ist, wird dies als Warnung an den Benutzer zurückgegeben. Der Datenwert des zu validierenden Objektes wird in diesem Fall nicht gegen die Validierungsgrenzen validiert, da eine Umrechnung nicht möglich ist.

Die *ConstraintValidator EntryRangeValidator* und *RangeValidator* fallen in den Geltungsbereich der QI „Sichere Widersprüche/Fehler“, „Mögliche Widersprüche/Warnungen“ und „Ausreißer bei stetigen Datenelementen“. Der *EntryRangeValidator* wird zur Plausibilitätsprüfung von 89 Eintragstypen von Medikation und Labor genutzt. Dies umfasst 201780 Objekte. Der *RangeValidator* wird zur Validierung von 97 Datenelementen genutzt. Dies umfasst 1122773 Objekte.

Mit der Etablierung der `@Range`- und `@ValidEntryRange`-Annotation und der darin enthaltenen Validierungseinheit wurde es möglich, die annotierten Labor- und Medikationsangaben während des Exports von Daten aus dem Register in die jeweilige Validierungseinheit umzurechnen. Hierfür wurden die Exportfunktion der Register diesbezüglich angepasst. In diesem Schritt wurde für Medikationsangaben zusätzlich eine automatische Berechnung der gegebene Dosis pro Kilogramm Körpergewicht etabliert und der berechnete Wert dem

⁴<http://jscience.org/>

Export hinzugefügt. Beide Anpassungen wurden vorgenommen um die Auswertbarkeit der Daten zu verbessern und Synergieeffekte der neuen Validerung zu nutzen.

7.11 Aktualität der gespeicherten Daten

Die Aktualität der gespeicherten Daten wurde von den Registerzentralen der Register CERTAIN und CERTAIN-LI als besonders wichtig erachtet. Daher wurde die Berechnung dieses QI realisiert, obwohl die Berechnung nicht auf KRhOKo anwendbar ist, da dort nicht die selben Endpunkte und geregelten Beobachtungszeitpunkte existieren. In Gesprächen mit den Registerzentralen von CERTAIN und CERTAIN-LI wurden die Daten der Patienten als aktuell erachtet, wenn diese bestimmte Kriterien erfüllten. Sollte der Patient lost-to-follow-up oder verstorben sein, wurden die Daten als aktuell erachtet, da keine neuen Daten zu erwarten sind. Gleichermaßen werden keine neuen Daten erwartet, wenn der Patient in der letzten dokumentierten Visite einen Transplantatsverlust erlitten hatte. Im Hinblick auf die vorausberechneten Visiten wurden die Daten als nicht mehr aktuell erachtet, wenn die berechnete Frist zur Dokumentation der aktuellsten Visite verstrichen war. Die Aktualität der gespeicherten Daten kann für jeden Patienten berechnet werden. Die durchschnittliche Aktualität wird für jede Einrichtung visualisiert in der Weboberfläche dargestellt, siehe Abbildung 20. Gleichermaßen wird der Wert in den monatlich generierten Einrichtungsreports der beiden Transplantationsregister festgehalten.

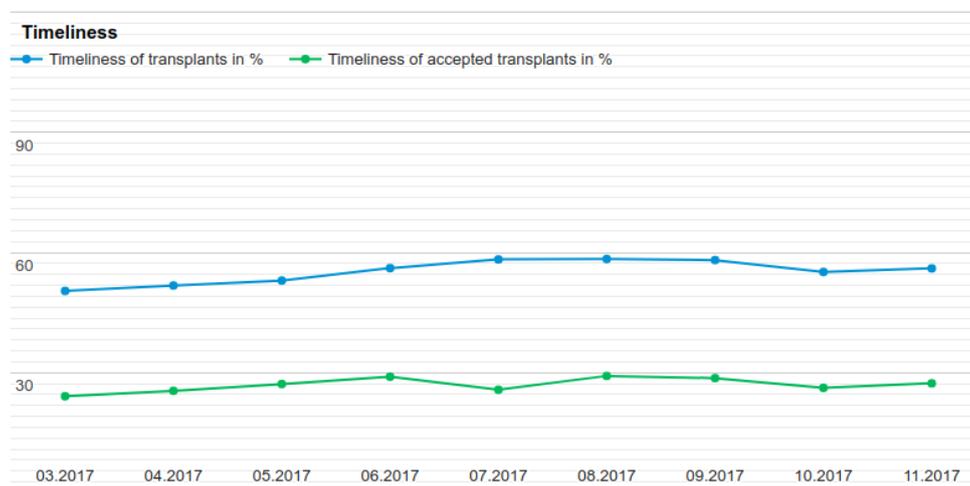


Abbildung 20: Aktualität der Daten für CERTAIN Einrichtung Heidelberg im November 2017

Für die Registerzentrale besteht zudem die Möglichkeit der Betrachtung der globalen Aktualität der gespeicherten Daten, siehe Abbildung 21.

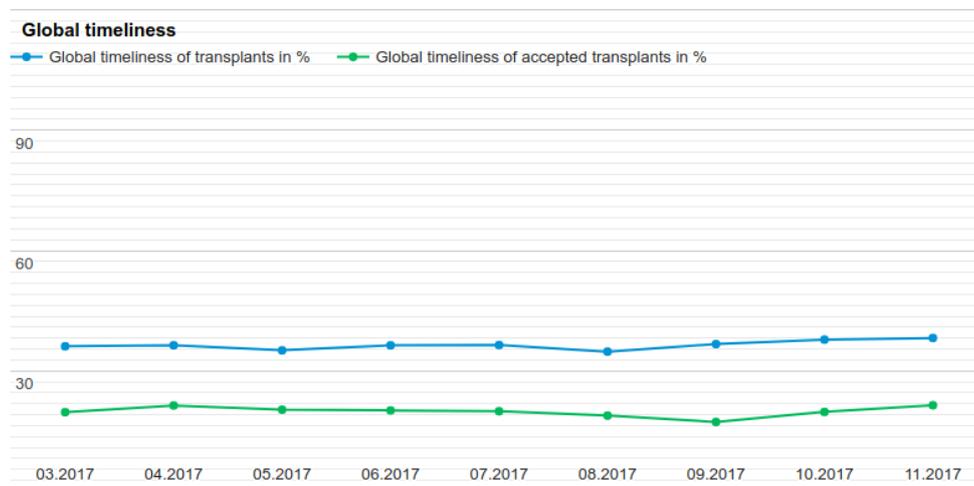


Abbildung 21: Globale Aktualität der Daten des CERTAIN-Registers November 2017

7.12 Rekrutierungsrate

Die Rekrutierungsrate wurde durch das CERTAIN-Registerpersonal als besonders wichtig erachtet, da eine komplette Vollständigkeit der Registerpopulation der teilnehmenden Einrichtungen angestrebt wird. Da eine automatisierte Berechnung für alle Einrichtungen aufgrund der unterschiedlichen Datenlage in den jeweiligen Ländern als nicht praktikabel erachtet wurde, wurde beschlossen lediglich die Rekrutierungsrate der Einrichtungen der ET-Länder zu berechnen, da für diese Einrichtungen alle Informationen zu transplantierten Patienten vorlagen. Für CERTAIN-LI und KRhOKo wurde die Umsetzung nicht angestrebt, da die Berechnung aufgrund der Notwendigkeit von externen Daten bezüglich der Inzidenzen nicht automatisiert werden konnte.

Da CERTAIN zu jeder durchgeführten Transplantation einer ET-Einrichtung eine Nachricht von ET erhält, konnten diese Informationen einfach aus dem Datenbestand extrahiert werden. Gleichermaßen konnte festgestellt werden, ob die Transplantation auf Basis der Nachricht in das Register eingegangen war, da diese Information in der erhaltenen Nachricht vermerkt wird. Abschließend wurde ein Modul entwickelt, das diese Überprüfung automatisch für alle von ET erhaltenen Nachrichten durchführt und die Rekrutierungsrate für jedes Kalenderjahr in der Einrichtungsübersicht für das Personal der Einrichtung darstellt, siehe Abbildung 22. Die generierten Graphiken enthalten zudem Informationen über die durch die Einrichtung abgelehnten Transplantationen. Eine Transplantation kann beispielsweise durch die Einrichtung abgelehnt werden, wenn der Patient bei Transplantation 18 Jahre oder älter war.

Zudem wurde für die Registerzentrale eine Gesamtübersicht der Rekrutierungsrate etabliert, siehe Abbildung 23.

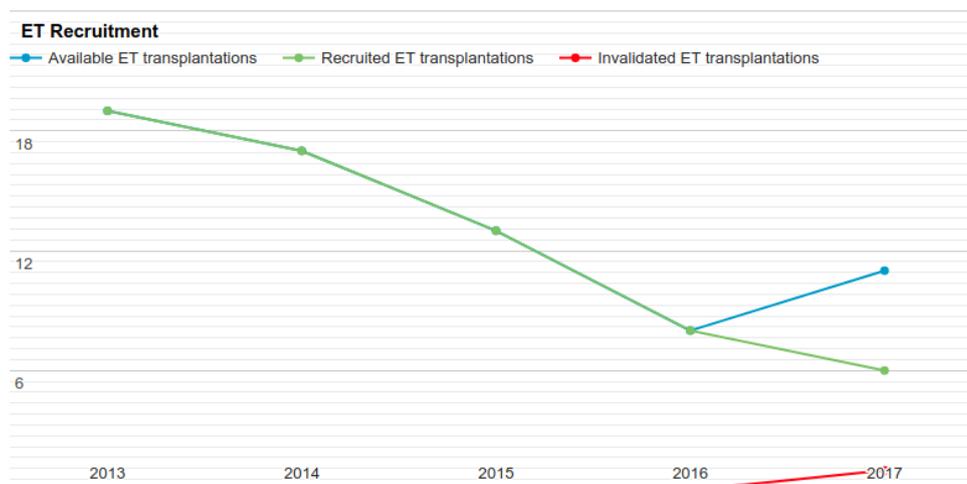


Abbildung 22: Rekrutierungsrate für CERTAIN Einrichtung Heidelberg im November 2017

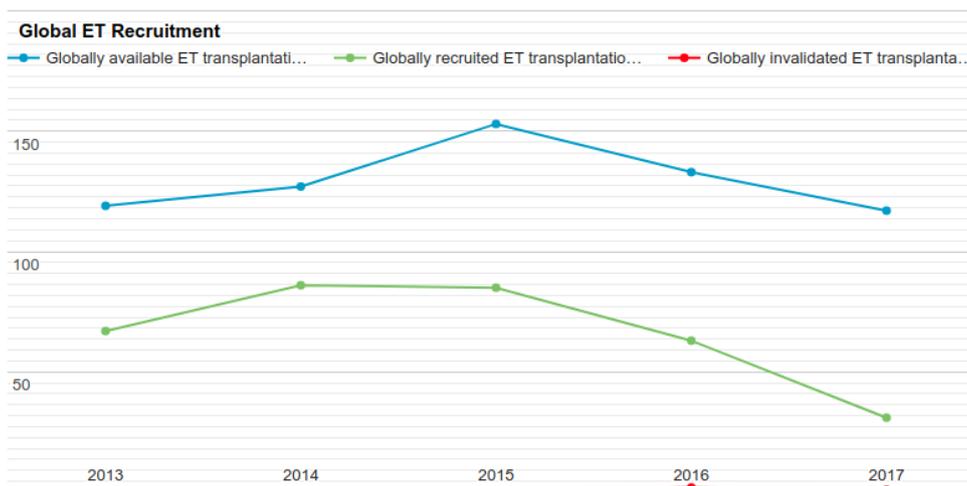


Abbildung 23: Globale Rekrutierungsrate der ET-Einrichtungen des CERTAIN-Register im November 2017

7.13 Test der Implementierung

Die entwickelten *ConstraintValidator* wurden gezielt mittels JUnit-Tests auf korrekte Funktion hin überprüft. Die testspezifischen Rahmenbedingungen der einzelnen Tests wurden mithilfe des Test-Frameworks Mockito simuliert. Insgesamt wurden für die neuen *ConstraintValidator* und die neue Einheitenumrechnung 198 Tests geschrieben.

7.14 Auswertungsmethoden

Zur weiteren Überprüfung der implementierten *ConstraintValidator* wurden mehrere Auswertungsmethoden implementiert, welche die Funktionalität der *ConstraintValidator* am Gesamtdatenbestand überprüfen sollten. Insbesondere sollte die Übereinstimmung der Validierungsgrenzen und -einheiten mit der Registerrealität geprüft werden.

Funktion `AggregateValidatedValues`

Die Funktion *AggregateValidatedValues* extrahiert alle durch *RangeValidator* und *EntryRangeValidator* validierten Datenwerte aus den übergebenen Visiten und erstellt für jedes validierte Datenelement eine eigene Datei. In dieser Datei werden die, auf die Validierungseinheit umgerechneten, Datenwerte des Datenelementes aggregiert. Es werden nur Datenwerte in die Datei geschrieben, die auf die Validierungseinheit umgerechnet werden konnten. Die Datei enthält zusätzlich die Validierungsgrenzen sowie die Validierungseinheit des Datenelementes. Die Datenwerte dieser Dateien können anschließend durch eigens entwickelte R-Skripte in Form von Boxplots und Stripcharts visualisiert werden. Die Validierungsgrenzen werden dabei in den generierten Grafiken dargestellt um die Lage der Werteverteilung innerhalb der Validierungsgrenzen einschätzen zu können. Abbildung 24 zeigt beispielhaft alle Kalium-Werte des CERTAIN-Registers im November 2017 und die zugehörigen Validierungsgrenzen der harten (rot) und weichen (orange) Validierungsgrenzen.

Funktion `Outliers`

Diese Funktion extrahiert alle Datenwerte aus durch *RangeValidator* und *EntryRangeValidator* validierten Datenelementen aus den übergebenen Visiten eines Datenbestandes. Anschließend werden diese Datenwerte nach Normalisierung auf die Validierungseinheit gegen die Wertebereiche geprüft, die in der jeweiligen *@Range* oder *ValidEntryRange* Annotation des Datenelementes vermerkt sind. Sollte ein Datenwert außerhalb der harten Validierungsgrenzen liegen, wird der Datenwert samt der zugehörigen Informationen zu Visite, Patient und betreuender Einrichtung ausgegeben. Diese Funktion sollte einen Überblick über Ausreißer des Datenbestandes geben, besonders im Hinblick auf häufiges Auftreten von Ausreißern bei einzelnen Datenelementen oder bei einzelnen Einrichtungen.

Funktion `InvestigateOutliers`

Die Funktion *InvestigateOutliers* extrahiert auf Basis der *@Range* oder *ValidEntryRange* Annotationen alle Datenwerte, die ihre harten Validierungsgrenzen verletzen. Diese Datenwerte werden zusammen mit ihrer Datenbank-ID in einer Datei gespeichert. Die Funktion bietet weiterhin die Möglichkeit, eine solche Datei einzulesen und den Zustand der in der Datei enthaltenen Datenwerte in der aktuellen Datenbank nachzuvollziehen. So kann fest-

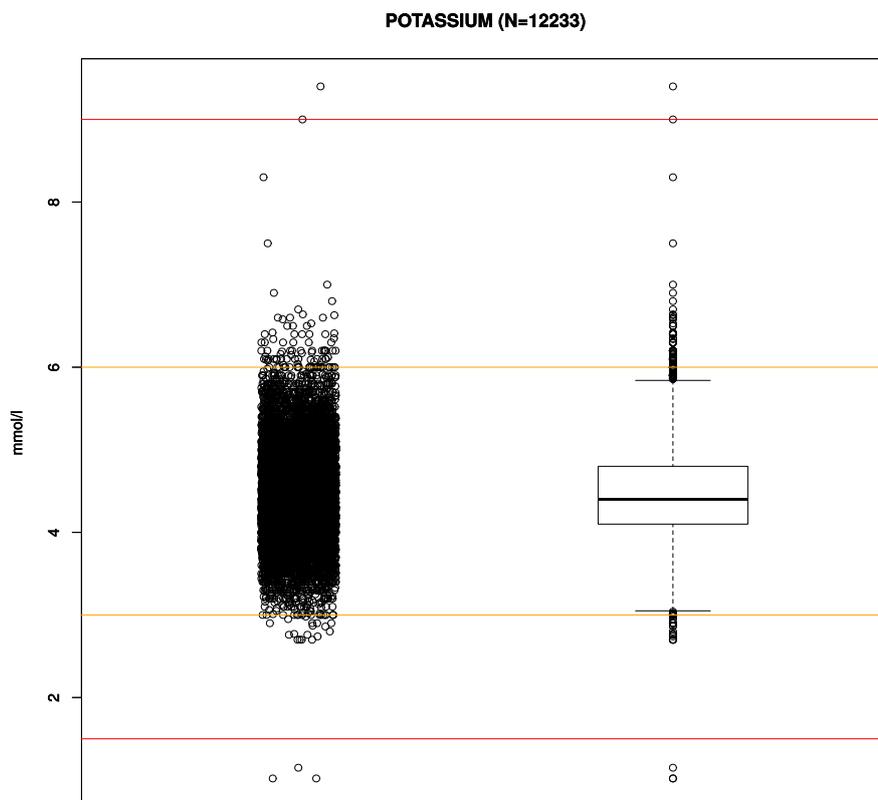


Abbildung 24: Aggregierte Kalium-Werte des CERTAIN-Registers November 2017

gestellt werden, ob und in welcher Form ein Datenwert korrigiert wurde. Beispielsweise kann festgestellt werden, ob der Datenwert gelöscht oder abgesenkt wurde, um die Fehlermeldung zu unterdrücken. Diese Funktion erstellt hierzu eine Zusammenfassung der erfolgten Änderungen an den Datenwerten der eingelesenen Datei.

StatusOfUnitsAndValues

Diese Funktion aggregiert die Datenwerte aller Datenelemente, die durch die Validatoren *RangeValidator* und *EntryRangeValidator* validiert werden. An die Funktion werden Visiten übergeben und die Klassen dieser Visiten nach Feldern oder Klassen durchsucht, welche mit *@Range* oder *ValidEntryRange* annotiert sind. Die Datenwerte jedes gefundenen Feldes und jeder Klasse werden auf Basis der Annotationen *@Value* und *@Unit* extrahiert und auf die Validierungseinheit umgerechnet, so möglich. Bei erfolgreicher Umrechnung wird der Datenwert zur Liste der Datenwerte dieses Datenelementes hinzugefügt. Auf Basis dieser Liste von Datenwerten wird die Werteverteilung des jeweiligen Datenelementes

berechnet. Abschließend werden die Werteverteilungen, Validierungsgrenzen sowie alle im Datenbestand angetroffenen kompatiblen und inkompatiblen Einheiten zu allen gefundenen, annotierten Datenelementen ausgegeben. Dies ermöglicht Betrachtungen zur Lage der Werteverteilungen innerhalb der Validierungsgrenzen sowie zum Auftreten von inkompatiblen Einheiten und deren Häufigkeit für jedes Datenelement.

Funktion StatusOfValidation

Diese Funktion ermöglicht es, alle Visiten eines Datenbestandes gegen die aktuellen Validierungsregeln zu prüfen. Die Verstöße gegen die Validierungsregeln jeder Visite werden dabei zunächst einzeln erfasst. Auf Basis dieser Verstöße gegen die Validierungsregeln werden die Verteilungen der Verstöße gegen die Validierungsregeln berechnet, sowie diese Verteilungen je nach Fehlern und Warnungen gefiltert aufgeschlüsselt. Die Funktion bietet zudem die Möglichkeit, nur die Verstöße gegen die Validierungsregeln bestimmter *ConstraintValidator* zu betrachten und die Verteilungen nur für diese *ConstraintValidator* zu berechnen. Durch diese Berechnungen können Aussagen zur Datenqualität des Datenbestandes gemacht werden, dessen Visiten die Basis der Berechnung darstellt.

