



Sektorenübergreifende Qualitätssicherung im Gesundheitswesen nach §137a SGB V

Konzept zur Sicherung der Dokumentationsqualität

Vor dem Hintergrund des Datenvalidierungsverfahrens in der
sektorenübergreifenden Qualitätssicherung

Abschlussbericht

Stand: 26. Oktober 2011

Impressum

Herausgeber:

AQUA – Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH

Thema:

Konzept zur Sicherung der Dokumentationsqualität – Datenvalidierungskonzept

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

22. Dezember 2009

Datum der Abgabe:

13. April 2011

(Redaktionelle Überarbeitung am 26. Oktober 2011)

Signatur:

11-SQG-006

Hinweis:

Aus Gründen der leichteren Lesbarkeit wird im Folgenden auf eine geschlechtsspezifische Differenzierung verzichtet. Entsprechende Begriffe gelten im Sinne der Gleichbehandlung für beide Geschlechter.

Anschrift des Herausgebers:

AQUA – Institut für angewandte Qualitätsförderung und
Forschung im Gesundheitswesen GmbH
Maschmühlenweg 8-10 · 37073 Göttingen

Telefon: (+49) 0551 - 789 52 -0

Telefax: (+49) 0551 - 789 52-10

office@aquainstitut.de

www.aquainstitut.de

Inhaltsverzeichnis

Tabellenverzeichnis	5
Abbildungsverzeichnis.....	6
Abkürzungsverzeichnis.....	7
Glossar.....	8
Hintergrund.....	10
Zusammenfassung	11
1. Einleitung	12
1.1. Grundlegende Zielrichtung des Konzeptes	13
1.2. Entwicklung statistischer Auffälligkeitskriterien.....	14
1.3. Umgang mit Leistungserbringern, bei denen statistische Auffälligkeiten ermittelt wurden	14
1.4. Ermittlung der in eine Stichprobe einzubeziehenden Leistungserbringer.....	15
1.5. Durchführung einer Stichprobenprüfung beim Leistungserbringer vor Ort.....	15
1.6. Evaluation des Verfahrens.....	16
2. Prüfkriterien der Dokumentationsqualität	17
2.1. Merkmale zur Auswahl von Prüfkriterien	17
2.1.1. Einsatz im nationalen Kontext	17
2.1.2. Einsatz im internationalen Kontext.....	17
2.1.3. Lehrmeinungen.....	18
2.1.4. Ressourcenverbrauch und Invasivität	18
2.1.5. Diskriminationsfähigkeit	18
2.2. Hintergrund der Prüfkriterienauswahl	19
2.3. Ausgewählte Prüfkriterien.....	19
2.3.1. Vollzähligkeit.....	20
2.3.2. Vollständigkeit	20
2.3.3. Konkordanz.....	21
2.3.4. Erlaubte Werte	22
2.3.5. Widerspruchsfreiheit	22
2.3.6. Werteverteilung	23
2.3.7. Bekannte Korrelationen	24
2.3.8. Fehlende Werte in optionalen Datenfeldern	24
2.3.9. Dubletten.....	25
2.4. Bewertung der Prüfkriterien	26
3. Datenprüfungen im Prozess der Datenerhebung	27
4. Datenvalidierungsverfahren.....	33

4.1. Auffälligkeitsprüfung.....	33
4.2. Stichprobenprüfung.....	34
5. Praktische Umsetzung des Datenvalidierungsverfahrens.....	36
5.1. Vorbereitende Maßnahmen	37
5.2. Durchführung	39
5.3. Umgang mit Leistungserbringern mit statistischen Auffälligkeiten.....	39
5.4. Evaluationen	41
5.4.1. Evaluation der QS-Verfahren	41
5.4.2. Evaluation des Datenvalidierungsverfahrens.....	41
5.4.3. Berichtswesen	43
6. Empfehlungen und Ausblick.....	44
Literaturverzeichnis	45

Tabellenverzeichnis

Tabelle 1: Umgang mit Auffälligen Daten in Abhängigkeit vom Zeitpunkt der Datenprüfung.....	15
Tabelle 2: Eigenschaften des Kriteriums Vollzähligkeit.....	20
Tabelle 3: Eigenschaften des Kriteriums Vollständigkeit.....	21
Tabelle 4: Eigenschaften des Kriteriums Konkordanz.....	22
Tabelle 5: Eigenschaften des Kriteriums Erlaubte Werte.....	22
Tabelle 6: Eigenschaften des Kriteriums Widerspruchsfreiheit.....	23
Tabelle 7: Eigenschaften des Kriteriums Werteverteilung.....	24
Tabelle 8: Eigenschaften des Kriteriums Bekannte Korrelationen.....	24
Tabelle 9: Eigenschaften des Kriteriums Fehlende Werte in optionalen Datenfeldern.....	25
Tabelle 10: Eigenschaften des Kriteriums Anteil der Dubletten.....	25
Tabelle 11: Zeitplan vorbereitende Maßnahmen.....	38
Tabelle 12: Zeitplan Evaluation und Bericht.....	43

Abbildungsverzeichnis

Abbildung 1: Datenfluss und Datenprüfungen bei Krankenhäusern	29
Abbildung 2: Datenfluss und Datenprüfungen bei kollektivvertraglich tätigen Arztpraxen.....	31
Abbildung 3: Datenfluss und Datenprüfungen bei selektivvertraglich tätigen Arztpraxen	32
Abbildung 4: Datenvalidierungsverfahren in der sektorenübergreifenden Qualitätssicherung	36

Abkürzungsverzeichnis

Abkürzung	Beschreibung
G-BA	Gemeinsamer Bundesausschuss
KV(en)	Kassenärztliche Vereinigung(en)
KZV(en)	Kassenzahnärztliche Vereinigung(en)
LAG(en)	Landesarbeitsgemeinschaft(en) für sektorenübergreifende Qualitätssicherung
LKG(en)	Landeskrankenhausgesellschaft(en)
LOS	Landesgeschäftsstelle für Qualitätssicherung
PCI	Koronarangiographie und Perkutane Koronarintervention
PDCA	<i>plan do check act</i> („Plane-Tue-Überprüfe-Handle“)
Qesü-RL	Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung
QS	Qualitätssicherung
VSt	Vertrauensstelle
SGB	Sozialgesetzbuch
KHEntgG	Gesetz über die Entgelte für voll- und teilstationäre Krankenhausleistungen

Glossar

Begriff	Beschreibung
Akkuratesse	(engl. <i>accuracy</i>); Dokumentationsqualitätskriterien, die Rückschlüsse auf die Übereinstimmung von Datensätzen mit den Referenzdaten zulassen.
Datensatz	Eine Menge von Datenfeldern, die einem Fall (beispielsweise einem Patienten) zugeordnet wird.
Diskriminationsfähigkeit	Geeignetheit eines Kriteriums, empirische Phänomene qualitativ sinnvoll zu unterscheiden.
Dokumentationsqualität	Maß, inwiefern die Prüfkriterien zur →Akkuratesse, Plausibilität und Datenstruktur/-organisation erfüllt sind.
Einrichtung	Organisatorische Einheit im Gesundheitswesen (beispielsweise Krankenhaus oder Vertragsarztpraxis), in der Leistungserbringer Patienten versorgen.
Follow-up-Datensätze	Alle Datensätze, die im Rahmen einer Behandlungsepisode einem Patienten zuzuordnen sind.
<i>fuzzy search</i>	Suche nach annähernd gleichen →Datensätzen („Unschärfe Suche“).
Klumpenstichprobe	Stichprobe, bei der zunächst übergeordnete Einheiten (z. B. Krankenhäuser oder Arztpraxen) gezogen werden, aus denen in der Folge die Einzeldatensätze stammen. In einer zweiten Stufe wird entweder eine Vollerhebung oder eine nahezu vollständige Erhebung der Elemente (z. B. behandelte Patienten) durchgeführt. Wenn in der zweiten Stufe tatsächlich eine Vollerhebung durchgeführt wird, handelt es sich faktisch um ein einstufiges Stichprobenverfahren. Bei einem zweistufigen Stichprobenverfahren werden auch in der 2. Stufe Stichproben gezogen.
Komplettheit	→Vollzähligkeit und →Vollständigkeit.
Konkordanz	Übereinstimmung der Stichprobendaten mit den Ursprungsdaten, hier der Patientendokumentation.
<i>oversampling</i>	Die (bewusste) Erhöhung der Wahrscheinlichkeit, dass Elemente mit bestimmten Eigenschaften (z. B. Patienten aus rechnerisch auffälligen Einrichtungen) in eine Stichprobe gezogen werden.
Pfadabhängigkeit	Ein stark vergangenheitsmitbestimmter Prozess, der nur kleine inkrementeller Entwicklungen erlaubt.
Reliabilität	Ein Maß für die Zuverlässigkeit eines Indikators, das anzeigt, inwieweit der Indikator stets das gleiche misst. Das „Ausmaß [...], in dem wiederholte Messungen eines Objektes mit einem Messinstrument die gleichen Werte liefern“ (Schnell 2008).
Routinedaten	Daten, die im Zusammenhang mit Patienten im medizinischen Alltag grundsätzlich anfallen – z. B. administrative und/oder Abrechnungsdaten. Für die Qualitätssicherung wäre der Zugriff auf bestimmte Routinedaten der Krankenkassen von Interesse, nämlich Stammdaten (z. B. Mortalität), Abrechnungsdaten zu Behandlungsleistungen (z. B. ICD, OPS) und statistische Angaben. Gründe hierfür sind die Vermeidung zusätzlichen Dokumentationsaufwandes und eine, zumindest in Teilen, bessere Vollständigkeit und Datenqualität als bei separater Erfassung. Allerdings ist die Datenverwendung an die in §284 SGB V genannten Zwecke gebunden. Derzeit besteht für Zwecke der Qualitätssicherung nur Zugriff auf Daten nach §21 Abs. 3a KHEntgG, die ab Mitte eines Folgejahres verfügbar sind.
<i>sampling frame</i>	Struktur und Umfang der Grundgesamtheit der →Datensätze, aus denen eine Stichprobe gezogen wird.
sektorenübergreifend	Gemäß §1 (3) der Qesü-RL sind QS-Verfahren, die nicht nur auf einen spezifischen Sektor bezogen Anwendung finden, sondern sektorale Grenzen in den jeweiligen Anforderungen unbeachtet lassen, sektorenübergreifend.
sektorenüberschreitend	Gemäß §1 (3) der Qesü-RL sind diejenigen →sektorenübergreifenden QS-Verfahren, bei denen mindestens zwei Sektoren am Behandlungsergebnis maßgeblich Anteil haben, sektorenüberschreitend. Sie bilden damit eine Teilmenge der sektorenübergreifenden Verfahren.
sektorgleich	Gemäß §1 (3) der Qesü-RL: QS-Verfahren, bei denen die Erbringung der gleichen medizinischen Leistung in unterschiedlichen Sektoren erfolgt. Auch diese Verfahren bilden damit eine Teilmenge der sektorenübergreifenden Verfahren.

Begriff	Beschreibung
Stellungnahmeverfahren	Gemäß § 17 (2) der Qesü-RL: Hier: Verfahren im Umgang mit auffällig gewordenen Leistungserbringern. Es kennt vier aufeinander aufbauende Handlungsstufen, an denen jeweils das Verfahren beendet werden kann, wenn zufriedenstellende Ergebnisse erreicht worden sind.
Strukturierter Dialog	Gemäß § 17 (2) der Qesü-RL: Ergebnisoffener und systematisch organisierter Dialog mit bestimmten Leistungserbringern.
Validität	Gibt an, ob die erhobenen Daten das abbilden, was sie abbilden sollen. Der Begriff Validität wird unterschiedlich konzipiert. Im vorliegenden Kontext der Prüfung der Dokumentationsqualität wird unterschieden zwischen: <ul style="list-style-type: none">▪ Inhaltsvalidität: diese wird gemessen über die Widerspruchsfreiheit und Plausibilität der Daten.▪ Konstruktvalidität: diese wird gemessen durch Vergleich zu einem externen Standard (Krankenakte, Routinedaten). Solche Kriterien werden insbesondere bei den Prüfungen zur Akkuratessse angewendet.
Vollständigkeit	Grad der nicht-fehlenden mandatorischen Werte innerhalb eines →Datensatzes.
Vollzähligkeit	Anteil vorhandener →Datensätze innerhalb einer definierten Grundgesamtheit.

Hintergrund

In Verbindung mit der Übernahme der Aufgaben als unabhängige Institution gemäß § 137a SGB V hat das AQUA-Institut den Auftrag, ein „Verfahren zur Überprüfung der Dokumentationsqualität (Datenvalidierungsverfahren)“ für die sektorenübergreifende Qualitätssicherung zu entwickeln (vgl. Anlage 1.1 der Aufgabenbeschreibung, Abschnitt 3.6).