8 Datenqualität des CERTAIN-Registers

Die im vorherigen Kapitel beschriebenen, neuen Validierungsregeln in Form der entwickelten *ConstraintValidator* sollten nach der Überprüfung auf ihre formale Korrektheit durch JUnit-Tests auf ihre Effektivität und Reichweite durch Validierung des Datenbestandes des CERTAIN-Registers untersucht werden. Zudem sollte untersucht werden, wie sich die erweiterte Validierung zur Eingabezeit auf die erfasste Datenqualität auswirkt. Im Rahmen der Untersuchungen zur Validierung in diesem Kapitel werden Verstöße gegen Validierungsregeln als Ereignisse bezeichnet. Schwere Verstöße gegen Validierungsregeln führen zu einer Fehlermeldung. Verstöße eines geringeren Schweregrades zu einem Warnhinweis. Visiten, bei denen Ereignisse aufgetreten sind, werden als auffällige Visiten bezeichnet. Zur Ermöglichung der differenzierteren Betrachtung der Validierungsergebnisse werden die aufgetretenen Ereignisse stets als Ereignisse pro Visite betrachtet. Hierdurch soll vermieden werden, dass Visiten mit einer hohen Anzahl an Ereignissen das Validierungsergebnis verzerren. Bei einer ausschließlichen Betrachtung der Gesamtzahl der Ereignisse kann zudem nicht entschieden werden, wie viele Visiten ohne Ereignisse Teil des Datenbestandes sind. Für die Auswertung der Datenqualität wurden mehrere Datenbestände miteinander verglichen, siehe Abbildung 10. Zum einen war es als Ausgangspunkt der Datenbestand D1 des CERTAIN-Registers R1 vom 17.11.2016. Dieser Datenbestand D1 wurde vor der Integration der Validierungsregeln V2 aus der Datenbank des Registers entnommen. Danach wurde das Register auf Version R2 aktualisiert und die Validierungsregeln V2 gingen in den Regelbetrieb des Registers ein. Nach der Aktualisierung wurde der Datenbestand D2 entnommen. Der letzte zum Vergleich herangezogene Datenbestand D3 wurde am 17.11.2017 aus der Datenbank des CERTAIN-Registers R3 entnommen. In alle Analysen gingen nur Visiten im Status „Akzeptiert“ ein, da diese Visiten den Datenqualitätsmanagementprozess durchlaufen hatten und durch die automatische Validierung sowie anschließend durch den DQM überprüft und als plausibel erachtet wurden.

Die Ergebnisse der durchgeführten Analysen werden im Folgenden beschrieben.

8.1 Datenqualität des Datenbestandes D1

Zunächst wurde der Stand der Datenqualität vor Etablierung der Validierungsregeln V2 anhand der Validierung des Datenbestands D1 gegen die Validierungsregeln V1 überprüft. Von insgesamt 14211 Visiten des Datenbestands waren 9004 im Status „Akzeptiert“. Ausschließlich diese Visiten wurden erneut durch die zu diesem Zeitpunkt aktiven Validierungsregeln V1 überprüft. 6795 dieser Visiten zeigten keine Ereignisse. Bei 2209 Visiten ergab sich durch die Überprüfung mindestens ein Ereignis. Die Aufschlüsselung der Verteilung der Ereignisse der 2209 Visiten kann der Tabelle 58 entnommen werden.

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	2209 (24,5%)	5,95 \pm 13,28	2,0 (1,0 - 167,0)	2,0	5,0
Fehlermeldungen	2209 (24,5%)	5,95 \pm 13,28	2,0 (1,0 - 167,0)	2,0	5,0
Warnhinweise					

Tabelle 58: Validierungsergebnis Datenbestand D1 geprüft gegen Validierungsregeln V1

Es wurden keine Warnhinweise ausgegeben, da es in der Version V1 der Validierungsregeln keine Validierungsregeln gab, die in diesem Ergebnis hätten resultieren können. Als wahrscheinliche Ursache der festgestellten Ereignisse wurde vom Registerpersonal des CERTAIN-Registers eine weitreichende Überarbeitung des Datenmodells des CERTAIN-Registers angegeben. Im Zuge dieser Überarbeitung wurden viele neue mandatorische Datenelemente hinzugefügt. Diese Datenelemente wurden jedoch nicht erneut durch die Eingabekräfte befüllt und führten daher zu Ereignissen.

8.2 Prüfung des Datenbestandes D1 gegen die Validierungsregeln V2

Die im Kapitel 7 Implementierung beschriebenen JUnit-Tests stellen die formale Korrektheit der entwickelten *ConstraintValidator* im Hinblick auf das angedachte Planungsszenario sicher. Die Registerwirklichkeit kann sich jedoch vom Planungsszenario in selten auftretenden Konstellationen der Daten unterscheiden. Um diese Konstellationen zu berücksichtigen und Ereignisse aufgrund unbekannter Datenkonstellationen ausschließen zu können, sollte zunächst ein kompletter Datenbestand der akzeptierten Visiten des CERTAIN-Registers testweise gegen die Validierungsregeln V2 validiert werden. Dieser Datenbestand wurde informell aus dem CERTAIN-Register entnommen und geht Datenbestand D1 voraus. Nach Abschluss der Entwicklung der Validierungsregeln V2 wurden diese gegen den zu diesem Zeitpunkt aktuellen Datenbestand des CERTAIN-Registers validiert. Dieses Zwischenergebnis soll beispielhaft durch die Validierung des Datenbestandes D1 gegen die neuen Validierungsregeln V2 der Registerversion R2 dargestellt werden.

Von 9004 akzeptierten Visiten waren nach Einführung der neuen Validierungsregeln V2 noch 1500 ohne Ereignisse. Die Überprüfung von 7504 Visiten ergab mindestens eine Fehlermeldung oder einen Warnhinweis. Die Aufschlüsselung der einzelnen Kategorien findet sich in Tabelle 59.

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	7504 (83,3%)	7,68 \pm 13,6	4,0 (1,0 - 231,0)	2,0	8,0
Fehlermeldungen	5997 (66,6%)	5,53 \pm 9,29	3,0 (1,0 - 168,0)	2,0	6,0
Warnhinweise	5081 (56,4%)	4,82 \pm 9,47	2,0 (1,0 - 169,0)	1,0	4,0

Tabelle 59: Validierungsergebnis Datenbestand D1 geprüft gegen Validierungsregeln V2

Im Vergleich mit dem Validierungsergebnis der bisherigen Validierungsregeln V1, Tabelle 58, zeigte sich ein starker Anstieg der Anzahl der Visiten mit Ereignissen. Während vor Etablierung der neuen Validierungsregeln V2 noch 75,5% der Visiten ohne Ereignisse waren, fiel diese Zahl durch die Überprüfung durch die neuen Validierungsregeln V2 auf 16,7%. Die durchschnittliche Anzahl von Ereignissen pro Visite mit Ereignissen nahm hierbei zu. Die Einführung der neuen Validierungsregeln V2 ergab zudem eine unerwartet hohe Anzahl von ausgegebenen Warnhinweisen. Auf der Basis des Validierungsergebnisses der stark gestiegenen Anzahl an Visiten mit Fehlermeldungen und der hohen Anzahl an Visiten mit Warnhinweisen wurde deutlich, dass eine genauere Betrachtung des Datenbestandes D1 und der Validierungsregeln V2 notwendig war.

Zunächst sollte hierfür eine Übersicht über alle Ausreißer erstellt werden, um die Validierungsgrenzen der Validierung der Wertebereiche des jeweiligen Datenelementes mit den Datenwerten des Datenbestands abzugleichen. Die Ausreißer wurden mittels der Auswertungsfunktion *Outliers* identifiziert. Das Ergebnis wurde einer augenscheinlichen Prüfung hinsichtlich der Häufung von Fehlermeldungen bei einzelnen Datenelementen oder Einrichtungen unterzogen. Die Durchsicht der Ausreißer zeigte, dass besonders für die Steroid-Substanzen „Prednisone“, „Prednisolone“, „Methylprednisolone“ und „Deflazacort“ Ausreißer in den Einträgen der fortlaufenden Medikationsangaben des extendierten Datensatz vorlagen. Nach Rücksprache mit dem medizinischen Personal des Registers wurde ersichtlich, dass im extendierten Datensatz Steroidstöße dokumentiert werden, die eine Normaldosis um ein Vielfaches übersteigen und somit außerhalb der initial festgelegten Validierungsgrenzen liegen. Die Validierungsgrenzen des extendierten Datensatzes wurden daher für diese vier Steroid-Substanzen auf der Basis der Rücksprache entsprechend weniger strikt formuliert, um diese legitimen Angaben zu erlauben. Der Validierungsplan wurde diesbezüglich angepasst.

Des Weiteren zeigte sich eine Vielzahl von Ausreißern für das Datenelement „pH“. Der Hauptteil dieser fehlerhaften Datenwerte entfiel auf eine Einrichtung, die für Laborwerte eine eigens für diese Einrichtung entwickelte Importfunktion auf Basis eines Dateiuploads nutzt. Durch Abgleich der festgelegten Dateistruktur einer Beispieldatei mit einer hochgeladenen Datei der Einrichtung zeigte sich, dass in der hochgeladenen Datei fälschlicherweise zwei Datenwerte statt einem als „pH“ markiert worden waren. Der falsch bezeichnete Datenwert konnte aufgrund seines Wertebereichs als pH-Wert des Urins identifiziert werden, der gemäß festgelegter Struktur als „pH urine“ hätte bezeichnet werden müssen. Diese Datenwerte konnten nicht korrigiert werden, da nicht festgestellt werden konnte, ab wann die Einrichtung die Änderung der Dateistruktur vorgenommen hatte. Die betroffene Einrichtung wurde über das festgestellte Problem informiert.

Zur besseren Einschätzung der Werteverteilungen der mit Validierungsgrenzen versehenen Datenelemente wurden alle Datenwerte der validierten Datenelemente durch die Funktion *AggregateValidatedValues* aus dem Datenbestand extrahiert und in Form von Box- und Scatterplots visualisiert. Die erstellten Grafiken wurden manuell auf Auffälligkeiten hin untersucht. So zeigte sich beim Datenelement *Protein* des extendierten Datensatzes eine unerwartete Verteilung mit dem Mittelwert der Verteilung unterhalb der unteren Validierungsgrenze, siehe Abbildung 25.

Anhand der vorab extrahierten Ausreißer konnten diese Datenwerte hauptsächlich einer Einrichtung zugeordnet werden, die eine Importfunktion für Laborwerte nutzt. Durch Überprüfung der Importfunktion konnte der Fehler auf eine während des Imports falsch zugewiesene Einheit zurückgeführt werden, „mg/dl“ statt „g/dl“. Da es sich hierbei um einen systematischen Fehler durch die Registersoftware handelte, der seit Beginn der Nutzung der Importfunktion bestand, konnten die irrtümlich zugewiesenen Einheiten ganzheitlich korrigiert werden.

In weiteren Fällen zeigten sich für einzelne Datenelemente Cluster von Datenwerten außerhalb der Validierungsbereiche. Diese Datenwerte konnten bei einrichtungsspezifischen Clustern auf falsche Einheitenangaben durch die Einrichtungen zurückgeführt werden, beispielsweise „ng/l“ statt „mg/l“. In diesen Fällen wurden die jeweiligen Einrichtungen informiert und anhand einer Liste der betroffenen Visiten und Datenelemente um Korrektur gebeten.

Neben der Überprüfung der Validierungsgrenzen war die Überprüfung der Einheitenvalidierung aufgrund der großen Anzahl an Warnhinweisen ein weiterer Punkt der durchgeführten Analysen. Mittels der Funktion *StatusOfUnitsAndValues* wurden die Werteverteilungen aller mit Validierungsgrenzen versehenen Datenelemente berechnet sowie die zur Validierungseinheit inkompatiblen Einheiten für jedes Datenelement gelistet. Bei den 9004 Visiten des Datenbestandes D1 wurden in insgesamt 317580 Objekten 18904 inkompatible Einheiten festgestellt, die durch die entwickelte Einheitenkonvertierung nicht umgerechnet werden konnten. Die neu entwickelte Einheitenkonvertierung konvertiert nur Einheiten,

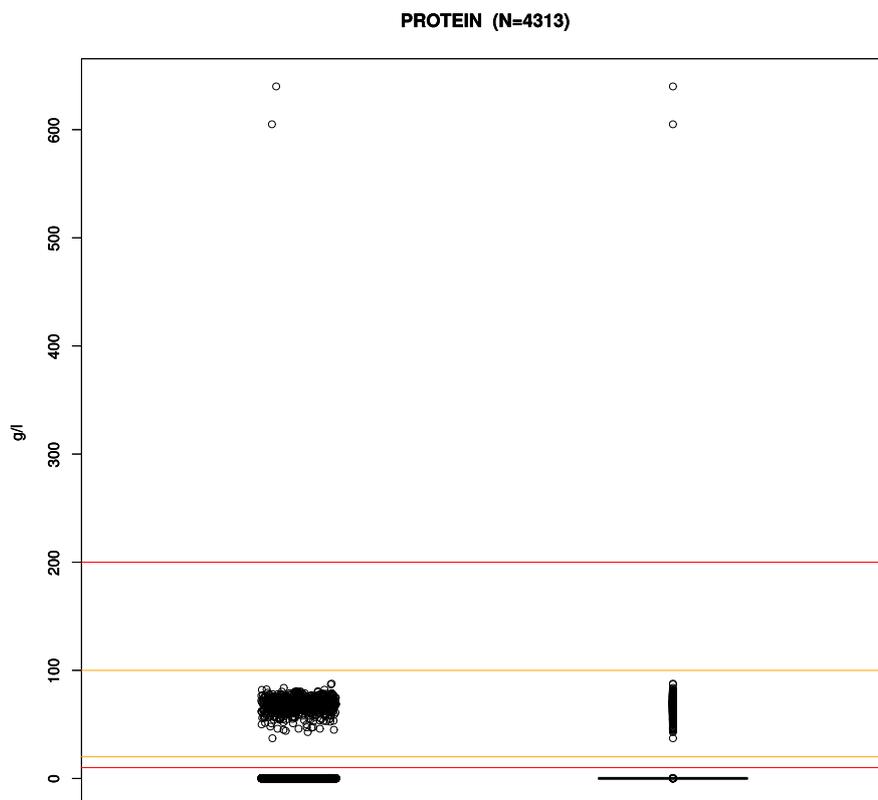


Abbildung 25: Aggregierte Protein-Werte des Datenbestandes D1

die in einer vorgegebenen Liste vorhanden sind. Nach Durchsicht der als inkompatibel bemängelten Einheiten und Rückmeldung bezüglich dieser Einheiten durch das medizinische Registerpersonal sollten einige dieser inkompatiblen Einheiten zugelassen werden. Die Liste der zugelassenen Einheiten wurde daher um einige zunächst als inkompatibel erachtete Einheiten ergänzt. Im selben Schritt zeigte sich, dass für einige Datenelemente wie „Chloride“ die Konvertierung von inkompatiblen und unerwünschten nicht-SI Einheiten wie beispielsweise „mEq/l“ in die Stoffmengenkonzentration „mmol/l“ durch eine einfache Umrechnung möglich ist. Alle durch Validierungsgrenzen validierten Datenelemente wurden daher auf mögliche Konvertierungen von inkompatiblen nicht-SI Einheiten hin überprüft und die Konvertierung nach Rücksprache mit dem medizinischen Personal des Registers für alle Datenwerte der diesbezüglich identifizierten Datenelemente durchgeführt.

Nach Durchführung der möglichen Konvertierungen verblieben dennoch einige inkompatible Einheiten. Einzelne dieser Einheiten konnten aufgrund von Tippfehlern, beispielsweise „mg7dl“ statt „mg/dl“ nicht durch die entwickelte Einheitenumrechnung konvertiert wer-

8.3 Datenqualität Datenbestand D2 nach Durchführung der Korrekturen geprüft gegen Validierungsregeln

den. Die verbliebenen, inkompatiblen Einheiten wurden daher auf mögliche Tippfehler überprüft und ein Migrationskript zur Korrektur identifizierter Tippfehler auf Datenbankebene erstellt. Die Korrekturen des Datenbankskripts wurden nach vorheriger Durchsicht durch das Registerpersonal ausgeführt. Die verbleibenden inkompatiblen Einheiten waren entweder nicht direkt in die SI-Einheit der Validierungseinheit umrechenbar oder aufgrund zu vieler Tippfehler nicht eindeutig identifizierbar.

Nach durchgeführter Korrektur waren noch 2368 von insgesamt 317580 validierten Einheiten nicht zur jeweils angegebenen Zieleinheit der Validierung der validierten Datenwerte kompatibel.

8.3 Datenqualität Datenbestand D2 nach Durchführung der Korrekturen geprüft gegen Validierungsregeln V2

Das Migrationskript der inkompatiblen Einheiten wurde zeitgleich mit der Einführung der neuen Validierungsregeln V2 in den Produktivbetrieb des CERTAIN-Registers ausgeführt. Der Stand der Datenqualität nach Etablierung der neuen Validierungsregeln V2 und Durchführung der beschriebenen Einheitenkorrekturen ist Tabelle 60 zu entnehmen.

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	7124 (79,1%)	5,3 \pm 8,53	3,0 (1,0 - 171,0)	2,0	6,0
Fehlermeldungen	5919 (65,7%)	4,99 \pm 8,85	3,0 (1,0 - 168,0)	2,0	6,0
Warnhinweise	3853 (42,8%)	2,14 \pm 1,72	2,0 (1,0 - 31,0)	1,0	3,0

Tabelle 60: Validierungsergebnis Datenbestand D2 geprüft gegen Validierungsregeln V2

Durch die vorgenommenen Korrekturen an den Validierungsgrenzen, der Liste der kompatiblen Einheiten und an den Einheiten selbst auf Datenbankebene konnte die Zahl der Visiten mit Ereignissen im Hinblick auf den Datenbestand D1, Tabelle 59, um 5,0% reduziert werden. Die Zahl der Visiten mit Fehlermeldungen fiel um 1,3%. In beiden Fällen fielen zudem die durchschnittliche Anzahl an Ereignissen pro auffälliger Visite und die Standardabweichung der beiden Verteilungen. Im Hinblick auf Visiten mit Warnhinweisen konnte deren Anteil um 24,2% reduziert werden. Die durchschnittliche Anzahl an Warnhinweisen pro Visite mit Warnhinweisen konnte mehr als halbiert werden. Die Standardabweichung fiel

ebenso ab, möglicherweise bedingt durch die Reduktion des Maximums an Warnhinweisen pro Visite von 169 auf 31.

Aufgrund von Benutzerfeedback wurden der *EntryOrderValidator* sowie der *SingleOrderedEntryPerDayValidator* auf eine genauere Unterscheidung hinsichtlich der verwendeten Substanzen und Handelsnamen der Medikamente durch eine Aktualisierung des Registers im Juli 2017 angepasst. Zeitgleich wurde die Validierung durch weitere Überprüfungen hinsichtlich der Kongruenz der Angaben von minimalem und extendiertem Datensatz erweitert. Zur besseren Vergleichbarkeit der Datenqualität der Datenbestände D2 und D3 wurde der Datenbestand D2 erneut validiert, dieses Mal gegen die Validierungsregeln V3. Das Ergebnis der Validierung der 9004 akzeptierten Visiten des Datenbestandes D2 gegen die Validierungsregeln V3 findet sich in Tabelle 61.