Der Gemeinsame Bundesausschuss hat in seiner Sitzung am 17. Dezember 2009 das AQUA-Institut formal beauftragt, ein „Konzept für das Verfahren zur Überprüfung der Dokumentationsqualität“ zu entwickeln.

Die Frist zur Bearbeitung dieses Auftrages war zunächst auf den 31. Juli 2010 festgelegt worden. Da zu diesem Zeitpunkt die Richtlinie über die Maßnahmen der sektorenübergreifenden Qualitätssicherung (Qesü-RL) noch nicht die Nichtbeanstandung durch das Bundesministerium für Gesundheit erfahren hatte, ist die Frist durch den G-BA verschoben worden.

Der Vorbericht ist vom AQUA-Institut am 9. Dezember 2011 fertiggestellt den beteiligten Institutionen nach § 137a SGB V (Abs. 3) zur Stellungnahme zur Verfügung gestellt worden. Auf Antrag wurde das Stellungnahmeverfahren um 4 Wochen auf insgesamt 8 Wochen verlängert, sodass die bis zum 3. Februar 2011 durch die berechtigten Institutionen eingereichten Stellungnahmen berücksichtigt werden konnten.

Das AQUA-Institut hat die Stellungnahmen gewürdigt und in dem zu erstellenden Abschlussbericht berücksichtigt. Die Abgabe des Abschlussberichtes erfolgte am 13. April 2011.

Zusammenfassung

Im vorliegenden Konzept zur Sicherung der Dokumentationsqualität werden systematisch entwickelte und wissenschaftlich fundierte Prüfkriterien der Dokumentationsqualität vorgestellt. In einem umfassenden Ansatz wird die Anwendung dieser Prüfkriterien zunächst im Kontext der Datenflüsse für Erhebungen durch Leistungserbringer in der sektorenübergreifenden Qualitätssicherung dargestellt. Daran anschließend fokussiert das Konzept auf das sogenannte Datenvalidierungsverfahren, das aus einer statistischen Auffälligkeitsüberprüfung und einer tiefergehenden Stichprobenprüfung besteht. In der Auffälligkeitsprüfung werden die Gesamtdaten auf rechnerische Auffälligkeiten geprüft. Kernelemente des Stichprobenverfahrens sind Überprüfungen der Qualität der QS-Dokumentationen anhand von Informationen aus der Patientenakte.

Neben konzeptionellen Grundlagen werden auch Fragen der Umsetzung, zum Umgang mit auffällig gewordenen Leistungserbringern und zur Evaluation des Verfahrens behandelt. Dabei wird auf Vorgaben der Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung (Qesü-RL) zurückgegriffen.

Über das Gesamtkonzept hinaus haben wir organisatorische Fragen bei seiner Durchführung geklärt. Die Zielsetzung dabei war, das Zusammenspiel der verschiedenen Akteure so klar wie zu diesem Zeitpunkt möglich darzulegen. Insgesamt ist davon auszugehen, dass eine sektorenübergreifende Qualitätssicherung zahlreiche neue Informationen generieren wird.

Das Konzept bildet ein Fundament für die weitere Ausgestaltung und Konkretisierung des Datenvalidierungsverfahrens in der sektorenübergreifenden Qualitätssicherung. Aufgrund der an vielen Stellen neuen Herausforderungen, sowohl inhaltlicher Art (wie die Überprüfung sektorenübergreifend zusammengeführter Daten) als auch in der Umsetzung (an der Durchführung beteiligte Institutionen sind zum Teil noch nicht existent), wird vor einer Umsetzung in den Regelbetrieb die Erprobung des Konzeptes vorgeschlagen. Dabei werden auch Detailfragen der Umsetzung geklärt.

1. Einleitung

Die Messung und Darstellung der Versorgungsqualität ist eine zentrale Aufgabe bei der Umsetzung der Qualitätssicherung. Zur Messung sollen Qualitätsindikatoren entwickelt werden, die sowohl Lern- und Verbesserungsprozesse in den Einrichtungen unterstützen als auch Grundlage für Wahlentscheidungen von Versicherten sein können. Damit die entwickelten Qualitätsindikatoren zielführend sind und nicht ihrerseits Fehlsteuerungen verursachen, sollen folgende Prozesse implementiert werden:

- eine wissenschaftlich fundierte Methode für die Entwicklung der Qualitätsindikatoren und
- ein Konzept zur Sicherung der Dokumentationsqualität.

Das hier vorgelegte Konzept beschäftigt sich mit letztgenannter Aufgabe.

Vorgaben der Beauftragung durch den Gemeinsamen Bundesausschuss

Gemäß den Vorgaben des G-BAs in der Auftragsbeschreibung sollte das Verfahren zur Überprüfung der Dokumentationsqualität darauf abzielen, „[...] einer systematischen Falschberichtserstattung und Datenmanipulation vorzubeugen“ (vgl. Anlage 1.1 der Aufgabenbeschreibung des AQUA-Instituts, Abschnitt 3.6).

Der Auftrag beinhaltet die Abfassung eines Konzeptpapiers, in dem insbesondere folgende Aufgaben berücksichtigt sein sollen:

- die Entwicklung statistischer Auffälligkeitskriterien;
- der Umgang mit Leistungserbringern, bei denen statistische Auffälligkeiten ermittelt wurden;
- die Ermittlung der in eine Stichprobe einzubeziehenden Leistungserbringer;
- die Durchführung einer Stichprobenprüfung beim Leistungserbringer vor Ort;
- die Evaluation des Verfahrens.

Das Konzept muss dem Stand der Wissenschaft entsprechen und mit einem angemessenen Verhältnis von Aufwand und Nutzen umsetzbar sein.

Umsetzung der Vorgaben im vorliegenden Konzept

Neben der Aufgabenbeschreibung des AQUA-Instituts sind zwischenzeitlich weitere, für die Umsetzung eines Konzeptes zur Sicherung der Dokumentationsqualität relevante Vorgaben entwickelt worden. Hierzu gehört insbesondere die Richtlinie zur einrichtungs- und sektorenübergreifenden Qualitätssicherung (Qesü-RL), in der ein Grundmodell des Datenflusses sowie Vorgaben zur Bewertung von Auffälligkeiten und zu Qualitätssicherungsmaßnahmen festgelegt sind. Weiterhin relevant für die Erstellung des Konzeptes sind die Ergebnisse der ersten Entwicklungsleistungen des AQUA-Instituts zu den Themen Katarakt, Konisation sowie Koronarangiographie und Perkutane Koronarintervention (PCI).

Bisher befindet sich noch keines der entwickelten sektorenübergreifenden Themen in der Umsetzung als QS-Verfahren. Auch die zur Umsetzung von sektorenübergreifenden QS-Verfahren erforderlichen Strukturen auf Landesebene – Landesarbeitsgemeinschaften (LAGen) und Vertrauensstelle (VSt) – sind zum gegenwärtigen Zeitpunkt noch nicht etabliert. Dort, wo für eine weitere Konkretisierung eines Konzeptes zur Sicherung der Dokumentationsqualität Abstimmungen mit diesen Strukturen und dem Gemeinsamen Bundesausschuss notwendig sind, kann das Konzept derzeit nur Grundlagen für die weitere Ausarbeitung schaffen, jedoch noch keine abschließenden Vorgaben formulieren.

Nachfolgend wird dargelegt, auf welche Weise die Vorgaben der Aufgabenbeschreibung umgesetzt werden.

1.1. Grundlegende Zielrichtung des Konzeptes

In Bezug auf die Zielrichtung des Konzeptes sind drei Aspekte von grundlegender Bedeutung:

- die Fokussierung auf systematische Falschberichterstattung und Datenmanipulation;
- die breite Ausrichtung des Konzeptes;
- die Berücksichtigung möglicher neuer Erhebungsinstrumente für Qualitätsindikatoren, nämlich Patientenbefragungen und Routinedaten.

Fokussierung auf relevante Fallkonstellationen vis-à-vis der breiten Ausrichtung des Konzeptes

In der Aufgabenbeschreibung ist als wesentliches Ziel die Vorbeugung vor systematischer Falschberichterstattung und Datenmanipulation formuliert. Das AQUA-Institut ist zu der Einschätzung gekommen, dass es sinnvoll ist, die Zielsetzung des Verfahrens zur Überprüfung der Dokumentationsqualität (Datenvalidierungsverfahren) breiter anzulegen, als dies der Auftragsbeschreibung zu entnehmen ist. Neben der Prävention vor Datenmanipulationen sollte es auch darauf abzielen, Fehldokumentationen in den QS-Daten zu identifizieren, deren Ursachen aufzudecken und notwendige Optimierungsmaßnahmen einzuleiten; und zwar aus folgenden Gründen:

- Erfahrungen mit dem Datenvalidierungsverfahren in der externen stationären Qualitätssicherung haben gezeigt, dass systematische Falschberichterstattungen nicht die primäre Ursache von fehlerhaften Dokumentationen sind (AQUA 2010b: 32). Vielmehr sind es überwiegend Schwierigkeiten, die auf den komplexen Prozess von Datengenerierung und -erfassung zurückzuführen sind (z. B. Missverständnis bei der Definition von Datenfeldern, Softwareprobleme).
- Zudem sollte das Datenvalidierungsverfahren so konzipiert sein, dass quantitative Aussagen zur Güte der Dokumentationsqualität getroffen werden können, da dadurch die Beurteilung der Validität von Qualitätsindikatoren verbessert wird.
- Die erweiterte Zielsetzung des Verfahrens entspricht auch den Grundgedanken der Durchführung von Qualitätssicherungsmaßnahmen nach § 17 der Qesü-RL.

Berücksichtigung möglicher neuer Erhebungsinstrumente für Qualitätsindikatoren

In den Abschlussberichten zu den ersten Entwicklungsleitungen des AQUA-Instituts zu den Themen Katarakt, Konisation und PCI sind neben den in der Qesü-RL bereits geregelten Datenerhebungen bei Leistungserbringern auch Patientenbefragungen und Routinedaten der Krankenkassen als Quellen für Teile der entwickelten Qualitätsindikatoren genannt. Die Konkretisierung der weiteren Umsetzung dieser Erhebungsinstrumente für die Bildung von Qualitätsindikatoren wird derzeit mit dem G-BA abgestimmt. Nach Klärung grundlegender Sachverhalte tritt hier auch die Frage auf, ob und in welcher Form ein Konzept zur Sicherung der Dokumentationsqualität für diese Erhebungsinstrumente erstellt werden soll. Das vorliegende Konzept fokussiert jedoch gemäß Aufgabenbeschreibung auf die Sicherung der Dokumentationsqualität der Datenerhebungen von Leistungserbringern zur Abbildung entwickelter Qualitätsindikatoren.

Anders verhält es sich mit der Einbeziehung von Routinedaten als mögliche Referenzquelle zur Überprüfung der Dokumentationen. Hierzu werden im vorliegenden Konzept Vorstellungen dargelegt, an welchen Stellen dies im Kontext einer sektorenübergreifenden Qualitätssicherung sinnvoll erscheint.

Hinsichtlich der Einbindung von Selbstzahlern (insbesondere Privatversicherte und Beihilfeberechtigte) in die sektorenübergreifende Qualitätssicherung insgesamt, bestehen darüber hinaus noch grundlegende Fragen. Im ambulanten Sektor bestehen für diese Patienten andere Abrechnungsmechanismen. Gerade für die Frage der Vollständigkeitsprüfung wirft dies Fragen auf, für die bisher noch keine Lösungen vorliegen.

1.2. Entwicklung statistischer Auffälligkeitskriterien

Die in Kapitel 2 dargestellte Entwicklung statistischer Auffälligkeitskriterien ist Kernbestandteil des vorliegenden Konzeptes. Es werden zunächst Merkmale definiert, anhand derer Prüfkriterien/Prüfwerkzeuge bewertet und ausgewählt werden können. Darauf basierend wird ein wissenschaftlich fundiertes System von Prüfkriterien für die Dokumentationsqualität entwickelt, welches auf die zu validierenden Daten angewandt werden soll. Auswahlkriterien sind z. B. Bewährtheit und das Verhältnis von Aufwand und Nutzen. Besondere Beachtung findet die Übertragung einzelner Prüfkriterien auf den sektorenübergreifenden Kontext. Zu beachten ist dabei, dass das Konzept zur Überprüfung der Dokumentationsqualität nicht die Gesamtheit der nach § 16 Qesü-RL erforderlichen Maßnahmen umfasst, da die Dokumentationsqualität nur einen Aspekt der Validität darstellt.

Die Umsetzung eines Konzeptes zur Sicherung der Dokumentationsqualität in einem sektorenübergreifenden Kontext ist nach unserem Kenntnisstand ein Novum. Mit dieser sektorenübergreifenden Ausrichtung stellen sich neue Fragen für die Dokumentationsqualität, beispielsweise:

- Lassen sich die richtigen Daten zusammenführen?
- Sind alle Follow-up-Datensätze vorhanden?
- Fließen bei mehreren Datensätzen die richtigen Datensätze in die Qualitätsmessung ein?

Ein Ziel des Konzeptes ist es, vorhandene, wissenschaftlich fundierte Kriterien zur Überprüfung der Dokumentationsqualität auf diesen Kontext zu übertragen.

Die Berücksichtigung des geforderten angemessenen Verhältnisses von Aufwand und Nutzen der Maßnahmen zur Sicherung der Dokumentationsqualität wird von folgenden Gedanken geleitet:

- Je früher ein Dokumentationsfehler erkannt oder vermieden werden kann, desto weniger aufwendig ist die Korrektur.
- Prüfkriterien, die auf eine Fehlererkennung im Nachhinein zielen, sollten möglichst wenig invasiv sein und auf vorhandenen Daten basieren.

Eine frühe Entdeckung von Dokumentationsfehlern setzt die Implementierung von Datenprüfroutinen möglichst bereits bei der Eingabe und im Zuge des Datenaustausches voraus. Vor dem Hintergrund der ausgewählten Kriterien zur Datenprüfung (vgl. Kapitel 2.3) und des in der Qesü-RL beschriebenen Datenflussmodells beschreibt Kapitel 3, welche Prüfungen auf welcher Ebene des Datenflusses umgesetzt werden können.

1.3. Umgang mit Leistungserbringern, bei denen statistische Auffälligkeiten ermittelt wurden

Der Begriff „statistische Auffälligkeiten“ verweist auf Prüfungen, die nicht auf Grundlage eines einzelnen Datensatzes, sondern eines Datenpools stattfinden. Im Kontext der sektorenübergreifenden Qualitätssicherung ist dies der sogenannte Bundesdatenpool, in dem die Daten aller Datenannahmestellen sektorenübergreifend zusammengeführt werden. Die hier durchgeführten Prüfungen werden nachfolgend unter dem Begriff Datenvalidierung zusammengefasst (vgl. Kapitel 4).

In Kapitel 5.3 wird erörtert, wie mit Leistungserbringern umgegangen werden soll, bei denen im Rahmen des Datenvalidierungsverfahrens (Basisprüfung oder Stichprobenprüfung) (statistische) Auffälligkeiten ermittelt wurden. Den Ausgangspunkt für das dort dargelegte Vorgehen bilden die Regelungen der Qesü-RL, auf die an geeigneter Stelle verwiesen wird.

Im Sinne des hier verfolgten, breiteren Ansatzes zur Sicherung der Dokumentationsqualität existieren unterschiedliche mögliche Zeitpunkte der Datenprüfung. In Tabelle 1 sind diese Zeitpunkte, sowie die Konsequenzen von Auffälligkeiten, zusammenfassend dargestellt. Zusätzlich ist ein Verweis auf die Kapitel des vorliegenden Konzeptes enthalten, in denen weitere Ausführungen dazu enthalten sind.

Tabelle 1: Umgang mit Auffälligen Daten in Abhängigkeit vom Zeitpunkt der Datenprüfung

Zeitpunkt der Datenprüfung	Konsequenzen von Auffälligkeiten
Beim Leistungserbringer, während der Dateneingabe, <i>vgl. Kapitel 3.</i>	Abhängig von Art der Auffälligkeit: z. B. Warnhinweis oder kein Dokumentationsabschluss möglich
Im Zuge des Datenaustauschprozesses (z. B. Export beim Leistungserbringer, Datenannahmestelle, Aufnahme von Daten in den Bundesdatenpool), <i>vgl. Kapitel 3.</i>	Abhängig von Art der Auffälligkeit: z. B. Protokoll, Aufforderung zur Korrektur oder Verweigerung der Datenannahme
Datenvalidierungsverfahren, also Umsetzung von Auffälligkeitsprüfungen bzw. Stichprobenprüfungen auf Grundlage der sektorenübergreifend gesammelten Daten eines definierten Zeitraums (Bundesdatenpool), <i>vgl. Kapitel 4.</i>	Nach Maßgabe der in der Qesü-RL festgelegten Regelungen, <i>vgl. Kapitel 5.3.</i> Weitere Einzelheiten sind mit den an der Durchführung beteiligten Institutionen abzustimmen

1.4. Ermittlung der in eine Stichprobe einzubeziehenden Leistungserbringer

Zum Design der Stichprobe wird in Kapitel 4.2 ein Grundkonzept dargestellt. Vor dem Hintergrund der sektorenübergreifenden Qualitätssicherung besteht die Herausforderung hier darin, eine möglichst aussagekräftige Stichprobe mit möglichst geringer Fallzahl zu konstruieren. Zunächst ist dabei die Frage des *sampling frames* zu diskutieren: Was sind eigentlich die Grundelemente der Stichprobe? Sind dies Einrichtungen, Patienten oder Eingriffe? Hier gilt es vor allem sinnvolle Nutzen-/Aufwand-Abwägungen vorzunehmen. Stichproben von Patienten wären zwar aus methodologischer Sicht am vorteilhaftesten, würden aber gerade bei sektorenübergreifenden Verfahren sehr aufwendig werden: Zum einen müssten möglicherweise Datensätze sektorenübergreifend geprüft werden, sodass zahlreiche Einrichtungen betroffen wären, zum anderen wären insbesondere bei den ambulanten Leistungserbringern viele Besuche in der Fläche notwendig, was zu hohen Reiseauslagen und hohem logistischen Aufwand – auch für die betroffenen Leistungserbringer – führen würde.

Wir empfehlen die Ziehung einer einrichtungsbezogenen Klumpenstichprobe. Innerhalb der gezogenen Einrichtungen, den Klumpen, sollen dann Zufallsstichproben gezogen werden. Ein solches Vorgehen ist wesentlich kostengünstiger als eine einstufige Zufallsstichprobe. Bei der Konstruktion der Stichprobe empfehlen wir zudem das *oversampling* rechnerisch auffälliger Einrichtungen, wobei hier in der Zukunft auch eine longitudinale Komponente denkbar wäre, bei der im Vor- oder Vorvorjahr rechnerisch auffällig gewordene Einrichtungen eine überdurchschnittliche Wahrscheinlichkeit, in die Stichprobe gezogen zu werden, zugewiesen bekämen. Derzeit soll die Stichprobe zu möglichst zwei Fünfteln aus Patienten aus rechnerisch auffällig gewordenen Einrichtungen bestehen. Außerdem soll mandatiert werden, aus jedem Bundesland mindestens eine Einrichtung pro QS-Verfahren in der Stichprobe zu untersuchen.

1.5. Durchführung einer Stichprobenprüfung beim Leistungserbringer vor Ort

Aufgrund der Fülle der zu beteiligen Akteure bei der sektorenübergreifenden Stichprobenerhebung besteht besonderer Koordinationsbedarf. Daher liegt ein Fokus auf der Identifizierung der relevanten Akteure. Dabei ist uns selbstverständlich bewusst, dass für ein transparentes Verfahren auch ein klares systematisches Vorgehen, das eindeutig dokumentiert ist, notwendig ist. Da an dieser Stelle aber noch Klärungsbedarf hinsichtlich der konkreten Prozesse besteht, ist diese Problematik hier lediglich skizziert. Ein vollständiges Handbuch, was zur transparenten und einheitlichen Durchführung des Verfahrens benötigt wird, kann an dieser Stelle noch nicht vorgelegt werden.

1.6. Evaluation des Verfahrens

Die Grundideen zur Evaluation des Verfahrens werden in Kapitel 5.4 vorgestellt. Dabei wird zwischen einer Evaluation des Verfahrens selbst und der Evaluation der QS-Verfahren im Zuge der Datenvalidierung unterschieden. Zentrale Frage des letztgenannten Punktes ist, wie sich aus den empirisch erhobenen Verteilungen der rechnerischen Auffälligkeiten Hinweise für Verbesserungen in den QS-Verfahren ergeben könnten. Darüber hinaus soll auch das Datenvalidierungsverfahren selbst einer Evaluation unterworfen werden, um in künftigen Zyklen dieses innovative und mutmaßlich noch fehleranfällige Verfahren zu optimieren. Ein elaboriertes Konzept zur Evaluation wird ebenfalls im Zuge der konkreten Planung der ersten sektorenübergreifenden Datenvalidierungsverfahren entwickelt werden.

2. Prüfkriterien der Dokumentationsqualität

Um eine aussagekräftige und transparente Prüfung der Daten- bzw. Dokumentationsqualität¹ durchführen zu können, müssen in einem ersten Schritt Merkmale identifiziert werden, mit denen unterschiedliche Prüfkriterien untereinander verglichen und bewertet werden können. In einem zweiten Schritt haben wir bekannte Prüfkriterien mithilfe dieser Merkmale analysiert und ausgewählt.

2.1. Merkmale zur Auswahl von Prüfkriterien

Für die Festlegung von Prüfkriterien haben wir folgende Eigenschaften geprüft:

- bisheriger Einsatz im nationalen Kontext;
- Einsatz im internationalen Kontext;
- Meinungen in der methodologischen Fachliteratur;
- geringer Ressourcenverbrauch;
- geringe Invasivität (Arbeit mit Daten, die möglichst nicht zusätzlich erhoben werden müssen, z. B. Zeitreihendaten);
- möglichst hohe Diskriminationsfähigkeit (Wie deutlich kann „gute“ von „schlechter“ Dokumentationsqualität unterschieden werden?);
- möglichst hohe Fehlerursachenerkennung:
 - Möglichkeit, systematische Fehler von Zufallsfehlern zu unterscheiden;
 - Möglichkeit, Rückschlüsse auf mögliche Fehlerursachen (Datenmanipulation, Softwarefehler, etc.) zu ziehen.

Allgemein ist zu beachten, dass die Erfüllung dieser Anforderungen zwar die Voraussetzung für eine gute Reliabilität und Validität der Daten ist, diese jedoch nicht impliziert.

Im Einzelnen wird die Auswahl der Merkmale, anhand derer Prüfkriterien bestimmt werden sollen, wie folgt begründet:

2.1.1. Einsatz im nationalen Kontext

Das AQUA-Institut hält den Rückgriff auf und den Einsatz von Prüfkriterien, die schon in praktischer Anwendung waren, für sinnvoll und angemessen. Obwohl eine „Pfadabhängigkeit“ – in etwa: das Festhalten an althergebrachten Prozessen ohne einen Funktionsvergleich dieser Prozesse mit alternativen Handlungssträngen – häufig innovationsbremsend wirkt (David 1997; Goldstone 1998), ist es sinnvoll, zumindest einen Teil der etablierten Prüfkriterien wieder zu verwenden. Dies hat drei Gründe: Zunächst bedeutet es für Anwender einen geringeren Lernaufwand, da sie mit den entsprechenden Kriterien bereits vertraut sind. Des Weiteren können so Ergebnisse der Prüfung über längere Zeiträume miteinander verglichen werden, was bei einem Einsatz neuerer Kriterien nicht möglich wäre. Drittens deutet der frühere Einsatz auch auf ein bestimmtes Maß qualitätssichernder Eigenschaften und einen gewissen Grad der Bewährtheit hin, weil man davon ausgehen kann, dass die Kriterien seinerzeit nicht willkürlich, sondern anhand rational begründbarer Merkmale selektiert worden sind.

2.1.2. Einsatz im internationalen Kontext

Der Vergleich mit der Anwendung von Prüfkriterien im internationalen Kontext ist ein gangbarer Weg, der Pfadabhängigkeit im nationalen Kontext entgegenzuwirken (Djelic et al. 2007: 180f). Auch internationale Studien haben sich, oft unabhängig von hiesigen nationalen Standards, bewährt. Internationale Vergleichsstudien kön-

¹ Die Begriffe *Datenqualität* und *Dokumentationsqualität* werden in der Fachliteratur nicht immer deutlich voneinander abgegrenzt. So sprechen Nonnemacher et al. (2007) von „Datenqualität“, um die in Zaiß et al. (2002: 50) unter „Dokumentationsqualität“ dargestellten Prüfkriterien zu erläutern. Wir benutzen hier den Begriff „Dokumentationsqualität“, zum einen, um Brüche mit der tradierten Terminologie der BQS zu vermeiden, zum anderen, weil der Begriff „Datenqualität“ bisweilen auch weiter gefasst wird und auch die hier nicht geprüfte *Reliabilität* und *Validität* eines Indikators mit einbezieht (Leidl 2003: 470).

nen so als Korrektiv wirken, indem sie Hinweise auf andere bewährte Kriterien geben, die in der bisherigen nationalen Qualitätssicherung nicht eingesetzt worden sind. Umgekehrt können sie aber auch nationale methodische Vorgehensweisen bestärken, ohne die Bürde der nationalen Pfadabhängigkeit zu tragen.