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
	9004				
Ereignisse	8065 (89,6%)	6,3 \pm 8,9	4,0 (1,0 - 179,0)	2,0	7,0
Fehlermeldungen	7435 (82,6%)	5,8 \pm 8,8	4,0 (1,0 - 177,0)	2,0	7,0
Warnhinweise	3826 (42,5%)	2,1 \pm 1,7	2,0 (1,0 - 31,0)	1,0	3,0

Tabelle 61: Validierungsergebnis Datenbestand D2 geprüft gegen Validierungsregeln V3

Im Vergleich zur Validierung des Datenbestandes D2 gegen die Validierungsregeln V2, siehe Tabelle 60, stieg die Anzahl der Visiten mit festgestellten Ereignissen um 13,2%. Ebenso stieg die durchschnittliche Anzahl an Ereignissen pro Visite an, während die Zahl der Visiten mit Warnhinweisen und deren durchschnittliche Zahl pro Visite nahezu unverändert blieben. Der Anstieg der Anzahl der Visiten mit Ereignissen ist auf einen Anstieg des Anteils der Visiten mit Fehlermeldungen um 25,6% zurückzuführen. Dieser Anstieg war ebenso mit einem Anstieg der durchschnittlichen Anzahl von Fehlermeldungen pro Visite bei Visiten mit Fehlermeldungen verbunden. Zurückgeführt werden kann der Anstieg der festgestellten Fehlermeldungen auf die Etablierung von weiteren *ConstraintValidator* abseits dieser Arbeit durch den Programmierer des CERTAIN-Registers. Diese *ConstraintValidator* überprüfen Zusammenhänge zwischen den Angaben des minimalen und des extendierten Datensatzes und gingen mit der Aktualisierung des Registers im Juli 2017 in den Produktivbetrieb ein. Ein weiterer Faktor für den Anstieg der Fehlermeldungen war, dass die Validierung der Wertebereiche auf weitere Datenelemente angewendet worden war.

8.4 Datenqualität ein Jahr nach Etablierung der Validierungsregeln V2

Im Zuge der vorangegangenen Auswertungen zeigte sich, dass bei 154 Visiten mehr als 20 Ereignisse auftraten. Von diesen 154 Visiten entstammten 107 aus der Heidelberger Einrichtung oder waren Teil der Dokumentation eines Patienten, der gemeinsam in Heidelberg und einer anderen Einrichtung betreut wurde. Zurückgeführt werden konnte diese hohe Anzahl an Ereignissen darauf, dass es sich bei diesen Visiten um Altdaten handelte, die zu Beginn des Registers eingegeben wurden. Seitdem durchlief das Datenmodell des Registers jedoch größere Änderungen während diese Visiten ungeachtet dessen nicht aktualisiert und somit viele fehlende Datenwerte durch die Validierung bemängelt wurden. Diese Visiten wurden von zwei Medizin-Doktoranden im Rahmen ihrer Tätigkeit als Hilfwissenschaftler des CERTAIN-Registers aufgearbeitet und die Ursachen der aufgetretenen Ereignisse so weit wie möglich korrigiert.

Zur Überprüfung der Datenqualität und der Effektivität der mit Registerversion R2 eingeführten Validierungsregeln V2 wurde im November 2017 der Datenbestand D3 aus der Datenbank des CERTAIN-Registers R3 entnommen. Dieser Datenbestand wurde durch die Validierungsregeln V3 der Registerversion R3 vom 17.11.2017 überprüft. Das Ergebnis der Validierung der 12339 akzeptierten Visiten findet sich in Tabelle 62.

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
	12339				
Ereignisse	7320 (59,3%)	4,5 \pm 6,5	3,0 (1,0 - 181,0)	1,0	6,0
Fehlermeldungen	5358 (43,4%)	4,5 \pm 7,0	3,0 (1,0 - 179,0)	1,0	6,0
Warnhinweise	4112 (33,3%)	2,1 \pm 1,7	1,0 (1,0 - 31,0)	1,0	2,0

Tabelle 62: Validierungsergebnis Datenbestand D3 geprüft gegen Validierungsregeln V3

Im vergangenen Jahr zwischen den Erhebungszeitpunkten der beiden zu vergleichenden Datenbestände D2 und D3 wurden 3335 Visiten durch den DQM in den Zustand „Akzeptiert“ versetzt. Während die Anzahl der akzeptierten Visiten damit um 37% stieg, sank die Anzahl der Visiten mit Ereignissen um 745 und somit anteilig um 33,8%. Die Anzahl der Visiten mit Fehlermeldungen sank ebenfalls um 2007 Visiten und anteilig um 47,4%. Die durchschnittliche Anzahl von Fehlermeldungen pro Visite sank dabei um etwas mehr als eine Fehlermeldung pro Visite. Die absolute Anzahl der Visiten mit Warnhinweisen stieg

um 286 leicht an. Aufgrund des Zuwachses von Visiten entspricht dies prozentual dennoch einer Verminderung um 21,6%. Es zeigte sich keine Änderung der durchschnittlichen Anzahl von Warnhinweisen bei den Visiten mit Warnhinweisen. Ebenso unverändert blieb die Standardabweichung dieser Verteilung.

8.5 Validierungsergebnis der Validierungsregeln V3

Aufgrund der Anpassungen der neuen *ConstraintValidator*, Anpassungen an den bestehenden *ConstraintValidator* und der Erstellung weiterer *ConstraintValidator* außerhalb dieser Arbeit wurden die Validierungsergebnisse der beiden Datenbestände D2 und D3 um Ereignisse aller *ConstraintValidator* bereinigt, die nicht Teil dieser Arbeit sind, siehe Tabellen 63 und 64.

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	7604 (84,5%)	4,8 \pm 4,6	5,0 (1,0 - 45,0)	2,0	6,0
Fehlermeldungen	6690 (74,3%)	4,3 \pm 4,3	3,0 (1,0 - 42,0)	2,0	5,0
Warnhinweise	3826 (42,5%)	2,1 \pm 1,7	2,0 (1,0 - 31,0)	1,0	3,0

Tabelle 63: Validierungsergebnis Datenbestand D2 geprüft gegen in dieser Dissertation entwickelte Validierungsregeln

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	6393 (51,8%)	3,8 \pm 3,9	3,0 (1,0 - 47,0)	1,0	5,0
Fehlermeldungen	3973 (32,3%)	3,9 \pm 4,0	3,0 (1,0 - 32,0)	1,0	5,0
Warnhinweise	4112 (33,3%)	2,1 \pm 1,7	1,0 (1,0 - 31,0)	1,0	2,0

Tabelle 64: Validierungsergebnis Datenbestand D3 geprüft gegen in dieser Dissertation entwickelte Validierungsregeln

Die durch *ConstraintValidator* dieser Arbeit aufgetretenen Ereignisse wurden im zwischen den Datenbeständen D2 und D3 vergangenem Jahr deutlich reduziert. Im Vergleich zum Datenbestand D2 sank die Anzahl der Visiten mit Fehlermeldungen deutlich. Im Zuge dessen sank auch die Anzahl der Visiten mit Ereignissen. Berücksichtigt werden muss bei dieser Entwicklung der Anstieg der Visiten um 37%. Der Anteil der Visiten mit Warnhinweisen sank zwar prozentual, die absolute Zahl stieg jedoch an.

8.6 Vergleich der Subgruppen der Visiten des Datenbestandes D3

Im Zuge von kontinuierlich stattfindenden Studien des CERTAIN-Registers werden die Datensätze eingeschlossener Patienten durch Studienmodule erweitert. Studienmodule erweitern die Visiten mit einem studienspezifischen Datensatz, der weitere Datenelemente enthält. Diese Studienmodule durchlaufen gleich den Visiten einen Datenqualitätsmanagementprozess, bei dem die Daten zunächst durch die Validierungsregeln und anschließend durch den DQM geprüft werden, bevor sie in Auswertungen eingehen. Gleichmaßen muss die Visite, zu der das Studienmodul selbst gehört, ebenfalls akzeptiert und frei von Ereignissen sein, bevor das ausgefüllte Studienmodul durch die Eingabekraft an den DQM weitergereicht werden kann. Festgestellte Fehlermeldungen zu einer Visite sind daher zwingend zu korrigieren, damit die Daten in Studien eingehen können. Somit sind auch die Fehlermeldungen zu korrigieren, die nachträglich durch die neuen Validierungsregeln V2 gefunden wurden. Durch die Analyse der Validierungsergebnisse der Visiten des Datenbestandes D3 sollte nachvollzogen werden, inwieweit eine nachträgliche Bearbeitung einer Visite zur Verbesserung der Datenqualität beiträgt.

Für diesen Vergleich wurden drei Subgruppen des Datenbestandes identifiziert. Zunächst Visiten, die seit Einführung der neuen Validierungsregeln V2 nicht mehr bearbeitet worden waren, siehe Tabelle 65. Diese Visiten geben die Datenqualität der existierenden Altdaten wieder und bilden die Ausgangsbasis des Vergleichs.

Eine weitere Subgruppe bilden Visiten, die bereits vor Einführung der neuen Validierungsregeln Teil des Datenbestands waren und nach Einführung der neuen Validierungsregeln V2 bearbeitet worden waren, Tabelle 66.

Die beiden Subgruppen unterscheiden sich hinsichtlich des Anteils der Visiten mit Ereignissen. Während in den Altdaten in 80% der Visiten Ereignisse gefunden wurden, sind es bei den editierten Altdaten 53,3%. Dies entspricht einer Reduktion der Visiten mit Ereignissen um 33,2% im Zuge der Bearbeitung der Visiten. Der Anteil der Visiten mit Fehlermeldungen zeigte sich in der Subgruppe der bearbeiteten Visiten um 53,5% niedriger. Darüber hinaus sanken die durchschnittliche Anzahl der Fehlermeldungen pro Visite mit Fehlermeldungen und die Standardabweichung. Ursache hierfür kann die Abnahme des Maximums von 177 Fehlern auf 28 Fehlermeldungen sein. Der Anteil der Visiten mit Warnhinwei-

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	4592 (80,0%)	5,2 \pm 7,6	4,0 (1,0 - 179,0)	2,0	7,0
Fehlermeldungen	3903 (68,0%)	5,1 \pm 7,8	3,0 (1,0 - 177,0)	2,0	6,0
Warnhinweise	2281 (39,7%)	2,1 \pm 1,7	1,0 (1,0 - 31,0)	1,0	3,0

Tabelle 65: Validierungsergebnis Datenbestand D3 nur Visiten berücksichtigt mit letzter Bearbeitung vor dem 17.11.2016

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	2108 (53,3%)	3,2 \pm 3,2	2,0 (1,0 - 30,0)	1,0	4,0
Fehlermeldungen	1247 (31,7%)	3,1 \pm 3,0	2,0 (1,0 - 28,0)	1,0	4,0
Warnhinweise	1378 (35,0%)	2,2 \pm 1,8	2,0 (1,0 - 30,0)	1,0	3,0

Tabelle 66: Validierungsergebnis Datenbestand D3 nur Visiten mit letzter Bearbeitung nach 17.11.2016

sen war in der Gruppe der bearbeiteten Visiten um 12% geringer. Die Verteilungen der Warnhinweise der beiden Datenbestände waren weiterhin ähnlich.

Die nach der Aktualisierung des Registers im November 2016 erstellten Visiten unterlagen während ihrer gesamten Existenz den neuen Validierungsregeln V2 dieser Arbeit und somit strengeren Validierungsregeln als die bisherigen Visiten. Daher sollte, insbesondere im Hinblick auf die Warnhinweise der neuen Einheitenvalidierung, nachvollzogen werden, ob diese Visiten Unterschiede gegenüber dem Gesamtdatenbestand aufweisen. Das Ergebnis der Validierung dieser Subgruppe des Datenbestands D3 findet sich in Tabelle 67.

Die neu angelegten Visiten wiesen einen um 51,4% geringeren Anteil an Visiten mit Warnhinweisen auf gegenüber der Subgruppe der bearbeiteten Visiten. Es zeigte sich eine geringfügig geringere durchschnittliche Anzahl von Warnhinweisen pro Visite mit Warnhinweisen. Zunächst waren für diese Subgruppe keine Visiten mit Fehlermeldungen erwartet worden,

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	620 (23,3%)	2,1 \pm 1,9	1,0 (1,0 - 22,0)	1,0	2,0
Fehlermeldungen	214 (8,0%)	2,0 \pm 1,8	1,0 (1,0 - 11,0)	1,0	3,0
Warnhinweise	453 (17,0%)	1,8 \pm 1,7	1,0 (1,0 - 22,0)	1,0	2,0

Tabelle 67: Validierungsergebnis Datenbestand D3 nur Visiten erstellt in Registerversion R2 oder R3

da diese Visiten zu jedem Zeitpunkt den neuen Validierungsregeln V2 unterlagen. Die festgestellten Fehlermeldungen waren auf die im Juli 2017 durch den CERTAIN-Programmierer eingeführten *ConstraintValidator* zurückzuführen. Da die Visiten der Subgruppe jedoch bereits vor Juli akzeptiert sein konnten, mussten die im Juli neu hinzugefügten Validierungsregeln nicht erfüllt sein.

Die Verteilung wurde erneut berechnet und dabei nur Visiten einbezogen, die nach Etablierung der weiteren *ConstraintValidator* im Juli 2017 angelegt worden waren, siehe Tabelle 68.

	Visiten N	Validierungsergebnis (Anzahl pro auffälliger Visite)			
		Mittelwert \pm SD	Median (Min - Max)	25. Pz	75. Pz
Ereignisse	184 (27,4%)	1,8 \pm 1,6	1,0 (1,0 - 14,0)	1,0	2,0
Fehlermeldungen	3 (0,45%)	1,0 \pm 0,0	1,0 (1,0 - 1,0)	1,0	1,0
Warnhinweise	181 (27,0%)	1,9 \pm 1,6	1,0 (1,0 - 14,0)	1,0	2,0

Tabelle 68: Validierungsergebnis Datenbestand D3 nur Visiten erstellt nach 01.07.2017

Die verbleibenden drei festgestellten Fehlermeldungen konnten nur durch nachträgliche Änderungen der Visiten durch das Eingabepersonal erfolgt sein, nachdem die Visite bereits durch den DQM akzeptiert worden war. In dieser Subgruppe zeigte sich ein Anteil von Visiten mit Warnhinweisen von 27,0%. Die Anzahl der durchschnittlichen Warnhinweise pro Visite mit Warnhinweisen blieb ebenso nahezu identisch.

8.7 Inkompatible Einheiten

Zur Einschätzung der Effektivität der Warnhinweise bezüglich inkompatibler Einheiten der Einheitenvalidierung wurden die 9004 Visiten des Datenbestandes D2 mit den 12339 Visiten des Datenbestandes D3 verglichen. Die inkompatiblen Einheiten der überprüften Objekte wurden mittels der Funktion *StatusOfUnitsAndValues* erfasst. Um in die Überprüfung einzugehen mussten die Objekte entweder in den Geltungsbereich des *RangeValidator* oder des *EntryRangeValidator* fallen. Ebenso mussten in den einbezogenen Objekten sowohl ein Zahlenwert als auch eine Einheit vorhanden sein. Für den Vergleich wurde die Einheitenvalidierung der Validierungsregeln V3 genutzt. Hierdurch stieg die Anzahl der einbezogenen Objekte gegenüber Überprüfung der inkompatiblen Einheiten in Kapitel 8.2 an. Die Ergebnisse dieser Überprüfung finden sich für den Datenbestand D2 in Tabelle 69 und für den Datenbestand D3 in Tabelle 70.

	minimaler Datensatz	extendierter Datensatz	Studiendatensatz	Gesamt
Anzahl Objekte	198227	133797	18921	350945
Inkompatible Einheiten	1386	966	16	2368

Tabelle 69: Inkompatible Einheiten Datenbestand D2

	minimaler Datensatz	extendierter Datensatz	Studiendatensatz	Gesamt
Anzahl Objekte	264210	175944	21934	462088
Inkompatible Einheiten	1214	1025	19	2258

Tabelle 70: Inkompatible Einheiten Datenbestand D3

Während die Gesamtzahl der überprüften Objekte um 31,7% anstieg, war im Hinblick auf die absolute Anzahl der inkompatiblen Einheiten eine Verminderung um 4,6% festzustellen. Betrachtet man den Anteil der inkompatiblen Einheiten an der Gesamtzahl der überprüften Objekte sank dieser von 0,67% um 27,5% auf 0,49%.

8.8 Korrekturen von Ausreißern bei numerischen Datenelementen

Als besonders kritisch wurde das Vorgehen der Einrichtungen bei der Korrektur der Datenwerte erachtet, die durch die neuen Validierungsgrenzen in Labor und Medikation in Form von Fehlermeldungen oder Warnhinweisen bemängelt wurden. Der Datenbestand D2 wurde daher mittels der Funktion *InvestigateOutliers* auf Ausreißer hin untersucht und der Zustand sowie die etwaige Korrektur dieser Ausreißer im Datenbestand D3 nachvollzogen. Im Datenbestand D2 wurden zunächst bei 350945 überprüften Objekten 4987 Ausreißer festgestellt. Dies entspricht einem Anteil der Ausreißer von 1,42%. Wie bereits beschrieben wurde im Zuge der Überprüfung der Validierungsregeln V2 anhand des Datenbestandes D1 festgestellt, dass eine Einrichtung sehr viele falsche pH-Werte aufgrund falscher Bezeichnung eines Datenwertes in einer Importdatei in das Register importiert hatte. Von den 4987 Ausreißern waren 1635 auf dieses Vorgehen der Einrichtung zurückzuführen. Bei derselben Einrichtung konnten zudem 901 Ausreißer für das Datenelement „Hämoglobin“ festgestellt werden. 510 dieser Ausreißer hatten einen Datenwert von 0.0, der mit dem Ableben des Patienten hätte einhergehen müssen. Im Zuge der Rücksprache mit einem der DQM des CERTAIN-Registers äußerte dieser die Vermutung, dass es sich hierbei, wie beim Import der pH-Werte, um einen weiteren falsch beschrifteten Datenwert handeln könnte. Dies konnte jedoch nicht abschließend geklärt werden. Bei wenigen der restlichen 2451 Ausreißer konnten diese in der Kombination von Datenelement und Einrichtung systematischen Fehlern bei der Eingabe der Daten zugeschrieben werden. Insgesamt entfielen die verbliebenen Ausreißer auf diverse Einrichtungen und Datenelemente so dass hieraus keine weiteren Schlüsse gezogen werden konnten.

Der Zustand der 4987 Ausreißer des Datenbestandes D2 wurde anschließend im Datenbestand D3 nachvollzogen. Von 4987 Ausreißern blieben 3274 im Laufe des vergangenen Jahres zwischen den Datenbeständen unberührt. 876 wurden gänzlich gelöscht. 659 dieser gelöschten Datenwerte entfielen auf fehlerhafte Hämoglobin- und pH-Werte der obigen Einrichtung. 96 der als gelöscht markierten Datenwerte entfielen auf die Heidelberger Einrichtung. Diese Datenwerte waren einem falschen Medikament zugeordnet und daher in das falsche Datenelement eingetragen worden. Die eingetragenen Datenwerte verletzten daher die Validierungsgrenzen und wurden durch das medizinische Personal gelöscht und anschließend in das richtige Datenelement eingetragen. Für den entwickelten Algorithmus waren sie hierdurch nicht mehr auffindbar und wurden als gelöscht markiert. Die restlichen Löschungen entfielen auf unterschiedliche Einrichtungen und Datenelemente, so dass hieraus keine weiteren Schlüsse gezogen werden konnten.