Um die Vergleichbarkeit zu gewährleisten, wurde bei der Auswahl entsprechender Studien darauf geachtet, dass sich diese auf Gesundheitssysteme mit einem dem deutschen ähnlichen ökonomischen Niveau beziehen.

2.1.3. Lehrmeinungen

Eine dritte Quelle für die Auswahl der Prüfkriterien ist der Forschungsstand in der wissenschaftlichen Literatur. Dieser hat den Vorteil, weniger durch politische Interessen geprägt zu sein, als vergleichbare Quellen von anderen Interessengruppen, die zwar die Auswahl von Prüfkriterien nicht bestimmen, aber doch beeinflussen. Dem entsprechend ist von akademischer Seite das größte Innovationspotenzial zu erwarten.

Hier wird zwischen Lehrbuchmeinungen und neuerer Fachliteratur unterschieden. Es wird im Folgenden deutlich werden, dass sowohl nationale als auch internationale Studien mit zentralen Verfahrensvorschlägen aus der Lehrbuchliteratur übereinstimmen. Dagegen werden in der neueren Fachliteratur weitere Prüfkriterien entwickelt, die zwar geringfügig aufwendiger in der Durchführung der Analyse sind, dafür aber bisweilen aussagekräftigere Ergebnisse liefern. An dieser Stelle existiert also ein großes Verbesserungspotenzial gegenüber altergebrachten Prozessen der Analyse der Dokumentationsqualität. Allerdings werden innovative Prüfkriterien aus der aktuellen Forschungsliteratur seltener bzw. oft überhaupt nicht angewendet oder empirisch überprüft. Deshalb sind Innovationen natürlich immer auch höhere Risiken immanent, da sie sich zwangsläufig noch nicht bewährt haben können. Im Folgenden wird deshalb ein Ausgleich zwischen Bewährtheit und Innovation angestrebt.

2.1.4. Ressourcenverbrauch und Invasivität

Selbstverständlich gehören zur Frage der Einsetzbarkeit von Prüfkriterien auch Überlegungen zu den ökonomischen und sozialen Aufwendungen, die beide so gering wie möglich gehalten werden sollen. Dabei wird in der ersten Phase auf Daten zurückgegriffen, die unabhängig von der Prüfung der Dokumentationsqualität erhoben worden sind, aber für weitere Zwecke wie der Prüfung zur Verfügung stehen. Dadurch entstehen keine zusätzlichen Kosten durch eine Erhebung, sodass allein die Analyseart den Aufwand bestimmt. Aufgrund des geringen Personalbedarfs fallen diese Kosten in der Gesamtanalyse nicht weiter ins Gewicht. Problematischer gestaltet sich der Aufwand in einer Stichprobenerhebung rechnerisch auffälliger Untersuchungsobjekte. Hier muss sowohl zusätzliches Krankenhaus- und Praxispersonal involviert als auch vonseiten der Auswertungsstelle Personal für die Erhebung vor Ort zu Verfügung gestellt werden. Dies bedeutet sowohl zusätzliche Personal- und Logistikkosten als auch Probleme beim Eingriff in den medizinischen Betrieb. Beides muss aber so gering wie möglich gehalten werden.

2.1.5. Diskriminationsfähigkeit

Zuletzt hängt die Geeignetheit eines Prüfkriteriums der Dokumentationsqualität davon ab, inwiefern es Rückschlüsse auf die Ursachen von Dokumentationsfehlern zulässt. Dabei wird zunächst zwischen systematischen Fehlern und Zufallsfehlern unterschieden (Sørensen et al. 1996). Beide Fehlerarten sollten minimiert werden. Dennoch sollte eine Priorisierung der Fehlererkennung vorgenommen werden. Zufallsfehler können nämlich leichter statistisch modelliert werden und entstehen lediglich durch Tipp-, Transkriptions- und Lesefehler, z. B. aufgrund schlechter Handschrift (Arts et al. 2002). Nicht spezifizierte systematische Fehler führen dagegen öfter zu statistischen Modellfehlern. Außerdem können mithilfe der Identifikation systematischer Fehler sowohl problematische Stellen im Erhebungsprozess – beispielsweise mangelhafte Datenfelddefinitionen und Ausfüllhinweise oder mangelhafter Ausschluss widersprüchlicher Daten – gefunden als auch potenzielle Fälle von Datenmanipulation aufgedeckt werden.

2.2. Hintergrund der Prüfkriterienauswahl

Es existierten verschiedene Typologien, Maße der Dokumentationsqualität zu systematisieren. So werden beispielsweise die Dimensionen der Akkuratessse, Komplettheit, Aktualität und der internen Konsistenz (Batini et al. 2006: 24,40) oder der Richtigkeit, Komplettheit und Konsistenz (Shortliffe et al. 1990: 515) unterschieden. Wir folgen hier der Typologie von Nonnemacher et. al (2007), die zwischen Richtigkeit, Plausibilität und Organisation unterscheidet. Dabei bevorzugen wir für das erste der Kriterien, der „Richtigkeit“ allerdings den Begriff „Akkuratessse“, weil letzterer verdeutlicht, dass bestenfalls die Übereinstimmung von Daten mit Referenzdaten gemessen werden kann.

In der Literatur finden sich zahlreiche Prüfkriterien, auf deren Ausschluss wir hier nicht einzeln eingehen können. Zumeist handelt es sich bei den ausgeschlossenen Kriterien um solche, die aufgrund logistischer Probleme oder gesetzlicher Bestimmungen nicht kostengünstig erhoben worden sind. Nur in den wenigsten dieser Fälle gibt es zudem zahlreiche Vergleichsstudien, die diese Prüfkriterien nutzen. Exemplarisch sei hier als Ausnahme das Kriterium der Aktualität genannt, das beispielsweise von der norwegischen Gesundheitsdirektion (Helsedirektoratet 2010) und vom britischen National Health Service (NHS 2007: 10) als Prüfkriterium verwandt wird. Die Umsetzung dieses Kriteriums liegt jedoch nicht bei den Leistungserbringern, sondern wird durch den gesetzlich vereinbarten Zeitplan fixiert, in den die Bundesauswertungsstelle nicht eingreifen kann.

Grundsätzlich ist allerdings festzuhalten, dass eine systematische Überprüfung der Dokumentationsqualität von Qualitätsindikatoren in Gesundheitssystemen überraschend selten vorgenommen wird. Eine Ausnahme sind hier die skandinavischen Länder, deren zentrale Organisationen zumindest rudimentär solche Analysen durchführen.

Auch zwischen den Sektoren gibt es Unterschiede. Im stationären Bereich existieren andere Mechanismen zur Überprüfung der Dokumentationsqualität als im ambulanten Bereich. In Letzterem gibt es vor allem Plausibilitätsprüfungen mit Blick auf möglichen Abrechnungsbetrug, insbesondere Zeitprüfungen (Ratzel et al. 2008: 10), die nicht zur Bewertung der Dokumentationsqualität von Qualitätsindikatoren geeignet sind.

Mit Blick auf diese Parameter haben wir insgesamt neun Kriterien ausgewählt, deren Nutzung wir für die systematische Überprüfung der Dokumentationsqualität empfehlen.

2.3. Ausgewählte Prüfkriterien

Aus der Fülle der von Nonnemacher et al. (2007) vorgeschlagenen Kriterien haben wir anhand der im vorangehenden Unterabschnitt herausgearbeiteten Merkmale folgende Prüfkriterien für unsere Empfehlung ausgewählt:

- Kriterien der Akkuratessse
 - Vollzähligkeit
 - Vollständigkeit
 - Konkordanz
- Kriterien der Plausibilität
 - Erlaubte Werte
 - Widerspruchsfreiheit
 - Werteverteilung
 - Bekannte Korrelationen
 - Fehlende Werte in optionalen Datenfeldern
- Kriterium der Organisation
 - Dubletten

Diese Kriterien werden in den folgenden Unterkapiteln ausführlicher dargestellt.

2.3.1. Vollzähligkeit

Zu den gängigsten Prüfkriterien der Dokumentationsqualität gehören Maße zur Komplettheit (completeness) der Daten (Batini et al. 2006: 24,40). Dieses Kriterium findet sich praktisch in allen neueren Lehrbüchern sowie auch in einschlägigen Richtlinien (Farquhar 2008: 10; IOM 2010) und Vergleichsstudien (z. B. Mc Colm 2009: 10f) wieder; manchmal sogar, neben der Analyse unerlaubter Werte, als einziges Prüfkriterium der Dokumentationsqualität (z. B. NHS 2009: 10). Da sowohl das AQUA-Institut (2010c: 11) als auch das BQS-Institut in seinen Studien Komplettheitskriterien nutzte und nutzt,² ist deren Weiterverwendung nicht infrage zu stellen.

Eines der zwei Komplettheitskriterien ist die Vollzähligkeit: der Anteil der tatsächlich dokumentierten Fälle unter den zu dokumentierenden Fällen. Sie wird erstmals bereits durch eine Soll-Ist-Statistik bei den Datenannahmestellen gemessen. Spezifisch für die sektorenübergreifende Vorgehensweise ist, dass die Bundesauswertungsstelle gegebenenfalls sektorenübergreifend einen Vergleich der Daten verschiedener Leistungserbringer vornehmen kann und so ein zusätzliches Maß an für die Vollzähligkeit erhält, das wesentlich resistenter gegen eine Manipulation durch einzelne Leistungserbringer ist. Dabei wurden im schwedischen Gesundheitssystem gute Erfahrungen in der Kreuzvalidierung verschiedener Datenbanken gemacht (Socialstyrelsen 2010). Dabei ist zu beachten, dass es leistungserbringerübergreifend zahlreiche andere Ursachen für vermeintlich mangelnde Vollzähligkeit geben kann, so zum Beispiel den Tod oder das freiwillige Ausscheiden von Patienten aus dem Behandlungszyklus, das Fehlen von Daten aus früheren potenziellen Untersuchungszyklen, die Möglichkeit noch zu erhebender Daten in der Zukunft usw. Aus diesem Grund eignen sich vor allem Daten, die eine frühere, nicht zu weit zurückliegende Behandlungssequenz implizieren, für den Vollzähligkeitsabgleich.

Hinsichtlich der Fehlererkennung kann die Vollzähligkeitsrate zwar dann Hinweise auf systematische Fehler geben, wenn sie unter den verschiedenen Leistungserbringern sehr unterschiedlich ist, auf die Ursachen solcher Unterschiede kann man allerdings nicht schließen. Tabelle 2 fasst die Eigenschaften des Vollzähligkeitskriteriums zusammen.

Tabelle 2: Eigenschaften des Kriteriums Vollzähligkeit

Anzahl der tatsächlich eingereichten Fälle im Vergleich zur erwarteten Fallanzahl		
Typ	Akkuratesse	
Fachliteratur	einheitlich, Lehrbuchkriterium	
Bewährtheit	national im Rahmen der externen stationären QS von AQUA und BQS verwendet, allerdings noch nicht im sektorenübergreifenden Kontext	
	internationaler Standard	
Bereich	leistungserbringerspezifisch	sektorenübergreifend
Aufwand	Erstellung einer Soll-Ist-Statistik, evtl. Abgleich mit Routinedaten	Erstellung einer Soll-Ist-Statistik; sektorenübergreifender Abgleich
Diskriminationsfähigkeit	problematisch, vor allem in QS-Verfahren mit kleiner Fallzahl kaum Einschätzung der Fehlerart möglich	
Fehlerursachenerkennung	insbesondere auf der Ebene der Leistungserbringer	
Zuständigkeit	Datenannahmestelle, Leistungserbringer	Bundesauswertungsstelle

2.3.2. Vollständigkeit

Das Pendant zur Vollzähligkeit ist die Vollständigkeit, welche den Anteil der fehlenden Werte bei einem obligatorisch zu erhebenden Datensatz bezeichnet (zu fehlenden Werten in optionalen Datenfeldern siehe unter 2.3.8). Auch dieses Kriterium bedarf keiner besonderen Legitimation. Es ist in allen aktuellen einschlägigen Lehrbüchern aufgeführt (bspw. Batini et al. 2006: 26) und wird im nationalen wie im internationalen Kontext durchgängig erhoben. Der Aufwand für die Generierung des Kriteriums fällt noch geringer als bei der Vollzähligkeit aus, kann man doch die Daten der mandatorischen Datenfelder auf gültige Inhalte überprüfen, ohne eine Soll-

² <http://www.bqs-qualify.com/online/public/leistungen/qualify/praktik/gk20> für die Vollzähligkeit und <http://www.bqs-qualify.com/online/public/leistungen/qualify/praktik/gk19> (abgerufen am 1. April 2011) für die Vollständigkeit.

Ist-Statistik zu erstellen. Unvollständige Daten weisen immer auf Fehler in der Softwarekonfiguration hin, denn diese sollte die Annahme unvollständiger Daten eigentlich verweigern.

Tabelle 3: Eigenschaften des Kriteriums Vollständigkeit

Anteil der fehlenden Werte („missing values“) bei mandatorischen Datenfeldereinträgen		
Typ	Akkuratesse	
Fachliteratur	einheitlich, Lehrbuchkriterium	
Erfahrungen	national im Rahmen der externen stationären QS von AQUA und BQS verwendet, allerdings noch nicht im sektorenübergreifenden Kontext	
	international Standard	
Bereich	leistungserbringerspezifisch	sektorenübergreifend
Aufwand	gering	
Diskriminationsfähigkeit	höher als bei der Vollzähligkeit, da andere Datenfelder vorhanden sind	
Fehlerursachenerkennung	Möglichkeit systematische Fehler zu erkennen, nach qualitativer Begutachtung mögliche Rückschlüsse auf Fehlerursache	
Zuständigkeit	Datenprüfprogramm der Datenannahmestelle	Bundesauswertungsstelle

2.3.3. Konkordanz

Ein besonders aussagekräftiges Plausibilitätskriterium ist die Konkordanz, d. h. die Übereinstimmung der Daten mit einer Referenzquelle, deren Angaben als Abbildung der Realität angenommen werden. Die übliche Referenzquelle für medizinische Daten ist dabei die Krankenakte (Arts et al. 2002: 605), wobei sich die Papierakten zwar als geringfügig valider als elektronische Akten herausgestellt haben (Stausberg et al. 2003: 473), aber dennoch keineswegs auch nur annähernd fehlerfrei sind (Stange et al. 1998; Stausberg 2007). Auch dieses Kriterium entspricht dem Lehrbuchstandard (Batini et al. 2006: 4ff; Nonnemacher et al. 2007: 15f). Da jedoch ein Vergleich der Krankenakte, besonders dann, wenn diese nur in Papierform vorliegt, einen hohen Ressourcenaufwand bedeutet (Besuch des Leistungserbringers, Gewährleistung des Datenschutzes, organisatorischer Aufwand des Leistungserbringers), wird dieses Kriterium nur selten überprüft. Trotzdem existieren bereits einige fruchtbare Anwendungen in Dänemark und Australien (McKenzie et al. 2005; Mosbech et al. 1995; Webster et al. 2010). Wir empfehlen daher zur Überprüfung eine Klumpenstichprobe statt einer Untersuchung der Grundgesamtheit durch die Datenauswertungsstelle. Die soll primär deshalb erfolgen, da so gängige Fehlerursachen erkannt werden können. Einen Spezialfall der Konkordanz stellen bildgebende Verfahren wie Röntgen, CT o.ä. dar, die als Befund in der Papierakte erscheinen und einem zusätzlichen Befundungs-Bias durch den Arzt unterliegen. Zu beachten ist des Weiteren, dass es hier einen *trade-off* zwischen der Genauigkeit der Daten und der Konkordanz gibt: Intervallskalierte Daten sind anfälliger für Inkongordanzen als kategoriale Daten (Willis et al. 2011: 3), was bedeutet, dass erstere mit Toleranzbereichen ausgestattet werden sollten.

Tabelle 4: Eigenschaften des Kriteriums Konkordanz

Übereinstimmung des Wertes eines Datenelements mit einer Referenzquelle	
Typ	Akkuratesse
Fachliteratur	einheitlich, Lehrbuchkriterium
Bewährung	national im Rahmen der externen stationären QS von AQUA und BQS verwendet, allerdings noch nicht im sektorenübergreifenden Kontext
	International vereinzelt angewandt
Bereich	leistungserbringerbezogen (einzelne Datensätze)
Aufwand	sehr aufwendig, weil als Referenzquelle zumeist die Krankenakte empfohlen wird
Diskriminationsfähigkeit	hängt von der Qualität der Referenzquelle ab
Fehlerursachenerkennung	hängt von der Qualität der Referenzquelle ab
Zuständigkeit	Bundesauswertungsstelle, Leistungserbringer

2.3.4. Erlaubte Werte

An der Grenze zwischen Akkuratesses- und Plausibilitätskriterien liegt der Anteil des Auftretens unzulässiger Werte (beispielsweise eine negative Altersangabe oder ein nicht-existierender ICD-Code). Auch dieses Kriterium ist Lehrbuchstandard (Nonnemacher et al. 2007: 23; Olsen 2003: 155ff), wird aber seltener in aktuellen Forschungsarbeiten rezipiert. Jedoch ist davon auszugehen, dass unerlaubte Datenfeldeinträge in der Regel wie fehlende Werte behandelt werden und somit zugleich mit der Vollständigkeit überprüft werden. Zudem weisen auch unerlaubte Werte auf eine mangelnde Konfiguration der Erhebungssoftware hin. Da die Prüfung auf unerlaubte Werte durch einfache Softwarelösungen vorgenommen werden kann, also wenig aufwändig ist, empfehlen wir eine Vorgehensweise wie bei der Vollständigkeitsprüfung, wobei ein Großteil der Prüfung bereits durch die Software beim Leistungserbringer vorgenommen wird. Ein Spezialfall ist die Prüfung des Pseudonyms, welche beim sektorenübergreifenden Datenabgleich auf Gültigkeit geprüft werden kann, wenn zwei ansonsten identische Datensätze unterschiedliche Pseudonyme aufweisen. Tabelle 5 charakterisiert dieses Kriterium noch einmal.

Tabelle 5: Eigenschaften des Kriteriums Erlaubte Werte

Anteil des Auftretens unzulässiger Werte		
Typ	Plausibilität	
Fachliteratur	Lehrbuchstandard	
Bewährung	national und international Standardprozedur, allerdings bisher noch nicht im sektorenübergreifenden Kontext angewendet	
Aufwand	Entwicklung von Filtersoftware	
Bereich	leistungserbringerbezogen	sektorenübergreifend hinsichtlich Pseudonymisierung
Diskriminationsfähigkeit	wie bei Vollständigkeit	
Fehlerursachenerkennung	wie bei Vollständigkeit	
Zuständigkeit	Datenprüfprogramm der Datenannahmestellen	Bundesauswertungsstelle

2.3.5. Widerspruchsfreiheit

Ein wesentlich geringerer Aufwand entsteht bei der Prüfung des Plausibilitätskriteriums der Widerspruchsfreiheit, bisweilen auch als „interne Validität“ bezeichnet. Bei diesem Lehrbuchkriterium (Chin et al. 2008: 304; Körner et al. 2006: 112; Nonnemacher et al. 2007: 17f) werden Wertepaare bzw. Wertetupel auf einander ge-

gegenseitig ausschließende Werte geprüft. Beispielsweise würde eine Schwangerschaft eines Mannes als Widerspruch gesehen. Auch hier übernimmt die Erhebungssoftware beim Leistungserbringer einen Großteil der Prüfung bereits vor der Datenübermittlung. Bei der sektorenübergreifenden Qualitätsprüfung ergibt sich für die Datenauswertungsstelle darüber hinaus noch Möglichkeit, Daten verschiedener Leistungserbringer auf ihre Konsistenz zu überprüfen. Dadurch gewinnt dieses Kriterium erheblich an Prüfkraft.

Bisher wurde, möglicherweise auch aufgrund einfacher Qualitätschecks bei der Datenerhebung, im nationalen Kontext dieses in Tabelle 6 beschriebene Kriterium zwar nicht erwähnt, aber praktisch angewandt (AQUA 2010d: 40ff). International wird es ebenfalls selten expliziert (z. B. Asch et al. 2004: 940). Vor allem aufgrund der erweiterten Möglichkeiten in der sektorenübergreifenden Prüfung empfehlen wir, nicht nur softwareseitig bei der Datenerhebung, sondern auch nach der Zusammenführung der Daten von der Bundesauswertungsstelle auf leistungserbringerübergreifende Widersprüche zu prüfen.

Tabelle 6: Eigenschaften des Kriteriums Widerspruchsfreiheit

Anteil der Wertetupel, die in sich widersprüchlich sind		
Typ	Plausibilität	
Fachliteratur	einheitlich, Lehrbuchkriterium	
Bewährung	national als Filter im Erhebungsinstrumenten	
	seltener auf internationaler Ebene geprüft	
Bereich	einrichtungsbezogen	sektorenübergreifend
Aufwand	Erstellung von Konsistenzregeln	
Diskriminationsfähigkeit	je nach Verteilung	
Fehlerursachenerkennung	je nach Verteilung	
Zuständigkeit	Softwareerkennung beim Leistungserbringer	Bundesauswertungsstelle

2.3.6. Werteverteilung

Ein weiteres gängiges Plausibilitätskriterium ist die Untersuchung der Werteverteilung. Darunter kann sowohl die Analyse von Ausreißern in der Werteverteilung fallen als auch die Form der Verteilung selbst, d. h. Überprüfung, ob die Verteilung der Werte der zur erwarteten Verteilung entspricht. Beispielsweise deuten überdurchschnittlich häufig durch 4 teilbare Atemfrequenzdaten auf Messungen im 15-Sekunden-Intervall hin, was die Genauigkeit der Daten beeinflusst. Derartige Kriterien werden standardmäßig in Lehrbüchern besprochen (bspw. Batini et al. 2006: 86; Maydanchik 2007: 73). Sie finden vereinzelt in empirischen Studien Anwendung. Die australisch-neuseeländische Gesellschaft für Intensivmedizin hat beispielsweise eine eigene Richtlinie für den Umgang mit Ausreißern (ANZICS 2008), während sich die AHRQ bisher lediglich explorativ mit solchen Daten befasst (Rand Health 2007: 45), obwohl diese Art der Datenprüfung mit recht geringem technischen Aufwand sehr aussagekräftige Ergebnisse erzielen kann. Dies wird zwar in der Auswertung der Qualitätsindikatoren selbst beachtet (Campbell et al. 2002: 358; Fink et al. 2007: 839), jedoch zumeist nicht als mögliches Maß der Dokumentationsqualität.

Tabelle 7: Eigenschaften des Kriteriums Werteverteilung

Plausibilität der Werteverteilung	
Typ	Plausibilität
Fachliteratur	Standardkriterium
Bewährtheit	im internationalen Kontext vereinzelt verwendet
Aufwand	gering
Diskriminationsfähigkeit	gut
Fehlerursachenerkennung	hoch für Genauigkeitsfehler und Dateneingabefehler
Zuständigkeit	Bundesauswertungsstelle, Leistungserbringer

2.3.7. Bekannte Korrelationen

Das in Tabelle 8 überblicksartig dargestellte Plausibilitätskriterium der Replikation bekannter Korrelationen ist bis jetzt weder Standard in der Lehrbuchliteratur, noch ist uns ein empirisches Beispiel seiner Anwendung im nationalen oder internationalen Kontext bekannt. Dennoch fand es Einzug in die Fachliteratur (Hasford et al. 1994; Nonnemacher et al. 2007: 26; Stausberg et al. 2006: 480). Dieses Kriterium eignet sich besonders dafür, Täuschungen aufzudecken. Auf der Kostenseite stehen hier zwar hohe Anforderungen an die Entwicklung geeigneter Korrelationsmatrizen zur Identifizierung bekannter Korrelationen, aber solche Matrizen können über Jahre hinweg verwendet werden. Zudem ist kein zusätzlicher Erhebungsaufwand gegeben, daher empfehlen wir die Einführung dieses Kriteriums.