Von den 4987 Ausreißern wurden 836 in ihrem Datenwert ihrer Einheit oder ihrem Parameter angepasst. Bei vier Datenwerten war die neue Einheit nicht mehr zur Validierungseinheit kompatibel, wodurch die Fehlermeldung zu einem Warnhinweis hinsichtlich einer

inkompatiblen Einheit wurde. Ein Datenwert verblieb auch nach einer erfolgten Änderung der Einheit weiterhin außerhalb der Validierungsgrenzen. 54 Datenwerte wurden soweit herabgesenkt, dass sie nur noch einen Warnhinweis auslösten. Von diesen 54 Datenwerten waren keine Datenwerte soweit herabgesenkt worden, dass sie innerhalb von 10% der jeweiligen harten Validierungsgrenze lagen. Zwei der korrigierten Ausreißer wurden soweit abgesenkt, dass sie innerhalb von 10% der weichen Validierungsgrenzen lagen. Bei 85 der Ausreißer wurde der Parameter des Eintrages der fortlaufenden Dokumentation angepasst, so dass dieser folglich anderen Validierungsgrenzen unterlag und aufgrund dessen keinen Ausreißer mehr darstellte. Von 836 angepassten Datenwerten erzeugten 778 danach keine Fehlermeldung oder keinen Warnhinweis mehr.

Im Datenbestand D3 existierten 3279 Ausreißer in 462088 Objekten, dies entspricht einem Anteil von 0,71%. Im zwischen den Datenbeständen vergangenen Jahr kamen vier Ausreißer hinzu. Diese Ausreißer wurden nachträglich in die Datenbestände eingetragen und konnten in drei Fällen auf Änderungen der jeweiligen Einheit zurückgeführt werden. Im letzten Fall wurde im Laufe des Jahres ein fehlerhafter Datenwert in ein Datenelement eingetragen, das vorher keinen Datenwert beinhaltete. Der Anteil der Ausreißer sank insgesamt innerhalb eines Jahres um 49,9%.

8.9 Clustering unterhalb der Validierungsgrenzen von numerischen Datenelementen

Da die harten Validierungsgrenzen der Validierung von numerischen Datenelementen zwingend erfüllt werden müssen, könnten die Eingabekräfte dazu verleitet werden, Datenwerte soweit abzusenken, bis diese die harten Validierungsgrenzen erfüllen, um die Visite anschließend in den Zustand „Freigegeben“ versetzen zu können. Ebenso kann bei den Eingabekräften hinsichtlich der weichen Validierungsgrenzen die Vorstellung entstehen, dass diese Datenwerte ebenfalls korrigiert werden müssten und somit der beschriebene Lösungsansatz durch das Eingabepersonal gewählt wird, bis der Warnhinweis nicht mehr angezeigt wird. Um ausschließen zu können, dass Datenwerte so weit abgesenkt wurden, bis diese keine Ereignisse mehr darstellen, wurde auf Basis der Methoden zur Aggregation der validierten numerischen Datenwerte (Funktion *AggregateValidatedValues*) ein R-Skript zur Prüfung der Werteverteilungen auf Clustering entwickelt. Auf der Basis der Datenwerte jedes Datenelementes wurde zunächst die Dichtefunktion der Verteilung anhand des Histogramms bestimmt. Anschließend wurden die lokale Maxima der Dichtefunktion berechnet, beispielhaft illustriert für das Datenelement „Triglycerides“ durch blaue Linien in Abbildung 26. Im Hinblick auf die festgestellte Anzahl der lokalen Maxima wurden zur Reduktion der Sensitivität nur Maxima akzeptiert, auf die mehr als 0,5% des Maximalwertes der Dichtefunktion entfielen.

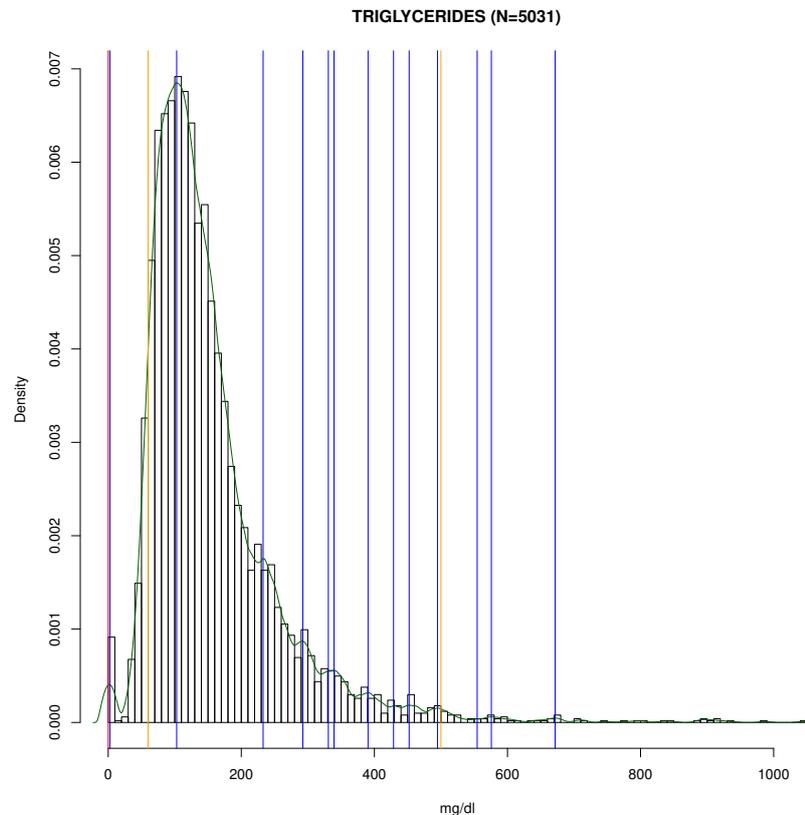


Abbildung 26: Lokale Maxima der Dichtefunktion

Anschließend wurde für die berechneten Maxima überprüft, ob diese in der Nähe der jeweiligen harten oder weichen Validierungsgrenzen liegen. Für Obergrenzen wurde als Einschlussbereich $Grenze * 0,95$ gewählt. Für Untergrenzen wurde der Einschlussbereich als $Untergrenze + Obergrenze - Obergrenze * 0,95$ definiert, um Untergrenzen von 0 zu berücksichtigen und für beide Formen von Grenzen Korridore derselben Größe zu überprüfen. Da die Dichtefunktion das Histogramm nicht exakt abbildet, sondern leicht versetzt sein kann, wurde zur besseren Erkennung der relevanten, lokalen Maxima der Einflussbereich um $Obergrenze * 1,005$ bzw. $Untergrenze * 0,995$ erweitert. Die in dieser Form bereinigten Maxima für das Beispiel „Triglycerides“ finden sich in Abbildung 27.

Die entwickelte Berechnung zur Erkennung der relevanten, lokalen Maxima zeigte sich anfällig für die Größe des Betrags zwischen den Validierungsgrenzen sowie für die Lage des Medians der Verteilung innerhalb dieses Bereichs. Durch große Unterschiede in der Größenordnung der Ober- und Untergrenze wurden umso größere Teile der Verteilung in den jeweiligen Einflussbereich einbezogen und somit auch mehr lokale Maxima. Durch die Nähe des Median zu den Validierungsgrenzen fallen mehr Datenwerte in den Einschluss-

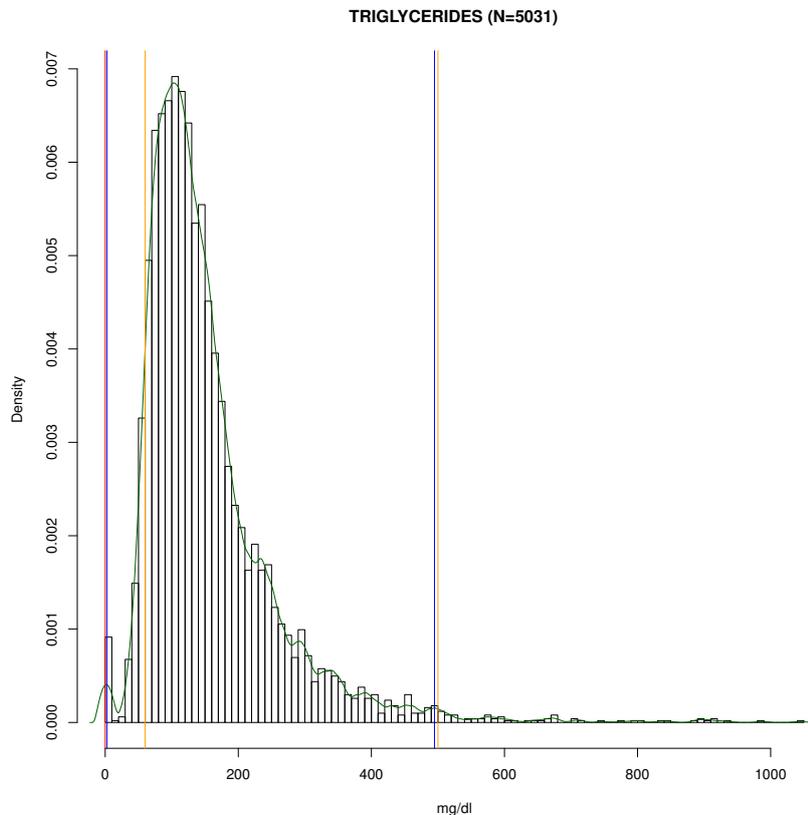


Abbildung 27: Lokale Maxima der Dichtefunktion in Nähe der Validierungsgrenzen

bereich der Berechnung der relevanten Maxima. Somit steigt die Wahrscheinlichkeit nicht relevante Maxima im Rahmen der Berechnung zu erfassen. Dies wurde weiter begünstigt durch uneinheitliche Dateneingabe der Zentren, beispielsweise indem Datenwerte in manchen Zentren nur ganzzahlig in anderen jedoch mit Nachkommastellen eingegeben wurden und somit zu einem lokalen Maximum führten. Eine Rundung aller verwendeten Datenwerte auf zwei Nachkommastellen zeigte keine Verbesserung hinsichtlich der Anzahl der lokalen Maxima. Eine Rundung auf weniger als zwei Nachkommastellen wurde verworfen da hierdurch eine zu starke Nivellierung der Datenwerte befürchtet wurde. Da sich die Größenordnungen der verschiedenen Datenelemente stark unterscheiden, konnte keine einheitliche Lösung gefunden werden. Die Berechnung zeigte sich zudem abhängig von der Klassenanzahl des Histogramms und der damit verbundenen Genauigkeit der Dichtefunktion. Insbesondere bei den Medikationsangaben zeigten sich multimodale Verteilungen die zu vielen lokalen Maxima führten.

Die Darstellungen der Verteilungen wurden für alle Datenelemente mit Darstellung der Validierungsgrenzen berechnet. Auf Basis des Algorithmus erkannte Auffälligkeiten wurden

dem medizinischen Registerpersonal des CERTAIN-Registers vorgelegt. Von 93 Verteilungen mit durch den Algorithmus identifizierten lokalen Maxima in Nähe der Validierungsgrenzen wurden sechs auf Basis der identifizierten lokalen Maxima durch das medizinische Personal als auffällig eingeschätzt. Das medizinische Personal schätzte die generierten Grafiken nach Durchsicht aller erzeugten Grafiken als guten Ansatzpunkt für weitere Analysen bei auffälligen lokalen Maxima ein. Die konkrete Untersuchung der Ursachen und Maßnahmen zur Adressierung dieser Ursachen können in weiteren Schritten erfolgen.

9 Diskussion und Ausblick

Im Rahmen dieser Arbeit wurden automatisierte Mechanismen zur Sicherung der Datenqualität von medizinischen Registern entwickelt. Im Laufe der Arbeit konnten mehrere Ergebnisse festgehalten werden:

- Analyse des Potentials der automatischen Validierung für das KRhOKo-Register
- Allgemeines Vorgehen zur Verbesserung der Datenqualität in Registern auf Basis der TMF-Leitlinie
- Nutzung des entwickelten Vorgehens für die drei CERTAIN-Register inklusive systematischer Analyse der QI hinsichtlich Anwendbarkeit, Implementierbarkeit sowie Aussagekraft und Festlegung der QI-Prioritäten und -Gewichte
- QI-Empfehlung und -Vorlage zur vereinfachten Anwendung des entwickelten, allgemeinen Vorgehens
- Konzept und Implementierung von Validierungsregeln zur automatisierten Qualitätssicherung
- Analyse des Einflusses der entwickelten Validierungsregeln auf die Datenqualität des CERTAIN-Registers

Von diesen Ergebnissen können insbesondere das allgemeine Vorgehen, die QI-Analyse, die QI-Empfehlung sowie das Konzept der Implementierung der Validierungsregeln von anderen Registern als Basis zur Verbesserung ihrer Datenqualität genutzt werden.

Im Folgenden werden zunächst die einzelnen Ergebnisse und anschließend die verwendeten Methoden im Detail diskutiert.

9.1 Diskussion der Ergebnisse

Analyse des Potentials automatischer Validierung für das KRhOKo-Register

Zu Beginn der Arbeit sollte das Potential der automatischen Validierung anhand der Validierungsregeln des neuen KRhOKo-Registers überprüft werden. Das Potential der Nutzung von automatischen Validierungsmechanismen als valides Werkzeug zur Sicherung der Datenqualität konnte durch die Validierung von 241 digitalisierten Fragebogen gegen die Validierungsregeln des KRhOKo-Registers bestätigt werden. Bei etwa 14% der digitalisierten Fragebogen konnten Verstöße gegen die Validierungsregeln festgestellt werden, die durch die Nutzung von Validierungsregeln zur Eingabezeit hätten verhindert werden können. Durch die angestrebte Niederschwelligkeit der Datenerhebung werden nur die notwendigen Sachverhalte validiert, um den Eingabekräften die Dateneingabe nicht unnötig

zu erschweren. Eine umfassendere Validierung hätte möglicherweise noch mehr Verstöße gegen die Validierungsregeln festgestellt, die bei einer elektronischen Erhebung der Daten hätten verhindert werden können.

Im Hinblick auf den verwendeten Digitalisierungsprozess ist festzuhalten, dass durch diesen eine Reihe von Fehlern in den digitalisierten Daten entstanden sind, die wiederum auch zu Verstößen gegen die Validierungsregeln führten. Der Import dieser digitalisierten Daten in das Register ist daher problematisch, besonders wenn die importierten Daten als Ausgangsbasis für spätere Follow-Ups durch Übernahme von Daten aus den vorherigen Follow-Ups genutzt werden sollen. Bei Fehlern im Digitalisierungsprozess, die zur Löschung oder Erzeugung von Datenwerten führen, die aber weiterhin eine valide Merkmalsausprägung annehmen, können keine Maßnahmen ergriffen werden, da die Fehler nicht bemerkt werden können. Die Ausnahme bilden hier Datenwerte, deren Bezug zu anderen Datenwerten durch Plausibilitätsprüfungen überprüft wird. Importierte Daten sollten daher besonders für die Eingabekräfte zur Überprüfung markiert werden und den gesamten Datenqualitätsmanagementprozess durchlaufen. Nicht erklärt werden konnte der Unterschied in der Anzahl der Digitalisierungsfehler zwischen Arzt- und Eltern- sowie Jugendlichenbogen. Da alle drei Fragebogentypen denselben Digitalisierungsprozess durchliefen, wäre eine ähnliche Fehlerhäufigkeit zu erwarten gewesen. Die Arztbogen wiesen jedoch eine deutlich geringe Anzahl an Fragebogen mit Digitalisierungsfehlern auf.

Die festgestellten Unterschiede in der Vollständigkeit zwischen Eltern- und Jugendlichenbogen wurden nur mit wenigen Maßnahmen zur Verbesserung der Vollständigkeit berücksichtigt, da nur eine geringe Anzahl der am Register teilnehmenden Einrichtungen sich vorab zur Nutzung der Patientenbeteiligung bereit erklärte. Die Effektivität der neuen Maßnahmen konnte aufgrund mangelnder Eingaben durch die Patienten nicht überprüft werden. Weitere Maßnahmen zur Verbesserung der Vollständigkeit wurden daher vor der anderweitigen Weiterentwicklung des KRhOKo-Registers zurückgestellt. Es wurde nicht weiter analysiert, worauf die Unterschiede zwischen den Benutzergruppen zurückzuführen sind und wie die Ursachen der Unterschiede adäquat für jede Benutzergruppe einzeln zu adressieren wären. Diese Punkte wären bei einem Anstieg der Nutzung der Patientenbeteiligung erneut zu betrachten.

Allgemeines Vorgehen zur Verbesserung der Datenqualität in Registern auf Basis der TMF-Leitlinie

Das entwickelte allgemeine Vorgehen stellt eine Erweiterung der TMF-Leitlinie und ein formales Vorgehen zur vereinfachten Anwendung der TMF-Leitlinie dar. Die identifizierten Kategorien von QI des Vorgehens gehen auf von Arts et al. (Arts et al. 2002) beschriebene Prozeduren zur Sicherung der Datenqualität zurück und stehen somit auf einem erprobten Fundament. Durch die QI-Kategorien „Registerplanung“ und „automatische Validierung“ wird eine Reduzierung des Aufwandes der Anwendung der TMF-Leitlinie erreicht, da die

QI dieser QI-Kategorien durch konkrete Maßnahmen zur Sicherung der Datenqualität abgedeckt werden sollen. Somit sinkt die Zahl der zu berechnenden QI und gleichermaßen der Aufwand der Berechnung des Gesamtscores. Die Kategorien „Registerplanung“ und „automatische Validierung“ zielen auf die Verhinderung des Eingangs fehlerhafter Datenwerte ab, die eine aufwändige Korrektur dieser Daten in späteren Phasen des Forschungsvorhabens erfordern würden. Auf der Basis der identifizierten QI-Kategorien ist zudem eine einfachere Auswahl der relevanten QI je nach bevorzugtem Ansatz (automatische Validierung, Reporting oder Datenverifizierung) möglich.

Die Auswahl der für ein Register relevanten QI auf Basis der Kriterien Anwendbarkeit, Implementierbarkeit und Aussagekraft berücksichtigt grundlegende Charakteristiken der QI. Bei anderen Forschungsvorhaben mögen diese Kriterien nicht alle Betrachtungspunkte erfassen. In diesen Fällen sollten die Kriterien erweitert werden, um die ganzheitliche Analyse der QI zu ermöglichen.

Die Nutzung der registerspezifischen QI-Analyse zur Priorisierung der QI ermöglicht Anwendern eine Transformation ihrer bisherigen Arbeit in Kenngrößen in Form der vergebenen Prioritäten, auf deren Basis sie die Verbesserung der Datenqualität gezielt vorantreiben können. Das Vorgehen wurde bewusst flexibel gehalten, um Eigenheiten des jeweiligen Registers berücksichtigen zu können und Anwendern die notwendige Freiheit bei der Festsetzung ihrer Prioritäten zur Umsetzung der QI zu geben.