Tabelle 8: Eigenschaften des Kriteriums Bekannte Korrelationen

Replikation bekannter Korrelationen	
Typ	Plausibilität
Fachliteratur	(noch?) nicht Standard der Lehrbuchliteratur, aber etabliert in der Fachliteratur
Bewährtheit	keine Vergleichsstudien bekannt
Aufwand	hoher Aufwand bei der Ermittlung der möglichen Korrelationen; mittlerer Aufwand bei der Datenaufbereitung
Diskriminationsfähigkeit	sehr hoch
Fehlerursachenerkennung	gutes Instrument zur Identifizierung manipulierter Daten, weil der Manipulationsaufwand hoch wäre
Zuständigkeit	Bundesauswertungsstelle, Leistungserbringer

2.3.8. Fehlende Werte in optionalen Datenfeldern

Strukturell ähnlich zu dem Kriterium der bekannten Korrelationen ist das in Tabelle 9 zusammengefasste Kriterium des Anteils der fehlenden Werte bei optionalen Datenfeldeinträgen (Nonnemacher et al. 2007: 21). Hier werden bekannte empirische Ausfallraten mit den tatsächlich auftretenden Ausfallraten verglichen, um statistisch auffällige Abweichungen (typischerweise um mehr als zwei Standardabweichungen bei Daten, bei denen keine a-priori-Standards existieren) zu identifizieren. Wie bei den bekannten Korrelationen ist uns auch hier keine Vergleichsstudie bekannt. Die von Nonnemacher, Weiland und Stausberg (2007: 21) zitierte Studie von Aronsky und Haug (2000) arbeitet zwar mit optionalen Datenfeldern, jedoch nicht, um damit Dokumentationsqualität zu messen. Aus empirischer Sicht ist daher noch keine Aussage zu diesem Kriterium zu treffen. Da der Aufwand, diese Daten zu testen, jedoch ein rein rechnerischer ist, spricht nichts dagegen, auch dieses Kriterium in die Bewertung mit aufzunehmen.

Tabelle 9: Eigenschaften des Kriteriums Fehlende Werte in optionalen Datenfeldern

Anteil der fehlenden Werte („missing values“) bei optionalen Datenfeldereinträgen	
Typ	Plausibilität/Akkuratesse
Fachliteratur	selten
Bewährtheit	nicht bekannt
Aufwand	gering
Diskriminationsfähigkeit	hängt von der Verteilung ab
Fehlerursachenerkennung	hängt von der Verteilung ab
Zuständigkeit	Bundesauswertungsstelle, Leistungserbringer

2.3.9. Dubletten

Unter Dubletten verstehen wir hier zwei oder mehr Datensätze, die mit hinreichender Wahrscheinlichkeit identischen Untersuchungsobjekten zugeordnet werden können. Obwohl das Organisationskriterium des Dublettenanteils sowohl in Lehrbüchern (Batini et al. 2006: 24; Nonnemacher et al. 2007: 29) als auch in der Fachliteratur (z. B. Winkler 2004) häufig diskutiert wird, wird es in empirischen Untersuchungen selten ausgewiesen – so auch nicht auf nationaler Ebene vom BQS-Institut. Dies könnte daran liegen, dass bei modernen Datensäuberungen im Vorfeld der Datensammlung häufig Dubletten eliminiert werden. Nichtsdestotrotz sollte dieses Kriterium nicht grundlos vernachlässigt werden. Zum einen ist der Kostenaufwand minimal: Es müssen lediglich geeignete Softwarelösungen erstellt werden. Zum Zweiten ist der Dublettenbestand empirisch gesehen nicht notwendigerweise gering: Beispielsweise fand eine kalifornische Studie zur Qualitätssicherung 3,6-9% Dubletten in verschiedenen Datenbeständen (Klein 2002). Hinzu kommt, dass mit bereits seit mindestens 10 Jahren verfügbaren, aber bisher weitgehend ungenutzten Techniken der „fuzzy searches“ im Data Mining auch nicht-exakte Dubletten gefunden werden können (Supekar et al. 2002), die eventuell Rückschlüsse auf Manipulationen – beispielsweise das „Auffüllen“ von Problem Daten mit unverdächtigen Daten aus anderen Fällen – zulassen würden. In der leistungserbringerübergreifenden Prüfung könnte dies bei widersprüchlichen Daten zwischen zwei Datenlieferanten sogar den Fehlerverursacher „identifizieren“ (Pseudonym), weil dieser annähernde Dubletten liefern würde. Auch deshalb obliegt die Prüfung des in Tabelle 10 dargestellten Kriteriums der Bundesauswertungsstelle, die als einzige verschiedene Datenbestände zusammenführen darf.

Tabelle 10: Eigenschaften des Kriteriums Anteil der Dubletten

Rate der mehrfach auftretenden Fälle	
Typ	Datenstruktur
Fachliteratur	Lehrbuchkriterium
Bewährtheit	seltener diskutiert
Aufwand	gering
Bereich	leistungserbringerbezogen mit sektorenübergreifenden Komponenten
Diskriminationsfähigkeit	gering, da aufgrund automatisierter Prozesse geringe Wahrscheinlichkeit; mögliche Erhöhung durch <i>fuzzy searches</i>
Fehlerursachenerkennung	evtl. Rückschlüsse auf Manipulation
Zuständigkeit	Bundesauswertungsstelle, Leistungserbringer
Bemerkung	sinnvoll zur gleichzeitigen Datensäuberung

2.4. Bewertung der Prüfkriterien

Zur Bewertung der Prüfkriterien schlagen Nonnemacher et al. (2007: 57) ein gleichsam sehr elaboriertes und starres Score-Schema vor. Dieses sieht Schwellenwerte vor, die auf Empfehlungen des Instituts für Medizinische Informatik, Biometrie und Epidemiologie (IMIBE) basieren. Allerdings sind diese Schwellenwerte keinesfalls international bewährt. Der NHS arbeitet beispielsweise im Fall der Vollzähligkeit mit vier Qualitätsstufen, die durch die Perzentile bei 97%, 90% und 80% Vollzähligkeit definiert sind (NHS 2007: 14), während die schwedische Gesundheits- und Sozialbehörde für zwei wichtige Datenbanken Vollzähligkeiten von 89% und 92% angibt (Socialstyrelsen 2010: 4). Insofern ist die vom IMIBE vorgeschlagene Marke von 90% durchaus plausibel. Aufgrund der organisatorisch sehr unterschiedlichen Strukturen der verschiedenen Gesundheitssysteme, empfehlen wir dennoch zunächst ein Benchmarking im ersten Jahr durchzuführen, anstatt derartige Werte 1:1 auf das deutsche Gesundheitssystem zu übertragen.

Noch problematischer sehen wir die Gewichtung der unterschiedlichen Kriterien. Keine der von uns betrachteten Datenbanken nimmt eine derartige Gewichtung vor, sondern bewertet die einzelnen Kriterien separat. Da es mit Komposita-Maßen von Dokumentationsqualität noch keine ausreichenden Erfahrungen gibt, schlagen wir deshalb vor, die Auswertung der verschiedenen Kriterien zunächst unabhängig voneinander durchzuführen und dabei auf allgemein übliche Auffälligkeitsmerkmale, beispielsweise Abweichung um mehr als zwei Standardabweichungen, zurückzugreifen. Mit Blick auf den notwendigen Aufwand werden die konkreten Bedingungen verfahrensspezifisch zu entscheiden sein.

3. Datenprüfungen im Prozess der Datenerhebung

Im Fokus des vorliegenden Konzeptes steht die Beschreibung des Datenvalidierungsverfahrens im Sinne des § 16 der Qesü-RL. Hierbei handelt es sich um ein Verfahren, bei dem retrospektiv Daten eines definierten Zeitraums geprüft werden. Vor dem Hintergrund des in der Anlage „Datenflussverfahren“ zur Qesü-RL dargestellten seriellen Datenflussmodells wird nachfolgend beschrieben, wie sich das Datenvalidierungsverfahren in die Datenflüsse der sektorenübergreifenden Qualitätssicherung eingliedert. Ferner soll dabei deutlich gemacht werden, dass die in Kapitel 2 erläuterten Prüfkriterien nicht sämtlich im Datenvalidierungsverfahren, sondern teilweise bereits zu früheren Zeitpunkten angewendet werden. Das serielle Datenflussmodell dient hier als Beispiel, weil es nach gegenwärtiger Planung zugleich das Grundmodell für die ersten umzusetzenden sektorenübergreifenden QS-Verfahren sein wird. Die dargestellten Aspekte lassen sich aber auch auf das in den tragenden Gründen zur Anlage „Datenflüsse“ der Qesü-RL beschriebene parallele Datenflussmodell übertragen.

Im Grundmodell des Datenflusses für die sektorenübergreifende Qualitätssicherung werden folgende Akteure unterschieden:

- Leistungserbringer
- Datenannahmestelle
- Vertrauensstelle
- Bundesauswertungsstelle

Die Primärerfassung der Daten (Dokumentation), die zum Zweck der sektorenübergreifenden Qualitätssicherung erhoben werden, erfolgt durch die Leistungserbringer. Bei dieser Erfassung werden die folgenden unterschiedlichen Daten erhoben:

- **Qualitätsdaten:** Daten, die die Behandlungen/Konsultationen in den verschiedenen Sektoren beschreiben. Dazu gehören u. a. Erkrankungen bei Aufnahme, Behandlungsanlässe, Angaben zu Verlauf und Ergebnis der Behandlung. Gemeinsam mit den Qualitätssicherungsdaten wird eine von den Leistungserbringern für den Dokumentationszeitraum und den einzelnen Patienten eindeutig generierte Fallnummer geliefert, die nur ihnen die Identifikation des Patienten in den Rückmeldeberichten ermöglicht. Die Fallnummer ist daher von essenzieller Bedeutung für die Umsetzung der Stichprobenprüfung (siehe Kapitel 4.2).
- **Patientenidentifizierende Angaben:** Daten, die eine eindeutige Identifizierung des Patienten ermöglichen, z. B. durch die Krankenversicherungsnummer. In der Qesü-RL werden diese Daten, zusammen mit den leistungserbringeridentifizierenden Daten, auch als personenbezogene Daten bezeichnet.
- **Leistungserbringeridentifizierende Daten:** Daten, die eine eindeutige Identifizierung des Leistungserbringers zulassen, z. B. durch die Betriebsstättennummer. In der Qesü-RL werden diese Daten, zusammen mit den patientenidentifizierenden Daten, auch als personenbezogene Daten bezeichnet.
- **Administrative Daten:** Daten, die zur Prüfung auf Plausibilität, Vollständigkeit und Vollzähligkeit geeignet sind sowie weitere, meldebezogene Daten. Die meldebezogenen Daten umfassen Daten, die insbesondere der Organisation des Datenflusses dienen, z. B. den meldenden Leistungserbringer bis zur Pseudonymisierung durch die Datenannahmestelle nach §9 Abs. 1 der Qesü-RL, Standort, Zeitstempel und Nummer des Datensatzes, das Thema der jeweiligen Datenlieferung sowie Angaben über Kommunikationspartner.

Die dokumentierten Daten werden von den Leistungserbringern verschlüsselt und an die Datenannahmestellen übermittelt. Nach erfolgreicher Prüfung der Daten, ersetzen die Datenannahmestellen die die Leistungserbringer (LE) identifizierenden Angaben durch ein LE-Pseudonym. Anschließend übermitteln sie die Daten mit entsprechender Verschlüsselung an die Vertrauensstelle.

In der Vertrauensstelle werden die patientenidentifizierenden Daten entschlüsselt, durch ein verfahrensspezifisches Patientenpseudonym ersetzt und mit entsprechender Verschlüsselung an die Bundesauswertungsstelle übermittelt.

In der Bundesauswertungsstelle erfolgt die personenbezogene Zusammenführung und Prüfung der Daten aus unterschiedlichen Sektoren bzw. zu unterschiedlichen Erhebungszeitpunkten. Auf Grundlage der zusammengeführten Daten werden Auswertungen nach bundeseinheitlichen Kriterien erstellt. Dies gilt sowohl für die landesbezogenen als auch die bundesbezogenen Verfahren. Die landesbezogenen Auswertungen werden von der Bundesauswertungsstelle an die Landesarbeitsgemeinschaften (LAG) weitergeleitet, denen auch die Durchführung der Stellungnahmeverfahren obliegt. Zudem muss die Bundesauswertungsstelle die geprüften Daten für die länderbezogenen Verfahren vorhalten und Nutzungsmöglichkeiten gewähren.

Nachfolgend werden die in Kapitel 2 erläuterten Prüfkriterien zu den Akteuren der sektorenübergreifenden Datenflüsse zugeordnet. Da die Aufgaben zum Teil variieren, müssen unterschiedliche vertragliche Zusammenhänge berücksichtigt werden:

- Stationäre Versorgung (Krankenhäuser)
- Ambulante, kollektivvertragliche Versorgung
- Ambulante, selektivvertragliche Versorgung

Die Zuordnung der Prüfkriterien erfolgt nach den Vorgaben der Qesü-RL. Soweit bestimmte Prüfungen dort noch nicht vorgesehen sind, wird nach dem Grundsatz verfahren, dass Datenprüfungen zum frühestmöglichen Zeitpunkt durchgeführt werden sollten.

Für die Zuordnung der in Kapitel 2 erläuterten Prüfkriterien der Dokumentationsqualität werden aus Gründen der besseren Darstellbarkeit folgende Abkürzungen verwendet:

- Akkuratesse
 - Vollzähligkeit (VZ)
 - Vollzähligkeit, sektorenübergreifend (VZ-SÜ)
 - Vollständigkeit (VS)
 - Vollständigkeit, sektorenübergreifend (VS-SÜ)
 - Konkordanz (Kon)
- Plausibilität
 - erlaubte Werte (eW)
 - erlaubte Werte, sektorenübergreifend (eW-SÜ)
 - Widerspruchsfreiheit (Wid)
 - Widerspruchsfreiheit, sektorenübergreifend (Wid-SÜ)
 - Werteverteilung (WV)
 - bekannte Korrelationen (Korr)
 - fehlende Werte (fW)
- Datenstruktur/-organisation
 - Dubletten (Dub)

Ablauf der Datenprüfung

In den Abbildungen 1 bis 3 sind die verschiedenen Datenprüfungen vor dem Hintergrund der Datenflüsse in der sektorenübergreifenden Qualitätssicherung dargestellt. Auf Ebene der Leistungserbringer erfolgen die ersten Datenprüfungen zur Datenstruktur, Plausibilität und Akkuratesse direkt bei der Eingabe. Vorgaben hierzu sind in den Spezifikationen der Datenerfassungen niedergelegt und werden von den Softwareanbietern in den Instrumenten der Datenerfassung umgesetzt. Geprüft werden Vollständigkeit (VS), erlaubte Werte (eW) und Widerspruchsfreiheit der Daten (Wid).

Beim Export der Daten sollen die Leistungserbringer das Datenprüfprogramm gemäß §4 Abs. 2 der Qesü-RL verwenden. Dieses Vorgehen ist gemäß §2 Abs. 1 der Anlage zu Teil 1 der Qesü-RL verpflichtend. Das Daten-

prüfprogramm prüft ebenfalls die Vollständigkeit (VS), erlaubte Werte (eW) und Widerspruchsfreiheit der Daten (Wid). Grundlage ist hier der Exportdatensatz und nicht die dokumentierten Daten (z. T. bestehen hier Unterschiede: z. B. wird für den Export das Geburtsdatum in Alter umgerechnet). Im kollektivvertraglichen Bereich basiert diese Prüfung auf einem Protokoll. Im Zuge der Datenannahme prüft die Datenannahmestelle Vollständigkeit, erlaubte Werte und Widerspruchsfreiheit. In einem separaten Prozess prüft die Datenannahmestelle im stationären Sektor zusätzlich die Vollzähligkeit (VZ) der Daten. Voraussichtlich findet diese Prüfung einmal jährlich statt. Im kollektivvertraglichen Bereich findet die Prüfung der Vollzähligkeit voraussichtlich quartalsweise zusammen mit der Abrechnung statt. Für den selektivvertraglichen Bereich müssen hierzu noch Regelungen getroffen werden.

Abhängig von der Art eines Fehlers erfolgt eine Rückmeldung an den Leistungserbringer per Protokoll mit einem Hinweis oder der Aufforderung zur Korrektur oder zur Verweigerung der Datenannahme. Details zur Rückprotokollierung müssen in der Qesü-RL noch festgelegt werden.

In den Abbildungen 1 bis 3 ist auf der rechten Seite ein Protokollfluss dargestellt, der es erlaubt, auch Auffälligkeiten und Fehler, die bei weiteren Akteuren im Datenfluss auftreten, in ein Gesamtprotokoll zu übernehmen. Dieser Protokollfluss hat zwei wesentliche Vorteile: Zum einen wird sichergestellt, dass Akteure über auftretende Fehler informiert werden, sofern diese für ihre Aufgaben relevant sein könnten und zum anderen erhält der Leistungserbringer eine eindeutige Rückmeldung, ob die Datenlieferung fehlerfrei ist.

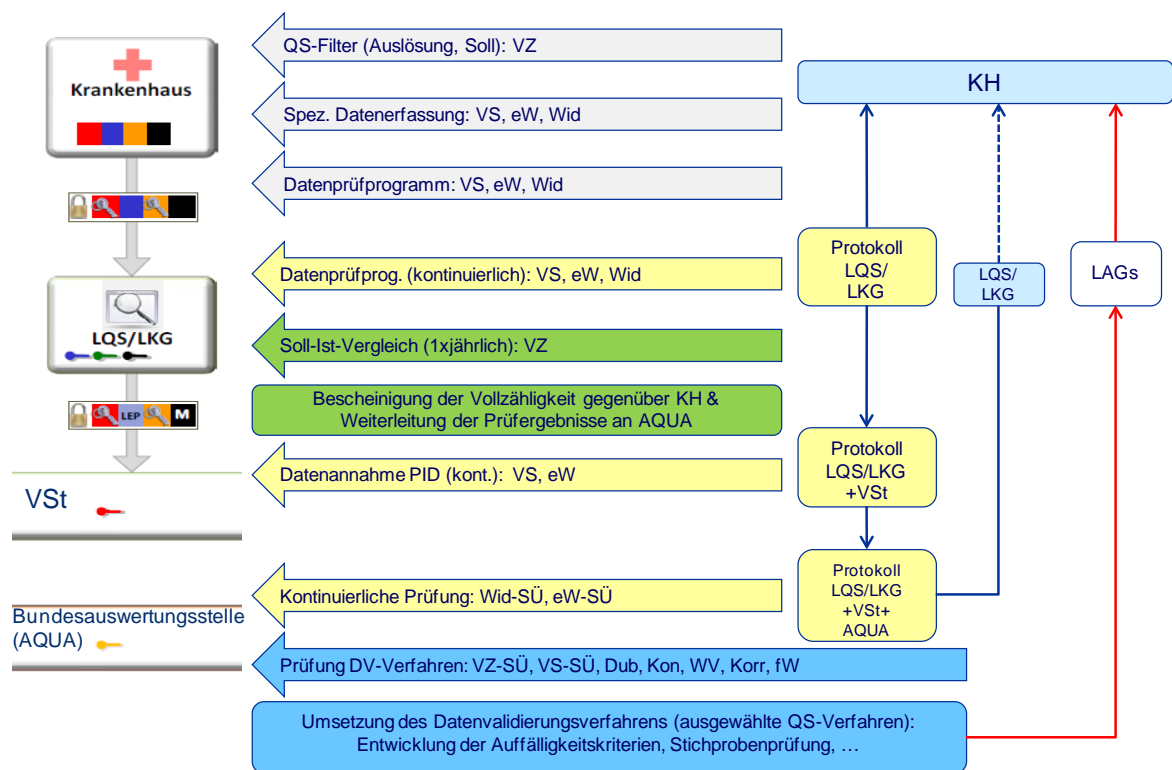


Abbildung 1: Datenfluss und Datenprüfungen bei Krankenhäusern³

Ebenfalls noch in der Qesü-RL festzulegen, sind die Prüfprozesse der Vertrauensstelle (VSt) und der Bundesauswertungsstelle. Hinsichtlich der Vertrauensstelle ist zu erwarten, dass diese vor der Pseudonymisierung prüfen wird, ob die personenidentifizierenden Daten (PID) vollständig (VS) und mit erlaubten Werten (eW) vorliegen. In der Bundesauswertungsstelle sind letztlich zwei Arten von Datenprüfungen zu unterscheiden:

- Datenüberprüfungen im Zuge des Datenaustausches
- Statistische Datenprüfungen auf Grundlage des Bundesdatenpools (Datenvalidierungsverfahren)

³ **Legende:** Dub – Dubletten; eW – erlaubte Werte; fW – fehlende Werte; Kon – Konkordanz; Korr – bekannte Korrelationen; SÜ – sektorenübergreifend; VS – Vollständigkeit; VZ – Vollzähligkeit; Wid – Widerspruchsfreiheit; WV – Werteverteilung.

Farblegende: grau – leistungserbringerinterne Prozesse; gelb – QS-Datenfluss; blau – Datenvalidierungsprozesse; grün – Vollzähligkeitsprüfung; hellblau – Rückprotokoll; rot – Stellungnahmeverfahren bei Auffälligkeit.

In der Bundesauswertungsstelle werden die über die Datenannahmestelle und die Vertrauensstelle weitergeleiteten Datenlieferungen angenommen und gegebenenfalls mit bereits vorhandenen Daten zu den jeweiligen Patienten verknüpft. Im Zuge der Verknüpfung wird geprüft, ob die zusammengeführten Datensätze insgesamt gesehen die erwarteten Werte enthalten (eW-SÜ) und widerspruchsfrei zueinander sind (Wid-SÜ). Dies ist beispielsweise nicht der Fall, wenn unterschiedliche Altersangaben für einen Patienten aus dem gleichen Erfassungsjahr vorliegen. Gegebenenfalls auftretende Auffälligkeiten werden im Rückprotokoll vermerkt, das an die Datenannahmestelle weitergeleitet wird. In welcher Form auch diese besonderen Auffälligkeiten an die Leistungserbringer weitergeleitet werden, muss noch geklärt werden. Beispielsweise tritt die Frage auf, welcher Leistungserbringer bei auftretenden Unstimmigkeiten zu benachrichtigen ist. Vermutlich ist es am praktikabelsten, wenn zunächst der Leistungserbringer informiert wird, der den letzten Datensatz zu einem Patienten gesendet hat. Dies muss jedoch im Zuge der weiteren Umsetzung sektorenübergreifender QS-Verfahren noch überprüft werden.

Statistische Datenprüfungen auf Grundlage des Bundesdatenpools (Datenvalidierungsverfahren)

Am Ende der Prüfkette schließlich steht das sogenannte Datenvalidierungsverfahren (siehe Kapitel 4). Hierunter sind Datenprüfungen zusammengefasst, die erst auf Grundlage zusammengeführter Daten eines definierten Zeitraums und mithilfe statistischer Analysen möglich sind. Es werden Auffälligkeitsprüfungen (siehe Kapitel 4) und Stichprobenprüfungen (siehe Kapitel 4.2) unterschieden. Folgende Prüfkriterien werden dabei angewendet:

- Prüfung auf Vollzähligkeit auf Grundlage sektorenübergreifend zusammengeführter Patientendaten (VZ-SÜ)
- Prüfung auf Vollständigkeit auf Grundlage sektorenübergreifend zusammengeführter Patientendaten (VS-SÜ)
- Prüfung auf Dubletten (Dub)
- Prüfung auf Konkordanz (Kon)
- Prüfung auf Auffälligkeiten bei der Werteverteilung (WV)
- Prüfung auf Auffälligkeiten bei Betrachtung bekannter Korrelationen (Korr)
- Prüfung auf fehlende Werte (fW)

Weitere Einzelheiten zum Datenvalidierungsverfahren sind in den folgenden Kapiteln dargestellt.

Besonderheiten der Datenprüfungen in der ambulanten, kollektivvertraglichen Versorgung

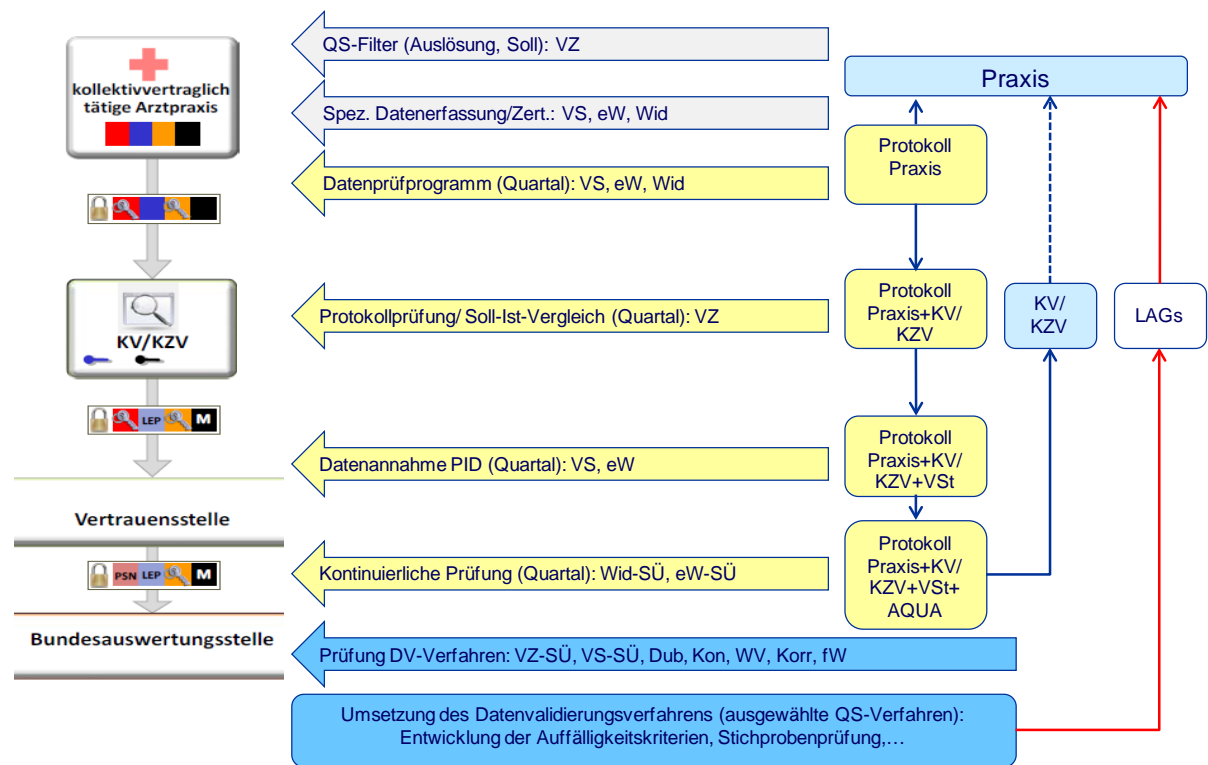


Abbildung 2: Datenfluss und Datenprüfungen bei kollektivvertraglich tätigen Arztpraxen⁴

Abbildung 2 stellt die Aufteilung der Datenprüfungen für ambulante, kollektivvertraglich tätige Leistungserbringer dar. Eine wesentliche Besonderheit besteht darin, dass gemäß der Qesü-RL die KVen und KZVen, als Datenannahmestellen in diesem Bereich, ihre Aufgaben im Rahmen der Datenprüfungen ohne Einblick in die Qualitätsdaten erfüllen müssen.