Die Vergabe der Gewichte an die Kategorien zur Gewichtung der einzelnen QI baut ebenso auf den Ergebnissen der Anwender auf und gibt diesen somit die Möglichkeit ihre eigenen Anforderungen und Einschätzungen genau abzubilden. Eine Gewichtung der QI der Kategorien „Registerplanung“ und „automatische Validierung“ ist nicht erforderlich, da diese QI aufgrund konkreter Maßnahmen zur Sicherung der Datenqualität immer als „unauffällig“ angenommen werden sollten und den Gesamtscore bei Berechnung somit positiv beeinflussen. Zur Unterstützung der Vergleichbarkeit mit den Gesamtscores anderer Register können diese dennoch gewichtet werden, sollten dann jedoch stets als „unauffällig“ in die Berechnung eingehen, da auffällige Werte dieser QI durch die ergriffenen Maßnahmen hinreichend plausibel ausgeschlossen werden können.

Nutzung des entwickelten Vorgehens für die drei CERTAIN-Register

Das entwickelte Vorgehen zeigte sich bei Anwendung für die drei CERTAIN-Register als gut geeignet um die relevanten QI zu identifizieren. Die Kriterien Anwendbarkeit, Implementierbarkeit und Aussagekraft ermöglichten es, ein differenziertes Bild jedes QI im Hinblick auf seine Anwendung in den drei Registern zu zeichnen.

Die Ergebnisse der Analyse der QI der TMF-Leitlinie hinsichtlich Anwendbarkeit, Implementierbarkeit und Aussagekraft sind durch die Erfahrungen mit der CERTAIN-Registerplattform und den Datensätzen der drei Register geprägt. Die Analyse der QI im Hinblick auf Gesichtspunkte, wie beispielsweise eine stringente Definition des Datensatzes und der

Eingabemasken sowie auf die automatische Validierung, betont Aspekte, die in der TMF-Leitlinie zumeist nur beiläufig als Einflussfaktoren der QI aufgegriffen wurden, und ergänzt hierdurch die Ausführungen der TMF-Leitlinie zu den einzelnen QI. Durch die unterschiedlichen Anforderungen der drei Register konnten die QI für die Analyse aus unterschiedlichen Blickwinkeln betrachtet werden. Daher sollten die Ergebnisse der QI-Analyse auch für andere Register anwendbar sein, da grundlegende Gemeinsamkeiten in den verwendeten Datentypen und -sätzen angenommen werden können. Limitierung der Übertragbarkeit der QI-Analyse stellt hierbei die grundlegende Analyse der QI für ein webbasiertes, mindestens jedoch ein elektronisches Register dar. Für ein papierbasiertes Register werden die hier beschriebenen Ausführungen nur eingeschränkte Relevanz besitzen. Dies wird insbesondere dadurch deutlich, dass eine direkte Rückmeldung an die Eingabekräfte zur Eingabezeit in der Analyse vieler QI aufgegriffen wird. Im Zuge der fortschreitenden Digitalisierung in der Medizin ist indessen mit einer Zunahme von rein elektronischen Registern zu rechnen. Die Ergebnisse der QI-Analyse bleiben daher auch für zukünftige Register relevant.

Die Anwendbarkeit der QI für die drei CERTAIN-Register musste als heterogen beurteilt werden, da sich die Geltungsbereiche der QI sehr unterscheiden. Dieses Bild wiederholte sich bei der Beurteilung der Implementierbarkeit, da viele unterschiedliche Daten in die Berechnungen einfließen können, die zuweilen schwer für die Berechnung zusammenzustellen sind. Die Aussagekraft wurde mit zwei Ausnahmen stets als mindestens neutral und überwiegend als gut beurteilt. Durch die Bewertung der Anwendbarkeit und Implementierbarkeit wurde zudem deutlich, dass eine Umsetzung einiger QI durch automatische Validierungsregeln weniger aufwändig sein kann, als die Implementierung der Berechnung der QI selbst. Dies ist in den drei Registern besonders bei den QI ersichtlich, welche das Fehlen von Datenwerten erfassen sollen. Durch die Schachtelung von Folgefragen muss in diesen Fällen für die Berechnung genau betrachtet werden, ob das jeweilige Datenelement überhaupt hätte ausgefüllt werden können. Durch die automatische Validierung kann das Vorhandensein der Datenwerte jedoch sichergestellt werden. Bei vielen QI musste festgestellt werden, dass manuelle Vorgänge zur Berechnung der QI notwendig gewesen wären. Beispielsweise wäre für die Berechnung der Rekrutierungsrate für CERTAIN-LI und KRhOKo eine manuelle Recherche hinsichtlich der Transplantations- und Erkrankungszahlen notwendig. Dies widersprach dem Ziel der Automatisierung der Sicherung der Datenqualität, so dass diese QI auch bei Anwendbarkeit auf die drei CERTAIN-Register wohl letztlich nicht umgesetzt werden.

Im Hinblick auf die Registerempfehlung der TMF-Leitlinie wurden die QI in dieser Analyse umfassender und eingehender analysiert. Statt aus dem Blickwinkel eines Registers konnten Einschätzungen zu den QI aus der Sicht von drei Registern dargelegt werden. Nicht alle QI konnten detailliert analysiert werden, da eine Anwendbarkeit für einige der QI in den drei Registern nicht gegeben war. Insgesamt wurden jedoch mehr QI differenziert analysiert als in der Registerempfehlung. Hierdurch kann die QI-Analyse als zusätzliches Hilfsmittel bei

der Anwendung der TMF-Leitlinie und des Ansatzes dieser Dissertation dienlich sein und bei der Auswahl von relevanten QI für andere Register helfen.

QI-Empfehlung und -Vorlage

Wie schon im Rahmen der Analyse der QI für die drei CERTAIN-Register festgestellt, haben die jeweiligen Datensätze und Rahmenbedingungen erheblichen Einfluss auf die Umsetzbarkeit von Maßnahmen zur Sicherung der Datenqualität. Grundsätzlich besitzen Register jedoch dasselbe Ziel der Datensammlung. Daraus folgen maßgeblich gleiche Anforderungen an hohe Datenqualität und Maßnahmen, um diese sicherzustellen. Die entwickelte QI-Empfehlung ermöglicht die Konzentration auf registerspezifische Eigenheiten, da die grundlegende Einordnung der einzelnen QI hinsichtlich ihres Ansatzpunktes und der Dringlichkeit ihrer Umsetzung bereits vorgenommen wurde. Die Basis der QI-Empfehlung stellt die Analyse der QI für die drei CERTAIN-Register dar. Andere Register können sich von den Rahmenbedingungen dieser Register unterscheiden.

Die Übertragbarkeit der QI-Empfehlung unterliegt daher verständlicherweise Grenzen, die durch eine registerspezifische Analyse der QI und anschließender Priorisierung auf Basis der QI-Empfehlung überwunden werden können. Die Nutzung der QI-Empfehlung erfordert daher eine kritische Hinterfragung der QI-Empfehlung im Hinblick auf die Eigenheiten des jeweiligen Registers. Durch die bereitgestellte Vorlage ist es Anwendern möglich, die Ergebnisse einer vorher durchgeführten registerspezifischen QI-Analyse mit der QI-Empfehlung zu vergleichen und gegebenenfalls Anpassungen an den Prioritäten vorzunehmen. Die Ergebnisse der durchgeführten QI-Analyse für die drei CERTAIN-Register und der QI-Empfehlung stehen den Anwendern dabei stets als Referenz zur Verfügung, um die Arbeiten zur Umsetzung des Ansatzes zu vereinfachen.

Durch die Nutzung der QI-Empfehlung ist es zu erwarten, dass Anwender bei reduziertem Aufwand zu guten Ergebnissen im Hinblick auf die Auswahl der für ihr Register relevanten QI kommen können.

Durch die QI-Empfehlung wird versucht, bereits frühestmöglich mit der Sicherung guter Datenqualität zu beginnen und hierdurch den Aufwand für das adaptive Feedback zu senken.

Konzept und Implementierung von Validierungsregeln zur automatisierten Qualitätssicherung

Die zur Kategorie „Automatische Validierung“ des entwickelten Vorgehens zugehörigen QI wurden ganzheitlich oder in Teilen in Software-Komponenten zur Sicherung der Datenqualität umgesetzt. Die Anforderung der generischen Umsetzung wurde ganzheitlich erfüllt. Alle entwickelten Software-Komponenten bauen entweder auf Annotationen oder Interfaces auf, die einfach in weitere Register der CERTAIN-Registerplattform einzubringen sind.

Anwender können die entwickelten und dargelegten Ansätze ebenso in ihre Projekte überführen, da die Ansätze und Software-Komponenten agnostisch gegenüber Klassenstrukturen und Datensätzen sind. Bei einer direkten Übernahme der Software-Komponenten aus der CERTAIN-Registerplattform in andere Register ist einzig eine Anpassung des Projektpackages in den Reflection-Methoden für das jeweilige neue Projekt erforderlich. Dies bedingt die Anpassung eines Strings, der den Namen des Projekt Packages enthält.

Strings wurden auch genutzt um Warnhinweise neben Fehlermeldungen als zweite Klasse von Verstößen gegen die Validierungsregeln in der CERTAIN-Registerplattform zu etablieren. Die Nutzung einer String-Konstante zur Identifikation von Warnhinweisen auf Basis der Validierungsnachricht ist als fehleranfällig zu erachten. Eine Nutzung der statischen *payload* der *Constraint Annotation* hätte das Fehlerpotential verringert, gleichzeitig aber die Trennung von harten und weichen Validierungsgrenzen in den beiden *ConstraintValidator* zur Validierung der Wertebereiche bedingt und somit insgesamt vier *ConstraintValidator* zur Validierung der Wertebereiche erfordert. Zudem hätte dieser Ansatz die aufwändige Filterung der Validierungsnachrichten erfordert, um bei einem Verstoß gegen die harten Validierungsgrenzen den zeitgleich festgestellten Verstoß gegen die weichen Validierungsgrenzen zu unterdrücken, so dass nur eine Fehlermeldung angezeigt werden würde. Diese Anpassungen wogen schwerer gegenüber der Anfälligkeit eines Strings zur Kodierung der Warnhinweisen, daher wurde der Ansatz beibehalten.

Analyse der Datenqualität des CERTAIN-Registers

Durch die Einführung der neuen Validierungsregeln wurden eine Vielzahl von Verstößen gegen die Validierungsregeln festgestellt, die vorher unbemerkt geblieben waren. Die Analysen des Datenbestandes D2 zeigten, dass selbst nach eingehender Kontrolle der neuen Validierungsregeln V2 und der durchgeführten Korrekturen am Datenbestand ein Großteil der Visiten mit Fehlermeldungen behaftet war. Die anhand des Datenbestandes D3 festgestellte Datenqualität zeigte ein deutlich verbessertes Bild der Datenqualität. Der Anteil der Visiten mit Verstößen gegen die Validierungsregeln war um 33,8% gesunken, der Anteil der Visiten mit Fehlermeldungen um 47,4%. Aus der Betrachtung des Validierungsergebnis, rein bezogen auf die neuen Validierungsregeln, konnte ein deutlicher Rückgang von Fehlermeldungen festgestellt werden, die durch die neuen Validierungsregeln ausgegeben worden waren. Der Anteil von Visiten mit Fehlermeldungen, die durch die neuen Validierungsregeln festgestellt wurden, fiel innerhalb eines Jahres um 56,7%. Die Datenqualität konnte somit durch die neuen Validierungsregeln innerhalb eines Jahres deutlich verbessert werden.

Die Bearbeitung von Visiten des Datenbestandes D2 innerhalb des vergangenen Jahres ging mit einer Verbesserung der Datenqualität der jeweiligen bearbeiteten Visite einher. Nach einer Bearbeitung wiesen Visiten mit Fehlermeldungen eine geringere durchschnittliche Anzahl von Fehlermeldungen auf. Im Hinblick auf den Anteil der Visiten mit Fehlermel-

dungen am Datenbestand konnte dieser im Vergleich zu den unbearbeiteten Altdaten von 68% um 53,5% auf 31,7% reduziert werden. Die Anzeige von Fehlermeldungen in bestehenden Visiten wurde vermutlich von den Eingabekräften als Anlass zur Korrektur dieser Daten wahrgenommen.

Das Validierungsergebnis der Subgruppe der Visiten, die nach Einführung der neuen Validierungsregeln V2 angelegt worden waren, war wider Erwarten nicht frei von Ereignissen. Dies konnte durch neue *ConstraintValidator*, die außerhalb dieser Arbeit entwickelt wurden, erklärt werden. Jedoch waren auch nach Ausschluss dieser *ConstraintValidator* durch Änderung des zeitlichen Rahmens der Betrachtung Visiten mit Fehlermeldungen Teil der Subgruppe. Diese Fehlermeldungen mussten daher durch Änderungen ausgelöst worden sein, die vorgenommen wurden, nachdem die Visiten im Zuge des Datenqualitätsmanagementprozesses geprüft worden waren. Bei der Änderung von Daten und dem Speichern der Daten werden diese erneut gegen die Validierungsregeln geprüft und das Ergebnis der Überprüfung dem Benutzer angezeigt. Visiten können durch die Eingabekräfte grundsätzlich immer bearbeitet werden, beispielsweise mit dem Ziel Daten zu einzelnen Studien zur Visite hinzuzufügen. Da die CERTAIN-Registerplattform die Änderungen an einzelnen Datenelementen nicht detailliert speichert, konnten die Verursacher nicht ausgemacht werden. Grundsätzlich ist es bedenklich, dass Benutzer entgegen der Fehlermeldungen des Registers nachträglich falsche Datenwerte in den Datenbestand einbringen. Diesbezüglich sollten weitere Beobachtungen erfolgen, und bei vermehrtem Auftreten falscher Datenwerte sollten Maßnahmen zur Verhinderung dieser Vorgänge ergriffen werden. Beispielsweise durch erneutes Prüfen der Visite durch den DQM sollten Daten durch die Einrichtung geändert worden sein.

Ein uneinheitliches Bild zeigte sich bei der Überprüfung der Änderungen an Ausreißern. Von 4987 Ausreißern wurden 876 gelöscht und von 836 Datenwerten, die Änderungen erfahren hatten, waren 778 danach nicht mehr zu beanstanden. Der Großteil der gelöschten Datenwerte entfiel auf die Einrichtung, bei der falsche pH- und Hämoglobinwerte importiert worden waren. Daher wurden hier korrekterweise offensichtlich falsche Daten gelöscht. Fraglich bleibt, welche Daten an dieser Stelle hätten importiert werden sollen und ob diese tatsächlich durch die Eingabekräfte der Einrichtung nachgetragen wurden. Dies konnte nicht überprüft werden. Im Hinblick auf das befürchtete Absenken von Datenwerten, bis diese entweder die harte oder weiche Validierungsgrenze verletzten, konnte kein diesbezügliches Fehlverhalten festgestellt werden. Lediglich zwei von 836 bearbeiteten Ausreißern wurden in den als kritisch erachteten Bereich von 10% innerhalb der jeweiligen Validierungsgrenzen abgesenkt. Da der Großteil der gelöschten Ausreißer auf eine Einrichtung entfällt, verzerrt dies das Bild der Korrektur der Ausreißer.

Im Vergleich der Datenbestände D2 und D3 hinsichtlich inkompatibler Einheiten konnte ein Absinken des Anteils von inkompatiblen Einheiten um 27,5% beobachtet werden. Da die durchschnittliche Anzahl von Warnhinweisen pro Visite zwischen den Datenbeständen

unverändert blieb, kann dieser Rückgang auf die Abnahme des Anteils von Visiten mit Warnhinweisen an der Gesamtzahl der Visiten zurückgeführt werden. Eine mögliche Ursache hierfür könnte sein, dass manche Einrichtungen die Warnhinweise ernster nehmen und die Ursachen dann korrigieren. Der Rückgang an inkompatiblen Einheiten ist als positiv festzuhalten.

Insgesamt zeigte sich im Zuge der Analysen eine deutlich verbesserte Datenqualität. Besonders hervorzuheben ist hierbei die Verbesserung der Datenqualität im Zuge der Bearbeitung einer Visite. Da im CERTAIN-Register kontinuierlich Studien durchgeführt und somit auch Visiten immer wieder editiert werden, scheint die Etablierung von weiteren Validierungsregeln auch weiterhin sinnvoll. Dies liegt darin begründet, dass diese Validierungsregeln beim Speichern einer bearbeiteten Visite geprüft werden und das Ergebnis dieser Überprüfung dem Benutzer mitgeteilt wird, der die Verstöße gegen die Validierungsregeln dann direkt korrigieren kann. Hinsichtlich der mit Ereignissen belasteten Altdaten scheint somit eine stetige Verbesserung der Datenqualität auf Basis weiterer Validierungsregeln plausibel, da die Eingabekräfte die Ursachen dieser Ereignisse im Zuge einer Bearbeitung korrigieren.

Im Hinblick auf die durchgeführte Prüfung zum Auftreten von auffälligen Werte-Clustern in Nähe der Validierungsgrenzen musste festgestellt werden, dass sich das Problem der Erkennung der Cluster aufgrund der Heterogenität der Verteilungen der Datenelemente als äußerst schwierig erwies. Ein präzises, allgemeingültiges Verfahren zur Erkennung der Cluster konnte nicht entwickelt werden. Der entwickelte Algorithmus wurde zwar von Seiten des medizinischen Personals des CERTAIN-Registers als grundsätzlich positiv hinsichtlich der Markierung möglicher Auffälligkeiten bewertet. Die hohe Anzahl an falsch positiven Auffälligkeiten macht jedoch unbedingt eine manuelle Prüfung des Ergebnisses des Algorithmus durch das medizinische Personal erforderlich.

Im Hinblick auf die Vielzahl der Datenelemente im Datensatz des CERTAIN-Registers muss festgehalten werden, dass die detaillierte Durchsicht aller Datenwerte für den DQM mit erheblichem Aufwand verbunden ist.

9.2 Diskussion der Methoden

Analyse des Potentials automatischer Validierung für das KRhOKo-Register

Für die Berechnung des erforderlichen Umfangs der Analyse der papierbasierten Fragebogen der Kerndokumentation wurde die in der TMF-Leitlinie beschriebene Fallzahlplanung genutzt, die auf der Basis von Literaturquellen definiert wurde. Der erfolgte Abgleich der Papierbogen mit den digitalisierten Bogen wurde nur durch eine Person vorgenommen und nicht durch Dritte verifiziert. Neben der Feststellung eines Unterschieds in der Vollständigkeit zwischen den Eltern- und Jugendlichenbogen wäre eine weitergehende Analyse zur Untersuchung der Ursachen für diesen Unterschied durchzuführen gewesen, um die Unterschiede gezielter als mit einer allgemeinen Vollständigkeitsanzeige adressieren zu können.

Da sich bereits zum Zeitpunkt dieser Analyse nur eine sehr geringe künftige Nutzung der Patientenbeteiligung abschätzen ließ, wurden keine Arbeiten unternommen um diese Punkte zu adressieren.

Allgemeines Vorgehen zur Verbesserung der Datenqualität in Registern auf Basis der TMF-Leitlinie

Die Identifikation der Kategorien des Vorgehens beruht nicht auf einem systematischen Literaturreview, welches vielleicht zur Favorisierung eines anderen Vorgehens geführt hätte. Die Arbeit von Arts et al. stellt jedoch immer noch einen relevanten Meilenstein dar, auf dem auch die TMF-Leitlinie aufbaut. Die Auswahl der Kriterien Anwendbarkeit, Implementierbarkeit und Aussagekraft erfolgte auf Basis der Ausrichtung dieser Dissertation hinsichtlich der Verbesserung der Datenqualität durch technische Maßnahmen. Andere Entscheidungsträger eines Registers mögen aufgrund ihrer Profession andere Kriterien wählen.

Nutzung des entwickelten Vorgehens für die drei CERTAIN-Register

Die Analyse der QI wurde zunächst für jeden QI einzeln auf Basis der während der Entwicklung der drei Register gemachten Erfahrungen erarbeitet. Diese Analysen der QI wurden dann mit den Registerzentralen diskutiert und bestätigt. Hierdurch orientieren sich die Ergebnisse der Analyse der QI an der Registerwirklichkeit der drei CERTAIN-Register. Eine allgemeingültigere Analyse hätte beispielsweise durch eine Delphi-Umfrage unter der Einbeziehung verschiedener Registerzentralen erreicht werden können. Die Analyse der QI gibt dennoch die Blickwinkel dreier Register auf die QI wieder und bietet ein gesichertes Mindestmaß an Heterogenität. Die Analyse der nicht auf die CERTAIN-Registerplattform anwendbaren QI musste dabei aufgrund mangelnder Einsicht in die Bedeutung dieser QI allgemein gehalten werden. Diese QI mögen in anderen Registern einen höheren Stellenwert einnehmen, der jedoch nicht auf der Basis der während der Entwicklung der CERTAIN-Registerplattform gemachten Erfahrungen wiedergegeben werden konnte.

Die Zusammenfassung der Blickwinkel dreier Register auf einen QI innerhalb einer Tabelle gestaltete sich für einzelne QI schwierig, bei denen größere Unterschiede zwischen den Registern festgestellt wurden. Dies führte stellenweise zu Kompromissen in der Bewertung der QI, die bei einer getrennten Bewertung der QI für jedes einzelne Register nicht hätten eingegangen werden müssen. Die identifizierten Kriterien Anwendbarkeit, Implementierbarkeit und Aussagekraft zeigten sich als geeignet um ein differenziertes Bild der QI zu zeichnen. Sollten die QI erneut im Rahmen einer Delphi-Umfrage analysiert werden, wäre eine Standardisierung der Kriterien ein weiterer Aspekt zur Verbesserung der Aussagekraft der QI-Analyse.

QI-Empfehlung und -Vorlage

Die QI-Empfehlung und die Priorisierung der QI innerhalb der QI-Empfehlung basieren auf der Verallgemeinerung der durchgeführten Analyse der QI für die drei CERTAIN-Register und unterliegen somit Beschränkungen in der Übertragbarkeit, die eine kritische Hinterfragung der QI-Empfehlung und der Priorisierungen durch den Anwender bedingen. Die Priorisierung der QI innerhalb der Kategorien wurde durch die drei Registerzentralen bestätigt und bezieht hierdurch unterschiedliche Ansichten mit ein, um eine einseitige Priorisierung zu vermeiden.

Konzept und Implementierung von Validierungsregeln zur automatisierten Qualitätssicherung

Die Validierungsregeln der entwickelten Software-Komponenten zeigten eine sehr unterschiedliche Reichweite. Einige der neuen *ConstraintValidator* wurden letztlich nur auf einige tausend Objekte angewendet. Eine Vorabanalyse der Reichweite jedes zur Umsetzung ausgewählten *ConstraintValidator* hätte zu einer effektiveren Umsetzung der QI geführt. So wurden zwar QI entwickelt, die derzeit nur eine geringe Reichweite erreichen. Es konnten jedoch für nahezu alle QI der Kategorie „Automatische Validierung“ *ConstraintValidator* entwickelt werden.

Da der Einfluss der zu entwickelnden Mechanismen später ausgewertet werden sollte, war ein möglichst großer Datenbestand für aussagekräftige Ergebnisse erforderlich. Die Aktivität der Dateneingabe des CERTAIN-LI-Registers war im Verhältnis zu der des CERTAIN-Registers nicht sehr hoch. Das KRhOKo-Register war zum Zeitpunkt der Entwicklung der *ConstraintValidator* noch nicht gestartet. Beide Register konnten daher nicht die erforderliche Größe der Datenbestände aufweisen, anhand derer aussagekräftige Schlüsse hätten gezogen werden können. Die Auswahl und Entwicklung der Validierungsregeln war daher durch die Möglichkeiten der Anwendung im CERTAIN-Register geprägt.

Die Nutzung der Java Bean Validation und die Entwicklung der Software-Komponenten auf Basis des Hibernate Validator erwiesen sich als geeignete Wahl, da mit JSR-380 durch technische Erweiterungen mehr Möglichkeiten hinsichtlich der Umsetzung von Warnhinweisen ermöglicht wurden. Durch beide Frameworks wurde zudem die Übertragbar- und Wiederverwendbarkeit der Software-Komponenten gewährleistet.

Analyse der Datenqualität des CERTAIN-Registers

Die Analysen zur Datenqualität der Datenbestände erfolgten mittels des Hibernate Validators und somit mit einem erprobten Werkzeug, das sich als geeignet für diesen Zweck herausstellte. Durch die detaillierte Ausgabe der Verstöße gegen die Validierungsregeln durch den Hibernate Validator können einzelne Sachverhalte oder *ConstraintValidator* gezielt betrachtet werden, um spezielle Fragestellungen hinsichtlich der Datenqualität beantworten zu können. Die speziell entwickelten Auswertungsmethoden zeigten sich als sehr nützlich

zur Feststellung von Problemen in den Datenbeständen und zur Prüfung der entwickelten *ConstraintValidator* gegen reale Datenbestände. Da diese speziellen Auswertungsmethoden sich auf die *ConstraintValidator RangeValidator* und *EntryRangeValidator* zur Erkennung der für ihre Berechnung einzubeziehenden Datenelemente verlassen, sind die Auswertungsmethoden bei Nutzung dieser *ConstraintValidator* ebenso auf andere Register übertragbar. Alle in die Analyse einbezogenen Datenbestände boten eine hinreichend große Anzahl an Visiten, so dass aussagekräftige Ergebnisse erzielt werden konnten.

Der entwickelte Algorithmus zur Nachverfolgung von Änderungen an Ausreißern zeigte sich weitgehend als adäquat, konnte jedoch nicht unterscheiden, wann ein Ausreißer von einem Datenelement in ein anderes übertragen wurde. Hierdurch wurden mehr Ausreißer als gelöscht identifiziert als tatsächlich gelöscht wurden. Eine Anpassung des Algorithmus diesbezüglich konnte nicht durchgeführt werden, da die Anzahl der Datenelemente, in die jeder Ausreißer übertragen werden kann, schlicht zu groß und das Problem damit zu komplex ist. Eine manuelle Durchsicht der Änderungen an den Ausreißern bleibt damit weiter unabdingbar, um schlüssige Aussagen treffen zu können.

Für die Analyse des Datenbestandes hinsichtlich des Vorkommens von Werte-Clustern wäre die Erkennung von Auffälligkeiten durch Prüfung auf das Vorkommen von Werten gewesen, die nicht der erwarteten Verteilung des Datenelementes folgen. Dies erwies sich für diese Arbeit nicht geeignet, da die Verteilungen der Datenelemente des Datenbestandes sehr heterogen waren und die Erkennung der Verteilungen nicht automatisiert werden konnte. Der im Rahmen dieser Arbeit entwickelte Ansatz zur Erkennung von Clustern blieb ebenfalls hinter den Erwartungen zurück. Literatur wurde zu diesem Problem gesichtet, eine befriedigende, automatisierbare Lösung konnte jedoch nicht gefunden werden und könnte Gegenstand weiterer Entwicklungen sein.

9.3 Ausblick

Die entwickelten Ergebnisse der QI-Analyse und der QI-Empfehlung sollten durch die Zusammenarbeit mit anderen Registern validiert und gegebenenfalls weiterentwickelt werden. Eine Delphi-Umfrage kann hierfür ein mögliches Werkzeug darstellen, mit dem die Einschätzungen anderer Register erfasst werden können. Diese sollte insbesondere für eine eingehendere Analyse der QI genutzt werden, die nicht auf die drei CERTAIN-Register angewendet werden konnten.

Im Zuge einer neuen Studie, die über das KRhOKo-Register durchgeführt werden soll, wird mit einer steigenden Anzahl an Fragebogen gerechnet, die von Patienten direkt über das Register ausgefüllt werden. Aufgrund dieser voraussichtlichen Zunahme der Nutzung der Patientenbeteiligung sollte versucht werden, die Gründe für die Unterschiede in der Vollständigkeit zwischen Eltern- und Jugendlichenbogen festzustellen und diese durch technische Maßnahmen zu adressieren.

Nicht alle Wertebereiche des Datenvalidierungsplans des CERTAIN-Registers werden bereits validiert. Da aus Gründen der kontrollierten Einführung der Einheitenvalidierung anfänglich nur SI-Einheiten umgerechnet werden sollten, wurden Datenelemente mit Baseinheiten, die nicht zu SI-Einheiten kompatibel sind, zunächst von der Validierung der Wertebereiche ausgeschlossen. Im Hinblick auf die erfolgreiche Reduktion der Ausreißer im Datenbestand des CERTAIN-Registers sollte auch die Validierung der verbleibenden Wertebereiche des Datenvalidierungsplans umgesetzt werden, um eine ganzheitliche Validierung von Laborwerten und Medikationsangaben zu erreichen. Dies sollte einhergehen mit einer genaueren Betrachtung der Korrektur der Ausreißer durch die jeweiligen Einrichtungen. Der Algorithmus der Funktion *InvestigateOutliers* sollte hierfür gegebenenfalls überarbeitet werden, um die erfolgte Änderung eines Ausreißers besser nachvollziehen zu können.

Im Zuge der Erweiterung der Validierung der Wertebereiche sollten auch die Eingabemöglichkeiten der Einheiten der validierten Datenelemente auf Einheiten beschränkt werden, die für das jeweilige Datenelement sinnvoll erscheinen. Da es auch in den neu angelegten Visiten immer wieder zu Tippfehlern oder zur Auswahl von nicht plausiblen Einheiten kommt, wie beispielsweise der Einheit Sekunden für den Blutdruck, besteht auch hier Handlungsbedarf.

Die Nutzung von Strings zur Unterscheidung zwischen Fehlermeldungen und Warnhinweisen bei Verstöße gegen die Validierungsregeln war bereits während der Implementierung als Übergangslösung identifiziert worden. Durch den Hibernate Validator in Version 6.x wurde es möglich, die *payload* der *Constraint Annotation* dynamisch anzupassen und die Verstöße gegen die Validierungsregeln dynamisch zu typisieren, so dass keine Identifikation von Warnhinweisen über Strings mehr erforderlich ist. Der String-basierte Ansatz sollte hierdurch abgelöst werden.

Die zur Validierung entwickelten Software-Komponenten können nicht nur prospektiv zur Sicherung guter Datenqualität während der Eingabe neuer Visiten genutzt werden. Durch die Validierung von Bestandsdaten gegen die neuen Validierungsregeln könnten diese Validierungsregeln auch zur automatischen Generierung von Queries an die Einrichtungen genutzt werden, um auch retrospektiv die Datenqualität zu verbessern. Dies bietet sich besonders für Register an, in denen eine nachträgliche Bearbeitung der Daten durch die Eingabekräfte und somit die erneute Anzeige von Verstößen gegen die Validierungsregeln für die Eingabekraft nicht zu erwarten ist.

Für alle drei Register sollte zudem sichergestellt werden, dass die Auswahllisten bei Datenelementen mit Auswahllisten auf Standards oder kontrollierten Vokabularen beruhen. Eine solche Überarbeitung könnte die Vergleichbarkeit der Daten der Register und die Qualität der erhobenen Daten verbessern.

Letztlich sollte auch der Einfluss der automatischen Validierung auf die Arbeit des DQM untersucht werden. Durch die vermehrte Überprüfung einfacher Sachverhalte könnte der

DQM sich zu sehr auf die Überprüfungen durch das System verlassen und bei der Überprüfung der Daten nachlässig werden.

Zu guter Letzt sollte auch weiterhin an der automatischen Validierung gearbeitet und diese auch künftig um die Überprüfung neuer Sachverhalte ergänzt werden. Da erst in Folge der neuen Validierung die beschriebenen Missstände erfasst werden konnten, ist nicht anzunehmen, dass die aktuelle Validierung alle möglichen Fehlerquellen beseitigt hat. Selbst wenn zur Überzeugung gelangt wird, dass die automatische Validierung komplett sei, besteht weiterhin die Möglichkeit doch noch das adaptive Feedback der TMF-Leitlinie aufzugreifen und die QI der Kategorie „Report“ für die drei Registern zu berechnen.

10 Zusammenfassung

Seltene Erkrankungen sind aufgrund der geringen Fallzahlen eine Herausforderung für die klinische Forschung. Register stellen eines der Instrumente zur systematischen Dokumentation von Verläufen von seltenen Erkrankungen dar. Ihre Aussagekraft hängt stark von der Qualität der Daten ab und somit von Mechanismen, welche die Qualität absichern sollen. Für einen Verbund aus drei Registern (CERTAIN, CERTAIN-LI und KRhOKo) sollten die Maßnahmen zur Sicherung die Datenqualität auf der Basis von automatischen Mechanismen verbessert werden.

Hierfür wurde ein allgemeines Vorgehen zur Verbesserung der Datenqualität auf Basis der Qualitätsindikatoren (QI) der TMF-Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern entwickelt. Dieses Vorgehen bündelt die 51 QI der Leitlinie in vier QI-Kategorien mit unterschiedlichen Ansatzpunkten zur Verbesserung der Datenqualität: Registerplanung, automatische Validierung, Report und Datenverifizierung. Anhand des allgemeinen Vorgehens kann eine Gewichtung der QI für die Berechnung des Gesamtscores der TMF-Leitlinie erfolgen. Diese stützt sich auf eine initial durchzuführende Analyse der QI der TMF-Leitlinie im Hinblick auf die Kriterien Anwendbarkeit, Implementierbarkeit und Aussagekraft. Auf Basis des Ergebnisses der QI-Analyse verbleiben nur die für das jeweilige Register relevanten QI, für die über eine Priorisierung der QI innerhalb der identifizierten QI-Kategorien die Gewichte bestimmt werden können. Das allgemeine Vorgehen wurde für die drei genannten Register vollständig für alle QI der Leitlinie durchgeführt und die Ergebnisse der QI-Analyse, der Priorisierung der QI und der Zuweisung der Gewichte wurden detailliert dargestellt. Die QI-Analyse umfasst alle QI der TMF-Leitlinie und bietet eine detaillierte Betrachtung der QI. Aus diesen Ergebnissen wurde eine QI-Empfehlung entwickelt, die das entwickelte Vorgehen allgemein für elektronische Register vereinfacht und eine beispielhafte Priorisierung der QI bereitstellt. Auf Basis der QI-Analyse für die drei Register wurden Software-Komponenten zur Verbesserung der Datenqualität in Form von generischen Java-Modulen für die QI der Kategorie „automatische Validierung“ entwickelt, diese in die drei Register eingefügt und als Teil der Maßnahmen zur Sicherung der Datenqualität während der Dateneingabe genutzt. Die hierfür entwickelten Konzepte und Implementierungen wurden allgemein gehalten, so dass eine Übertragbarkeit der Konzepte und Java-Module auf andere Register gegeben ist. Die Effektivität dieser neuen Software-Komponenten konnte durch die umfangreiche Auswertung der Datenqualität des Datenbestandes des CERTAIN-Registers im Rahmen dieser Dissertation bestätigt werden.

Somit konnte hier ein allgemeines Vorgehen zur Verbesserung der Datenqualität in medizinischen Registern durch automatisierte Mechanismen sowie ein Beispiel einer konkreten Umsetzung dieses Vorgehens und der Mechanismen dargestellt werden. Die QI-Empfehlung und die Ergebnisse der QI-Analyse für die drei Register stehen für andere zur Verbesserung ihrer Datenqualität als Basis zur Verfügung. Gleiches gilt für die Konzepte der entwickelten Software-Komponenten.

Glossar

Automatische Validierung	Die Durchführung der Validierung eines Datenwertes durch elektronische, automatisierte Vorgänge. In dieser Dissertation findet die automatische Validierung während der Eingabe der Daten in das Register statt. Das Registersystem gibt auf der Basis der erfolgten Validierung Rückmeldung an den Benutzer.
Beobachtungseinheit	Je nach Vorhaben können Beobachtungseinheiten Personen, Patienten, Behandlungsfälle, Meldungen, Erhebungsbogen, etc. darstellen Nonnemacher et al. 2014, S, 9.
Class-level-Annotation	Eine Java-Annotation mit der eine Klasse annotiert werden kann.
Datenbestand	Die gesammelten Daten einer Einrichtung in einem Register.
Datenelement	Datenelemente speichern Merkmalsausprägungen als Datenwert.
Datensatz	Ein Datensatz ist eine Sammlung von Modulen und Datenelementen.
Digitalisierte Daten	Digitalisierte Daten sind Daten, die beispielsweise über einen Papierfragebogen erhoben und durch einen manuellen oder maschinellen Prozess in ein digitales Format überführt wurden. Im Rahmen dieser Dissertation handelt es sich bei digitalisierten Daten um die digitalisierten Fragebogen der papierbasierten Kerndokumentation.
Ereignis	Ein Ereignis beschreibt in dieser Arbeit einen Verstoß gegen eine Validierungsregel.
Extendierter Datensatz	Der extendierte Datensatz der beiden Transplantationsregister CERTAIN und CERTAIN-LI umfasst optionale Datenelemente, die von geringerer Bedeutung für die medizinischen Auswertungen zu Forschungsfragen sind. Hierzu zählen auch die Einträge der fortlaufenden Dokumentation.

Field-level-Annotation	Eine Java-Annotation mit der ein Feld innerhalb einer Klasse annotiert werden kann.
Minimaler Datensatz	Der minimale Datensatz der beiden Transplantationsregister CERTAIN und CERTAIN-LI umfasst alle elementar wichtigen Datenelemente und besteht nahezu vollständig aus Pflichtfeldern.
Studiendatensatz	Ein Studiendatensatz innerhalb der beiden Transplantationsregister CERTAIN und CERTAIN-LI umfasst Datenelemente, die spezifisch für konkrete Forschungsfragen im Rahmen von Studien erhoben werden und die sich nicht im minimalen oder extendierten Datensatz des jeweiligen Registers finden. Studiendatensätze erweitern den Datensatz einer Visite.
Validierung	Der Vorgang der Validierung beschreibt die Überprüfung der Gültigkeit eines Datenwerts.
Verifizierung	Der Vorgang der Verifikation beschreibt die Überprüfung einer Sache auf Richtigkeit.
Verstoß gegen Validierungsregeln	Ein Verstoß gegen Validierungsregeln ist das Ergebnis der Validierung eines Objektes durch einen ConstraintValidator, sollte das Objekt der Prüfung durch den ConstraintValidator gegen die zu gehörige Validierungsregel nicht standhalten.

Literatur

- Abate, M., Diegert, K. und Allen, H. (1998). **A hierarchical approach to improving data quality**. In: *Data Quality* 4.1, S. 365–369.
- Amoroso, W. P., Borrebach, D., Kuntz, T. E., Kamons, L. A., Martin, J. und Remaley, N. (1989). **The PoP-Parent Software System: a distributed data processing and clinical trial management system for the NHLBI BARI**. In: *Controlled Clinical Trials* 10.3, S. 339.
- Aronsky, D. und Haug, P. J. (2000). **Assessing the quality of clinical data in a computer-based record for calculating the pneumonia severity index**. In: *Journal of the American Medical Informatics Association* 7.1, S. 55–65.
- Arts, D. G., De Keizer, N. F. und Scheffer, G.-J. (2002). **Defining and improving data quality in medical registries: a literature review, case study, and generic framework**. In: *Journal of the American Medical Informatics Association* 9.6, S. 600–611.
- Arts, D. G., Keizer, N. F. de und Jonge, E. de (2001). **Data quality measurement and assurance in medical registries**. In: *Studies In Health Technology And Informatics* 1, S. 404–404.
- Barrie, J. und Marsh, D. (1992). **Quality of data in the Manchester orthopaedic database**. In: *BMJ* 304.6820, S. 159–162.
- Bautista, M. P., Bonilla, G. A., Mieth, K. W., Llinás, A. M., Rodríguez, F. und Cárdenas, L. L. (2017). **Data Quality in Institutional Arthroplasty Registries: Description of a Model of Validation and Report of Preliminary Results**. In: *The Journal of Arthroplasty* 32.7, S. 2065–2069.
- Bellgard, M., Beroud, C., Parkinson, K., Harris, T., Ayme, S., Baynam, G., Weeramanthri, T., Dawkins, H. und Hunter, A. (2013). **Dispelling myths about rare disease registry system development**. In: *Source Code for Biology and Medicine* 8.1, S. 21.
- Beretta, L., Aldrovandi, V., Grandi, E., Citerio, G. und Stocchetti, N. (2007). **Improving the quality of data entry in a low-budget head injury database**. In: *Acta Neurochirurgica* 149.9, S. 903–909.
- Bernard, E. (2013). **Bean Validation specification**. <http://beanvalidation.org/1.1/spec/>. [Stand 31.01.2018].
- Bernard, E., Benson, M., Lee, A., Midwinter, R., Morling, G., Burns, E., Purushothaman, J., Ramu, J., Taube, M., Petracek, G., Ferentschik, H. und Santos, M. N. (2013). **The Java Community Process(SM) Program - JSRs: Java Specification Requests - detail JSR 349**. <https://jcp.org/en/jsr/detail?id=349>. [Stand 31.01.2018].
- Bernard, E., Bevin, G., Boness, U., Carreira, J., Ericsson, A., Fennell, T., Harrop, R., Hookom, J. J., Pemberton, N. K., Peterson, S., Petracek, G., Prasanna, D. R., Raible, M., Santos, M. N. und Thomschke, S. (2009). **The Java Community Process(SM)**

- Program - JSRs: Java Specification Requests - detail JSR 303.** <https://jcp.org/en/jsr/detail?id=303>. [Stand 31.01.2018].
- Bernard, E. und Peterson, S. (2009). **JSR 303: Bean Validation.** <http://beanvalidation.org/1.0/spec/>. [Stand 31.01.2018].
- Bobrowski, M., Marré, M. und Yankelevich, D. (1999). **A Homogeneous Framework to Measure Data Quality.** In: *IQ*, S. 115–124.
- Bray, F. und Parkin, D. M. (2009). **Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness.** In: *European Journal of Cancer* 45.5, S. 747–755.
- Brennan, P. und Stead, W. (2000). **Assessing data quality: from concordance, through correctness and completeness, to valid manipulatable representations.** In: *Journal of the American Medical Informatics Association: JAMIA* 7.1, S. 106–107.
- Brunkhorst, L. C., Fichtner, A., Höcker, B., Burmeister, G., Ahlenstiel-Grunow, T., Krupka, K., Bald, M., Zapf, A., Tönshoff, B. und Pape, L. (2015). **Efficacy and safety of an everolimus-vs. a mycophenolate mofetil-based regimen in pediatric renal transplant recipients.** In: *PloS One* 10.9, e0135439.
- Califf, R. M., Karnash, S. L. und Woodlief, L. H. (1997). **Developing systems for cost-effective auditing of clinical trials.** In: *Controlled Clinical Trials* 18.6, S. 651–660.
- CERTAIN-LI (2018). **CERTAIN-LI Cooperative European Paediatric Transplant Initiative Liver: About.** <http://www.certainli-registry.eu/index.html>. [Stand 24.01.2018].
- Cnattingius, S., Ericson, A., Gunnarskog, J. und Källén, B. (1990). **A quality study of a medical birth registry.** In: *Scandinavian Journal of Social Medicine* 18.2, S. 143–148.
- Cook, S. F., Visscher, W. A., Hobbs, C. L., Williams, R. L., Committee, P. I. C. I. et al. (2002). **Project IMPACT: results from a pilot validity study of a new observational database.** In: *Critical Care Medicine* 30.12, S. 2765–2770.
- Couchoud, C., Lassalle, M., Cornet, R. und Jager, K. J. (2013). **Renal replacement therapy registries—time for a structured data quality evaluation programme.** In: *Nephrology Dialysis Transplantation* 28.9, S. 2215–2220.
- Donabedian, A. (1988). **The quality of care: how can it be assessed?** In: *Jama* 260.12, S. 1743–1748.
- Fine, L. G., Keogh, B. E., Cretin, S., Orlando, M. und Gould, M. M. (2003). **How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback study on the UK Cardiac Surgery Experience.** In: *BMJ* 326.7379, S. 25–28.
- Gassman, J. J., Owen, W. W., Kuntz, T. E., Martin, J. P. und Amoroso, W. P. (1995). **Data quality assurance, monitoring, and reporting.** In: *Controlled Clinical Trials* 16.2, S. 104–136.

- Gaus, W. (2003). **Dokumentation und Datenverarbeitung bei klinischen Studien.** Bd. 1. BoD–Books on Demand.
- Gissler, M., Ulander, V.-M., Hemminki, E. und Rasimus, A. (1996). **Declining induced abortion rate in Finland: data quality of the Finnish abortion register.** In: *International Journal of Epidemiology* 25.2, S. 376–380.
- Gliklich, R. E., Dreyer, N. A., Leavy, M. B. et al. (2014). **Registries for evaluating patient outcomes: a user’s guide.** 13. Government Printing Office.
- Goldberg, J., Gelfand, H. M. und Levy, P. S. (1980). **Registry evaluation methods: a review and case study.** In: *Epidemiologic Reviews* 2.1, S. 210–220.
- Habbig, S., Volland, R., Krupka, K., Querfeld, U., Dello Strologo, L., Noyan, A., Yalcinkaya, F., Topaloglu, R., Webb, N. J. A., Kemper, M. J., Pape, L., Bald, M., Kranz, B., Taylan, C., Höcker, B., Tönshoff, B. und Weber, L. T. (2017). **Dyslipidemia after pediatric renal transplantation—The impact of immunosuppressive regimens.** In: *Pediatric Transplantation* 21.3. e12914, e12914–n/a.
- Håkansson, I., Lundström, M., Stenevi, U. und Ehinger, B. (2001). **Data reliability and structure in the Swedish National Cataract Register.** In: *Acta Ophthalmologica* 79.5, S. 518–523.
- Hasford, J. (1994). **Arzneimittelprüfungen und good clinical practice: Planung, Durchführung und Qualitätssicherung.** MMV Medizin Verlag, München.
- Hassey, A., Gerrett, D. und Wilson, A. (2001). **A survey of validity and utility of electronic patient records in a general practice.** In: *BMJ* 322.7299, S. 1401–1405.
- Hentschel, S. und Katalinic, A. (2008). **Das Manual der epidemiologischen Krebsregistrierung.** W. Zuckschwerdt Verlag, München.
- Hlaing, T., Hollister, L. und Aaland, M. (2006). **Trauma registry data validation: essential for quality trauma care.** In: *Journal of Trauma and Acute Care Surgery* 61.6, S. 1400–1407.
- Höcker, B., Aguilar, M., Schnitzler, P., Pape, L., Bald, M., König, J., Marks, S. D., Genc, G., Büscher, A., Kemper, M. J., Billing, H., Pohl, M., Dello Strologo, L., Webb, N. J. A., Rieger, S., Mankertz, A., Krupka, K., Bruckner, T., Fichtner, A. und Tönshoff, B. (2018a). **Vaccination titres pre- and post-transplant in paediatric renal transplant recipients and the impact of immunosuppressive therapy.** In: *Pediatric Nephrology*.
- Höcker, B., Aguilar, M., Schnitzler, P., Pape, L., Dello Strologo, L., Webb, N. J. A., Bald, M., Genc, G., Billing, H., König, J., Büscher, A., Kemper, M. J., Marks, S. D., Pohl, M., Wigger, M., Topaloglu, R., Rieger, S., Krupka, K., Bruckner, T., Fichtner, A. und Tönshoff, B. (2018b). **Incomplete vaccination coverage in European children with end-stage kidney disease prior to renal transplantation.** In: *Pediatric Nephrology* 33.2, S. 341–350.

- Höcker, B., Zencke, S., Pape, L., Krupka, K., Köster, L., Fichtner, A., Dello Strologo, L., Guzzo, I., Topaloglu, R., Kranz, B., König, J., Bald, M., Webb, N. J. A., Noyan, A., Dursun, H., Marks, S., Ozcakar, Z. B., Thiel, F., Billing, H., Pohl, M., Fehrenbach, H., Schnitzler, P., Bruckner, T., Ahlenstiel-Grunow, T. und Tönshoff, B. (2016a). **Impact of Everolimus and Low-Dose Cyclosporin on Cytomegalovirus Replication and Disease in Pediatric Renal Transplantation.** In: *American Journal of Transplantation* 16.3, S. 921–929.
- Höcker, B., Zencke, S., Krupka, K., Fichtner, A., Pape, L., Strologo, L. D., Guzzo, I., Topaloglu, R., Kranz, B., König, J. et al. (2016b). **Cytomegalovirus infection in pediatric renal transplantation and the impact of chemoprophylaxis with (val-) ganciclovir.** In: *Transplantation* 100.4, S. 862–870.
- Hoffmann, W., Latza, U., Terschüren, C., Berger, K., Bergmann, M., Dahl, A., Fendrich, K., Gierer, S., Holle, R., Jahn, I. et al. (2004). **Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP)–Langversion.**
- Hogan, W. R. und Wagner, M. M. (1997). **Accuracy of data in computer-based patient records.** In: *Journal of the American Medical Informatics Association* 4.5, S. 342–355.
- Holle, R. (1995). **Methoden zur Konstruktion und Evaluierung klinischer Scores.** Habilitationsschrift. Inst. für Med. Biometrie u. Informatik, Abt. Med. Biometrie.
- Hüßers, N. (2004). **Einführung in die (analytische) Qualitätssicherung.**
- Jacke, C., Kalder, M., Etter, J., Koller, M., Mueller, U., Wagner, U. und Albert, U. (2010). **Optimierung der Datenqualität von Kohortenstudien mittels Qualitätsindikatoren.** In: *Monitor Versorgungsforschung*. Bd. 3 (Sonderausgabe). 54.
- Jensen, A. R., Overgaard, J. und Storm, H. (2002). **Validity of breast cancer in the Danish Cancer Registry. A study based on clinical records from one county in Denmark.** In: *European Journal of Cancer Prevention* 11.4, S. 359–364.
- Jung, R. und Winter, R. (2000). **Data Warehousing Strategie: Erfahrungen, Methoden, Visionen.** Springer, Berlin.
- Katalinic, A. (2005). **Qualitätssicherung von Daten in epidemiologischen Krebsregistern.** Vortrag gehalten auf dem Workshop Medizinische Daten und Versorgung, 16.02.2005, Essen.
- Khosla, R., Verma, D., Kapur, A., Khosla, S. et al. (2000). **Efficient source data verification.** In: *Indian Journal of Pharmacology* 32.3, S. 180–186.
- Köster, L., Krupka, K., Höcker, B., Rahmel, A., Samuel, U., Zanen, W., Opelz, G., Süßal, C., Döhler, B., Plotnicki, L. et al. (2015). **Integrating data from multiple sources for data completeness in a web-based registry for pediatric renal transplantation—the CERTAIN Registry.** In: *Studies in Health Technology and Informatics* 216, S. 1049–1049.

- Kühn-Stoffers, P. (2004). **Script zur Vorlesung Grundlagen der Analytik II. Teil 3 Qualitätssicherung in der Analytik.**
- Laporte, R. E., Mazumdar, S. et al. (1994). **Approaches to quality control with an application to a new cancer registry in a developing country.** In: *Journal of Clinical Epidemiology* 47.7, S. 779–786.
- Leys, C., Ley, C., Klein, O., Bernard, P. und Licata, L. (2013). **Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median.** In: *Journal of Experimental Social Psychology* 49.4, S. 764–766.
- Lim, T.-O., Goh, A., Lim, Y.-N. und Morad, Z. (2008). **Use of renal registry data for research, health-care planning and quality improvement: What can we learn from registry data in the Asia–Pacific region?** In: *Nephrology* 13.8, S. 745–752.
- Lin, C.-M., Lee, P.-C., Teng, S.-W., Lu, T.-H., Mao, I.-F. und Li, C.-Y. (2004). **Validation of the Taiwan Birth Registry using obstetric records.** In: *Journal of the Formosan Medical Association= Taiwan yi zhi* 103.4, S. 297–301.
- Lindquist, M. (2004). **Data quality management in pharmacovigilance.** In: *Drug Safety* 27.12, S. 857–870.
- Logan, J. R., Gorman, P. N. und Middleton, B. (2001). **Measuring the quality of medical records: a method for comparing completeness and correctness of clinical encounter data.** In: *Proceedings of the AMIA Symposium*. American Medical Informatics Association, S. 408.
- Lu, J.-H., Hung, J.-H., Lin, F.-M., Shen, W.-Y., Chen, S.-J., Hwang, B., Wu, S.-I. und Chao, Y.-M. Y. (1995). **A Quality Study of a Computerized Medical Birth Registry.** In: *Journal of Obstetrics and Gynaecology Research* 21.1, S. 23–30.
- Maruszewski, B., Lacour-Gayet, F., Monro, J. L., Keogh, B. E., Tobota, Z. und Kansy, A. (2005). **An attempt at data verification in the EACTS Congenital Database.** In: *European Journal of Cardio-Thoracic Surgery* 28.3, S. 400–404.
- Massey, J. T. und Hoffman, K. L. (1989). **Monitoring data quality through comparisons between data systems.** In: *Statistics in Medicine* 8.3, S. 367–377.
- Mei, Z., Yip, R. und Trowbridge, F. (1998). **Improving trend of growth of Asian refugee children in the USA: evidence to support the importance of environmental factors on growth.** In: *Asia Pacific Journal of Clinical Nutrition* 7, S. 111–116.
- Messenger, J. C., Ho, K. K., Young, C. H., Slattery, L. E., Draoui, J. C., Curtis, J. P., Dehmer, G. J., Grover, F. L., Mirro, M. J., Reynolds, M. R. et al. (2012). **The National Cardiovascular Data Registry (NCDR) data quality brief: the NCDR data quality program in 2012.** In: *Journal of the American College of Cardiology* 60.16, S. 1484–1488.
- Minden, K., Niewerth, M., Listing, J., Zink, A., Pediatric Rheumatologists, G. S. G. of et al. (2002). **Health care provision in pediatric rheumatology in Germany–**

- national rheumatologic database.** In: *The Journal of Rheumatology* 29.3, S. 622–628.
- Morling, G. (2017). **Bean Validation specification.** <http://beanvalidation.org/2.0/spec/>. [Stand 31.01.2018].
- Morling, G., Ebberts, H., Yoshitomi, T., Molteni, M., Kaltepoth, C., Keil, W., Demichiel, L., Bernard, E. und Santos, M. N. (2017). **The Java Community Process(SM) Program - JSRs: Java Specification Requests - detail JSR 380.** <https://jcp.org/en/jsr/detail?id=380>. [Stand 31.01.2018].
- Mullooly, J. P. (1990). **The effects of data entry error: an analysis of partial verification.** In: *Computers and Biomedical Research* 23.3, S. 259–267.
- Naumann, F. und Rolker, C. (2000). **Assessment Methods for Information Quality Criteria.** In: *In Proceedings of the International Conference on Information Quality (IQ)*. Citeseer.
- Newhouse, M. (1985). **Data entry design and data quality.** In: *Controlled Clinical Trials*. Bd. 6. 3, S. 229–229.
- Nielsen, G. L., Sørensen, H. T., Pedersen, A. B. und Sabroe, S. (1996). **Analyses of data quality in registries concerning diabetes mellitus—a comparison between a population based hospital discharge and an insulin prescription registry.** In: *Journal of Medical Systems* 20.1, S. 1–10.
- Nonnemacher, M., Nasseh, D. und Stausberg, J. (2014). **Datenqualität in der medizinischen Forschung: Leitlinie zum Adaptiven Datenmanagement in Kohortenstudien und Registern.** Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
- Nonnemacher, M., Weiland, D. und Stausberg, J. (2007). **Datenqualität in der medizinischen Forschung: Leitlinie zum Adaptiven Datenmanagement in Kohortenstudien und Registern.** Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
- O’Kelly, M. (2004). **Using statistical techniques to detect fraud: a test case.** In: *Pharmaceutical Statistics* 3.4, S. 237–246.
- O’Reilly, G. M., Gabbe, B., Moore, L. und Cameron, P. A. (2016). **Classifying, measuring and improving the quality of data in trauma registries: a review of the literature.** In: *Injury* 47.3, S. 559–567.
- Ose, C., Weiland, D. und Jockel, K.-H. (2004). **Low Budget GCP—am Beispiel Monitoring. How to practice low cost monitoring in clinical trials.** In: *Informatik Biometrie und Epidemiologie in Medizin und Biologie* 35.1, S. 54–62.
- O’Sullivan, D., Needham, C., Bangs, A., Atkin, K. und Kendall, F. (1996). **Postmarketing surveillance of oral terbinafine in the UK: report of a large cohort study.** In: *British Journal of Clinical Pharmacology* 42.5, S. 559–565.
- Parkin, D., Muir, C., Whelan, S., Gao, Y., Ferlay, J. und Powell, J. (1992). **IARC Scientific Publications No. 120.** In: *Cancer Incidence in Five Continents* 6.

- Plotnicki, L., Kohl, C., Höcker, B., Krupka, K., Rahmel, A., Pape, L., Hoyer, P., Marks, S., Webb, N., Söylemezoglu, O. et al. (2013). **The CERTAIN Registry: a novel, web-based registry and research platform for pediatric renal transplantation in Europe.** In: *Transplantation proceedings*. Bd. 45. 4. Elsevier, S. 1414–1417.
- Pogash, R., Boehmer, S., Forand, P., Dyer, A. und Kunselman, S. (2001). **Data management procedures in the Asthma Clinical Research Network.** In: *Controlled Clinical Trials* 22.6 Suppl, 168S.
- Pollock, B. (1994). **Quality assurance for interventions in clinical trials. Multicenter data monitoring, data management, and analysis.** In: *Cancer* 74.9 Suppl, S. 2647.
- Pommerening, K., Sax, U., Müller, T., Speer, R., Ganslandt, T., Drepper, J. und Semler, S. C. (2009). **Das TMF-Datenschutzkonzept für medizinische Datensammlungen und Biobanken.** In: *GI Jahrestagung*, S. 1744–1757.
- Prins, H., Kruisinga, F., Büller, H. und Zwetsloot-Schonk, J. (2000). **Availability and accuracy of electronic patient data for medical practice assessment.** In: *Studies in Health Technology and Informatics* 77, S. 484.
- Prokein, J., Steinmann, S., Wiese, B. und Schäfer, I. (2011). **Bestimmung der Datenqualität und Qualitätsberichterstattung in epidemiologischen Studien am Beispiel der Studie MultiCare1.** In: *Mainz//2011. 56. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), 6. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi)*. German Medical Science GMS Publishing House.
- Schmidt, C. O., Stausberg, J., Pritzkeleit, R., Schrader, T. und Nonnemacher, M. (2012a). **TMF-Projekt V020-04: Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID).**
- (2012b). **TMF-Projekt V020-04: Open European Nephrology Science Center (OpEN.SC).**
- (2012c). **TMF-Projekt V020-04: Study in Health in Pomerania (SHIP).**
- Schmidt, J., Gertzen, H., Aschenbrenner, K. und Ryholt-Jensen, S. (1995). **Detecting Fraud Using Auditing and Biometrical Methods.** In: *Applied Clinical Trials* 4, S. 40–56.
- Svolba, G. und Bauer, P. (1999). **Statistical quality control in clinical trials.** In: *Controlled Clinical Trials* 20.6, S. 519.
- Swart, E., Ihle, P., Gothe, H. und Matusiewicz, D. (2005). **Routinedaten im Gesundheitswesen.** Verlag Hans Huber, Bern.
- Teperi, J. (1993). **Multi method approach to the assessment of data quality in the Finnish Medical Birth Registry.** In: *Journal of Epidemiology & Community Health* 47.3, S. 242–247.

- Teppo, L., Pukkala, E. und Lehtonen, M. (1994). **Data quality and quality control of a population-based cancer registry: experience in Finland.** In: *Acta Oncologica* 33.4, S. 365–369.
- Teufel, U., Ganschow, R., Kathemann, S., Lainka, E., Grabhorn, E., Sturm, E., Luck, W., Schulz-Juergensen, S., Krupka, K., Melter, M., Wenning, D. und Tönshoff, B. (2015). **Das CERTAIN-LI Register: Ein Web-basiertes Register für die pädiatrische Lebertransplantation in Deutschland.** In: *Pädiatrie – Kinder- und Jugendmedizin hautnah*. Bd. 27. 2, S. 82–83.
- TMF (2018). **Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. - Über uns.** http://www.tmf-ev.de/Ueber_uns.aspx. [Stand 31.01.2018].
- Tolonen, H., Dobson, A., Kulathinal, S. und Project*, W. M. (2006). **Assessing the quality of risk factor survey data: lessons from the WHO MONICA Project.** In: *European Journal of Cardiovascular Prevention & Rehabilitation* 13.1, S. 104–114.
- Tönshoff, B., Krupka, K., Köster, L., Fichtner, A., Pape, L. und Höcker, B. (2015). **CERTAIN Registry.** In: *Der Nephrologe* 10.6, S. 480–487.
- Topp, M., Langhoff-Roos, J. und Uldall, P. (1997). **Validation of a cerebral palsy register.** In: *Journal of Clinical Epidemiology* 50.9, S. 1017–1023.
- Veer, S. N. van der, Keizer, N. F. de, Ravelli, A. C., Tenkink, S. und Jager, K. J. (2010). **Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers.** In: *International Journal of Medical Informatics* 79.5, S. 305–323.
- Vestberg, K., Thulstrup, A. M., Sørensen, H. T., Ottesen, P., Sabroe, S. und Vilstrup, H. (1997). **Data quality of administratively collected hospital discharge data for liver cirrhosis epidemiology.** In: *Journal of Medical Systems* 21.1, S. 11–20.
- Wang, R. Y., Kon, H. B. und Madnick, S. E. (1993). **Data quality requirements analysis and modeling.** In: *Data Engineering, 1993. Proceedings. Ninth International Conference on.* IEEE, S. 670–677.
- Weiss, R. B. (1998). **Systems of protocol review, quality assurance, and data audit.** In: *Cancer Chemotherapy and Pharmacology* 42.1, S88–S92.
- Weiss, R. B., Vogelzang, N. J., Peterson, B. A., Panasci, L. C., Carpenter, J. T., Gavigan, M., Sartell, K., Frei, E. und McIntyre, O. R. (1993). **A successful system of scientific data audits for clinical trials: a report from the Cancer and Leukemia Group B.** In: *JAMA* 270.4, S. 459–464.
- Weitz, M., Laube, G. F., Schmidt, M., Krupka, K., Murer, L., Müller, D., Hoppe, B., Büscher, A., König, J., Pohl, M., Jungraithmayr, T., Thiel, F., Billing, H., Grenda, R., Rubik, J., Kaabak, M. M., Yalcinkaya, F., Topaloglu, R., Webb, N., Dello Strologo, L., Pape, L., Nadalin, S. und Tönshoff, B. (2018). **Outcome of renal transplantation in small infants: a match-controlled analysis.** In: *Pediatric Nephrology*.

- Whitney, C. W., Lind, B. K. und Wahl, P. W. (1998). **Quality assurance and quality control in longitudinal studies.** In: *Epidemiologic Reviews* 20.1, S. 71–80.
- Winter, M., Herrmann, C. und Helfert, M. (2003). **Datenqualitätsmanagement für Data-Warehouse-Systeme—Technische und organisatorische Realisierung am Beispiel der Credit Suisse.** In: *Data warehouse Management.* Springer, S. 221–240.
- Zink, A. und Albrecht, K. (2016). **Wie häufig sind muskuloskeletale Erkrankungen in Deutschland?** In: *Zeitschrift für Rheumatologie* 75.4, S. 346–353.

Eigene Veröffentlichungen

ZEITSCHRIFTENAUFsätze

Höcker, B., Zencke, S., Pape, L., Krupka, K., **Köster, L.**, Fichtner, A., Dello Strologo, L., Guzzo, I., Topaloglu, R., Kranz, B., König, J., Bald, M., Webb, N. J. A., Noyan, A., Dursun, H., Marks, S., Ozcahar, Z. B., Thiel, F., Billing, H., Pohl, M., Fehrenbach, H., Schnitzler, P., Bruckner, T., Ahlenstiel-Grunow, T. und Tönshoff, B. (2016a). **Impact of Everolimus and Low-Dose Cyclosporin on Cytomegalovirus Replication and Disease in Pediatric Renal Transplantation.** In: *American Journal of Transplantation* 16.3, S. 921–929

Tönshoff, B., Krupka, K., **Köster, L.**, Fichtner, A., Pape, L. und Höcker, B. (2015). **CERTAIN Registry.** In: *Der Nephrologe* 10.6, S. 480–487

MEDLINE-GEFÜHRTES ABSTRAKT

Köster, L., Krupka, K., Höcker, B., Rahmel, A., Samuel, U., Zanen, W., Opelz, G., Süsal, C., Döhler, B., Plotnicki, L. et al. (2015). **Integrating data from multiple sources for data completeness in a web-based registry for pediatric renal transplantation—the CERTAIN Registry.** In: *Studies in Health Technology and Informatics* 216, S. 1049–1049

Teilergebnisse der vorliegenden Arbeit wurden im folgenden Konferenzbeitrag (Abstrakt) publiziert:

Köster, L., Geisemeyer, N., Niewerth, M., Minden, K., Krupka, K., Höcker, B., Tönshoff, B. und Knaup, P. (2016). **Analyse der Datenqualität der papierbasierten Erhebung für ein bundesweites Register in der Kinderrheumatologie.** In: *HEC 2016: Health – Exploring Complexity. Joint Conference of GMDS, DGEpi, IEA-EEF, EFMI.* German Medical Science GMS Publishing House, Düsseldorf

Diese Publikation basiert auf den Ergebnissen des Kapitels 4 Überprüfung des Ansatzes der automatischen Validierung. Mein Eigenanteil an der Publikation besteht aus der Durchführung der Auswertung und des Schreibens des Manuskriptes.

Anhang

A Zusammenfassung QI-Analyse der CERTAIN-Register

In diesem Kapitel wird die in Kapitel 6 dargelegte Analyse der einzelnen QI der TMF-Leitlinie zusammengefasst. Nicht anwendbare QI wurden aus Gründen der ganzheitlichen Übersicht beibehalten.

Registerplanung

TMF-ID	Name	Anwendbarkeit	Implementierbarkeit	Aussagekraft	Prorität CERTAIN/-LI	Prorität KRhOKo
1029	Dubletten im Datenbestand	+ +	+	+ +	1	1
1036	Synonyme	o	+ +	+ +	1	1
1037	Homonyme	- -	+ +	+ +	1	1
1020	Werte aus Standards	o	o	+	2	2
1021	Unerlaubte Werte bei qualitativen Datenelementen	- -	-	+	2	2
1022	Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings	- -	-	+	2	2
1023	Unerlaubte Werte zur Kodierung von fehlenden Modulen	- -	o	+	2	2
1024	Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen					

Automatische Validierung

TMF-ID	Name	Anwendbarkeit	Implementierbarkeit	Aussagekraft	Prorität CERTAIN/-LI	Prorität KRhOKo
1004	Sichere Widersprüche/Fehler	++	+	++	1	1
1014	Fehlende Werte bei mandatorischen Datenelementen	++	-	++	1	1
1026	Beobachtungseinheiten mit unbekanntem Primärtumor ⁵	-	++	++	1	1
1001	Übereinstimmung mit Vorwert	o	-	+	2	2
1003	Widerspruchsfreiheit	++	+	+	2	2
1005	Mögliche Widersprüche/Warnungen	++	+	+	2	2
1012	Fehlende Module	++	++	+	2	2
1015	Fehlende Werte bei optionalen Datenelementen	++	-	o	2	2
1018	Ausreißer bei stetigen Datenelementen	++	o	+	2	2
1027	Nachweis bekannter Korrelationen	-	o	+	2	2
1013	Fehlende Werte bei Datenelementen	++	-	o	3	3
1008	Untersuchungen am Wochenende	-	--	--	Keine Anwendung ⁶	Keine Anwendung ⁶

⁵Anwendung als „Beobachtungseinheiten mit unbekannter Primärerkrankung“

⁶Auf eine Anwendung für die drei CERTAIN-Register wird aufgrund der geringen Aussagekraft und des Implementierungsaufwandes verzichtet.

Report

TMF-ID	Name	Anwendbarkeit	Implementierbarkeit	Aussagekraft	Prorität CERTAIN/-LI	Prorität KRhOKo
1028	Aktualität der gespeicherten Daten	++	++	++	1	1
1030	Rekrutierungsrate	++	o	++	1	1
1041	Datenquellen pro Beobachtungseinheit	-	++	++	1	Keine Anwendung
1047	Übereinstimmung mit Verfahrensregeln	+	-	++	1	1
1035	Endlosüberlebende	+	++	o	2	Keine Anwendung
1006	Wertevertelung	++	+	+	2	3
1052	Wertevertelung von Parametern zwischen Zentren	++	+	+	2	3
1016	Datenelemente mit Wert unbekannt o.ä.	+	-	+	2	2
1025	Datenelemente mit unspezifischen Werten	o	--	+	2	2
1034	Vorzeitig aus dem Register ausscheidende Beobachtungseinheiten	++	o	+	2	2
1040	Zurückgewiesene Meldungen ⁷	++	++	+	2	2
1042	Beobachtungseinheiten mit Follow-up	++	++	+	2	2
1007	Bevorzugung bestimmter Endziffern	+	++	o	3	3
1009	Wertevertelung der durch Untersucher erfassten Parameter					

⁷Anwendung als „Zurückgewiesene Datensätze“

1010	Verteilung der durch Geräte erfassten Parameter					
1011	Verteilung von Befunden					
1019	Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten					
1050	Umfang der Metadaten bei Untersuchungen					
1051	DCO-Rate (Death Certificate Only)					
1031	Verweigerungsrate von Untersuchungen					
1032	Verweigerungsrate von Modulen					
1033	Verweigerungsrate von einzelnen Datenelementen					
1038	Einzelmeldungen pro Beobachtungseinheit					
1039	Solitäre Meldungen von Pathologen					
1048	Repräsentativität der Registerdaten					
1017	Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten	++	-	-	Keine Anwendung ⁸	Keine Anwendung ⁸

⁸Auf eine Anwendung für die drei CERTAIN-Register wird aufgrund der geringen Aussagekraft und des Implementierungsaufwandes verzichtet.

Datenverifizierung

Aufgrund der begrenzten Ressourcen der Register sowie der Anzahl der teilnehmenden Einrichtungen und deren räumlicher Verteilung wird generell auf eine Umsetzung dieser QI verzichtet.

TMF-ID	Name	Anwendbarkeit	Implementierbarkeit	Aussagekraft	Prorität CERTAIN/-LI	Prorität KRhOKo
1043	Genauigkeit der Registerdaten	-	--	++	Keine Anwendung	Keine Anwendung
1044	Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente	++	--	++	Keine Anwendung	Keine Anwendung
1045	Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten	++	--	+	Keine Anwendung	Keine Anwendung
1046	Vollständigkeit der Registerdaten	-	--	+	Keine Anwendung	Keine Anwendung
1002	Konkordanz	+	--	o	Keine Anwendung	Keine Anwendung

B QI-Empfehlung für elektronische Register

Registerplanung

TMF-ID	Name	Anwendbarkeit	Implementierbarkeit	Aussagekraft	Vorgeschlagene Priorität	Priorität
1029	Dubletten im Datenbestand				1	
1036	Synonyme				1	
1037	Homonyme				1	
1020	Werte aus Standards				2	
1021	Unerlaubte Werte bei qualitativen Datenelementen				2	
1022	Unerlaubte Werte bei qualitativen Datenelementen zur Kodierung von Missings				2	
1023	Unerlaubte Werte zur Kodierung von fehlenden Modulen				2	
1024	Unerlaubte Werte bei quantitativen Datenelementen zur Kodierung von Unter- oder Überschreitungen von Messbarkeitsgrenzen				2	

Automatische Validierung

TMF-ID	Name	Anwendbarkeit	Implementierbarkeit	Aussagekraft	Vorgeschlagene Priorität	Priorität
1004	Sichere Widersprüche/Fehler				1	
1014	Fehlende Werte bei mandatorischen Datenelementen				1	
1026	Beobachtungseinheiten mit unbekanntem Primärtumor ⁹				1	
1001	Übereinstimmung mit Vorwert				2	
1003	Widerspruchsfreiheit				2	
1005	Mögliche Widersprüche/Warnungen				2	
1012	Fehlende Module				2	
1015	Fehlende Werte bei optionalen Datenelementen				2	
1018	Ausreißer bei stetigen Datenelementen				2	
1027	Nachweis bekannter Korrelationen				2	
1013	Fehlende Werte bei Datenelementen				3	
1008	Untersuchungen am Wochenende				Nicht anwenden ¹⁰	

⁹Sollte es sich um kein Tumorregister handeln, kann die Anwendung als „Beobachtungseinheiten mit unbekannter Primärerkrankung“ erfolgen

¹⁰Eine Anwendung des QI wird aufgrund der geringen Aussagekraft und des Implementierungsaufwandes nicht empfohlen.

Report

TMF-ID	Name	Anwendbarkeit	Implementierbarkeit	Aussagekraft	Vorgeschlagene Priorität	Priorität
1028	Aktualität der gespeicherten Daten				1	
1030	Rekrutierungsrate				1	
1041	Datenquellen pro Beobachtungseinheit				1	
1047	Übereinstimmung mit Verfahrensregeln				1	
1048	Repräsentativität der Registerdaten ¹¹				1	
1035	Endlosüberlebende				2	
1006	Werte Verteilung				2	
1052	Werte Verteilung von Parametern zwischen Zentren				2	
1016	Datenelemente mit Wert unbekannt o.ä.				2	
1025	Datenelemente mit unspezifischen Werten				2	
1034	Vorzeitig aus dem Register ausscheidende Beobachtungseinheiten				2	
1039	Solitäre Meldungen von Pathologen ¹¹				2	
1040	Zurückgewiesene Meldungen ¹²				2	
1042	Beobachtungseinheiten mit Follow-up				2	
1007	Bevorzugung bestimmter Endziffern				3	

¹¹Anwendung vorhabensspezifisch

¹²Sollte es sich um kein Tumorregister handeln, kann die Anwendung als „Zurückgewiesene Datensätze“ erfolgen

1009	Werte Verteilung der durch Untersucher erfassten Parameter				3	
1010	Werte Verteilung der durch Geräte erfassten Parameter				3	
1011	Werte Verteilung von Befunden				3	
1019	Werte, die die Messbarkeitsgrenzen von Verfahren unter- oder überschreiten				3	
1050	Umfang der Metadaten bei Untersuchungen				3	
1051	DCO-Rate ¹³ (Death Certificate Only)				3	
1031	Verweigerungsrate von Untersuchungen				3	
1032	Verweigerungsrate von Modulen				3	
1033	Verweigerungsrate von einzelnen Datenelementen				3	
1038	Einzelmeldungen pro Beobachtungseinheit				3	
1017	Datenelemente mit bestehenden Einträgen bei allen Beobachtungseinheiten				Nicht anwenden ¹⁴	

¹³Anwendung vorhabensspezifisch

¹⁴Eine Anwendung des QI wird aufgrund der geringen Aussagekraft und des Implementierungsaufwandes nicht empfohlen.

Datenverifizierung

TMF-ID	Name	Anwendbarkeit	Implementierbarkeit	Aussagekraft	Vorgeschlagene Priorität	Priorität
1043	Genauigkeit der Registerdaten				1	
1044	Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Datenelemente				1	
1045	Übereinstimmung der Registerdaten mit den Originaldaten bezogen auf Beobachtungseinheiten				2	
1046	Vollständigkeit der Registerdaten				2	
1002	Konkordanz				3	

Lebenslauf

PERSONALIEN

Name: Lennart Köster
 Geburtsjahr: 1986
 Geburtsort: Heidelberg
 Staatsangehörigkeit: Deutsch
 Familienstand: Ledig

SCHULISCHER WERDEGANG

1992 - 1996 Landhaus-Schule Heidelberg
 1996 - 2005 Helmholtz-Gymnasium Heidelberg
 24.06.2005 Allgemeine Hochschulreife

UNIVERSITÄRER WERDEGANG

WS 2005 Beginn des Studiums der Medizinischen Informatik an der Universität Heidelberg und Fachhochschule Heilbronn
 14.02.2011 Diplomprüfung im Studiengang Medizinische Informatik

BERUFLICHER WERDEGANG

10/2011 - 12/2013 Wissenschaftlicher Mitarbeiter
 Horst-Görtz Institut für IT-Sicherheit
 Ruhr-Universität Bochum
 02/2014 - 09/2017 Wissenschaftlicher Mitarbeiter
 Institut für Medizinische Biometrie und Informatik
 Sektion Medizinische Informatik
 Universitätsklinikum Heidelberg
 10/2017 - 05/2018 Wissenschaftlicher Mitarbeiter
 Stiftung Deutsches Rheuma-Forschungszentrum Berlin
 seit 05/2018 Informatiker
 Rheumatologische Fortbildungsakademie GmbH

Danksagung

Ich danke Petra Knaup für die Betreuung dieser Arbeit sowie die immer positiven und konstruktiven Diskussionen zur Arbeit.

Zudem danke ich den Mitarbeitern der Register CERTAIN, CERTAIN-LI und KRhOKo für ihre Mithilfe bei der Betrachtung der QI der TMF-Leitlinie und die Zusammenarbeit im Rahmen dieser Dissertation.

Eidesstattliche Versicherung

1. Bei der eingereichten Dissertation zu dem Thema
MECHANISMEN ZUR AUTOMATISIERTEN QUALITÄTSSICHERUNG UND VALIDIERUNG
DER DATEN IN MEDIZINISCHEN REGISTERN
handelt es sich um meine eigenständig erbrachte Leistung.
2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.
3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt
4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.
5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Ort und Datum

Unterschrift