⁴ **Legende:** Dub - Dubletten; eW - erlaubte Werte; fW - fehlende Werte; Kon - Konkordanz; Korr - bekannte Korrelationen; SÜ - sektorenübergreifend; VS - Vollständigkeit; VZ - Vollzähligkeit; Wid - Widerspruchsfreiheit; WV - Werteverteilung
Farblegende: grau - leistungserbringerinterne Prozesse; gelb - QS-Datenfluss; blau - Datenvalidierungsprozesse; hellblau - Rückprotokoll; rot - Stellungnahmeverfahren bei Auffälligkeit

Besonderheiten der Datenprüfungen in der ambulanten, selektivvertraglichen Versorgung

Abbildung 3 stellt die Aufteilung der Datenprüfungen für ambulante, selektivvertraglich tätige Leistungserbringer dar. Eine wesentliche Besonderheit besteht darin, dass gemäß der Qesü-RL die Vertrauensstelle hier gleichzeitig die Aufgaben der Datenannahmestelle übernimmt. Dafür, wie die Vertrauensstelle als Datenannahmestelle erkennen kann, welche Leistungserbringer Daten liefern sollten, werden noch Regelungen zu treffen sein.

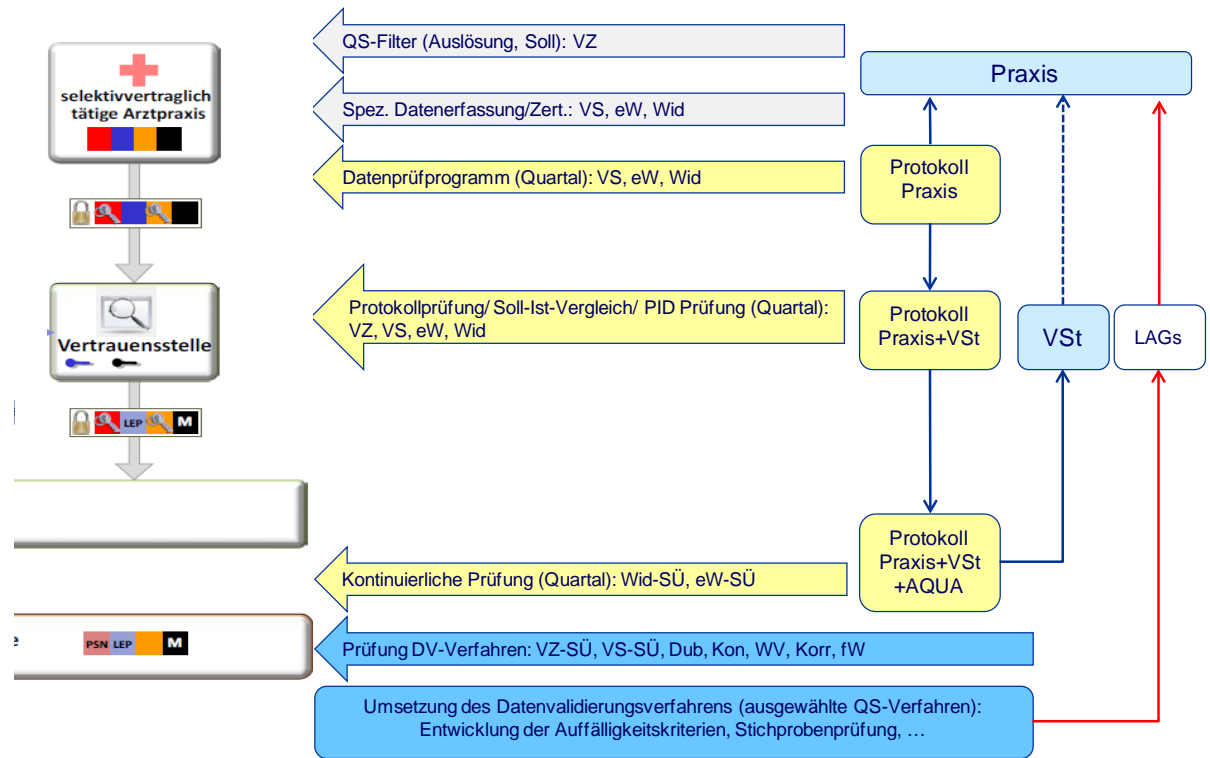


Abbildung 3: Datenfluss und Datenprüfungen bei selektivvertraglich tätigen Arztpraxen⁵

⁵ **Legende:** Dub – Dubletten; eW – erlaubte Werte; fW – fehlende Werte; Kon – Konkordanz; Korr – bekannte Korrelationen; SÜ – sektorenübergreifend; VS – Vollständigkeit; VZ – Vollständigkeit; Wid – Widerspruchsfreiheit; WV – Werteverteilung
Farblegende: grau – leistungserbringerinterne Prozesse; gelb – QS-Datenfluss; blau – Datenvalidierungsprozesse; hellblau – Rückprotokoll

4. Datenvalidierungsverfahren

Im Unterschied zu kontinuierlich stattfindenden Datenprüfungen bei der Dateneingabe oder im Zuge des Datenaustauschs, beschreibt das Datenvalidierungsverfahren Prüfungen, die auf Grundlage zusammengeführter Daten eines definierten Zeitraums durchgeführt werden. Innerhalb des Datenvalidierungsverfahrens werden Auffälligkeitsprüfungen und Stichprobenprüfungen unterschieden. Erstere beziehen sich auf den gesamten Datenpool ausgewählter QS-Verfahren, Letztere auf eine Auswahl von Leistungserbringern und/oder Patienten innerhalb ausgewählter QS-Verfahren.

Abbildung 4 stellt die Grundzüge des Datenvalidierungsverfahrens dar. Weitere Ausführungen zur Umsetzung enthält das Kapitel 5, nähere Aussagen zur Evaluation finden sich in Kapitel 5.4. In den nachfolgenden Abschnitten sind zunächst konzeptionelle Grundlagen der Auffälligkeits- und der Stichprobenprüfung dargestellt.

Grundsätzlich ist das Verfahren so angelegt, dass es jeweils auf die spezifischen Gegebenheiten eines bestimmten Themas angepasst werden kann. Folglich müssen für jedes QS-Verfahren, in dem ein Datenvalidierungsverfahren umgesetzt wird, spezifische konzeptionelle Arbeiten erfolgen. Unter Abwägung von Aufwand und Nutzen ist es jedoch nicht erforderlich, jährlich in jedem QS-Verfahren eine umfassende Datenvalidierung durchzuführen. Allerdings sollte langfristig in jedem QS-Verfahren eine Datenvalidierung stattgefunden haben, um Aussagen über die Güte der Daten treffen und ggf. Optimierungsmaßnahmen einleiten zu können. Zudem sollte in den Verfahren, in denen sich im Rahmen eines Datenvalidierungsverfahrens massive Dokumentationsprobleme gezeigt haben, zeitnah eine weitere Überprüfung der Dokumentationsqualität stattfinden.

Die primäre Zielsetzung und die konzeptionelle Ausgestaltung der beiden Verfahrensbestandteile werden in den folgenden Abschnitten beschrieben.

4.1. Auffälligkeitsprüfung

Die Auffälligkeitsprüfung hat zum Ziel, jene statistischen Auffälligkeiten zu identifizieren, die auf eine fehlerhafte Dokumentation hinweisen. Datenbasis ist der Bundesdatenpool für ein ausgewähltes QS-Verfahren. Um ein transparentes Vorgehen zu gewährleisten, sollen systematische Auffälligkeitskriterien formuliert werden. Diese sollen so konstruiert sein, dass Auffälligkeiten auf Ebene der Leistungserbringer bestimmt werden. Diesem Vorgehen unterliegt die Annahme, dass die Dokumentationsqualität zwischen den Leistungserbringern variiert und sich eine mangelnde Dokumentationsqualität folglich auch bei einzelnen Leistungserbringern niederschlägt. Für die rechnerisch auffälligen Leistungserbringer wird ein Stellungnahmeverfahren eingeleitet (vgl. § 17 Abs. 2 Qesü-RL), dessen Ziel es ist, Ursachen für die Auffälligkeit zu eruieren und gegebenenfalls Verbesserungsmaßnahmen einzuleiten.

Prüfkriterien der Auffälligkeitsprüfung

Grundsätzlich sollen alle drei der in Kapitel 2.3 vorgestellten Arten von Prüfkriterien (Akkuratesse, Plausibilität, Organisation) entwickelt werden. Allerdings kann dabei im Einzelnen das Konkordanzkriterium nicht berücksichtigt werden, weil die dafür notwendigen Referenzdaten erst in einer Stichprobe erhoben werden. Außerdem ist zu beachten, dass Auffälligkeitsprüfungen umso relevanter sind, je mehr unabhängige Datenlieferanten vorhanden sind. Schließlich kann das Kriterium der Replikation bekannter Korrelationen nur bei Verfahren und Leistungserbringern durchgeführt werden, die über eine ausreichend große Fallzahl verfügen.

Wie in Kapitel 2.4 begründet, sollen die verbliebenen sieben bzw. acht einzelnen Prüfkriterien separat ausgewertet werden. Sollte dies zu unbefriedigenden Ergebnissen führen, könnte nach einem in Kapitel 5.4 skizzierten Evaluationsprozess darüber nachgedacht werden, Indizes zu konstruieren, wie dies von Nonnenmacher et al. (2007: 55ff) vorgeschlagen und von Stausberg (2007: 1042) bekräftigt wurde. Ob es sich dabei um ein komplexes Schema wie das von Nonnenmacher et al. handeln sollte oder um ein einfacheres, wie es etwa von Gribben et al. (2001: 32) empfohlen wird, oder ob ggf. ein Index völlig neu zu konzipieren ist, müsste dann ebenfalls im Evaluationsprozess geklärt werden.

Kriterien zur Verstetigung von Auffälligkeitskriterien

Kriterien, die sich in den Auffälligkeitsprüfungen besonders bewährt haben, sollen auch in nachfolgenden Verfahrensjahren Anwendung finden (Verstetigung). Zur Bewertung der Auffälligkeitskriterien orientieren wir uns, in der Reihenfolge ihrer Relevanz, an den Messgrößen Trefferquote, Bedeutsamkeit und Praktikabilität. Als *Trefferquote* wird die Rate an aufgedeckten Fehldokumentationen bezeichnet, d. h. das Verhältnis zwischen rechnerischen Auffälligkeiten, die aufgrund von Fehldokumentationen entstanden sind, und den gesamten rechnerischen Auffälligkeiten. Je höher die Trefferquote ist, desto effektiver ist das Auffälligkeitskriterium darin, fehlerhafte Dokumentationen aufzudecken. Die *Bedeutsamkeit* eines Auffälligkeitskriteriums wird daran gemessen, ob dieses im Zusammenhang mit Qualitätsindikatoren steht, die für die öffentliche Berichterstattung bedeutsam oder anderweitig für die Qualitätsindikatoren (z. B. Basisauswertung) oder für die Risikoadjustierung relevant sind. Die *Praktikabilität* eines Auffälligkeitskriteriums wird daran bemessen, wie hoch die Anzahl an Auffälligkeiten ist, die sich durch die Anwendung auf den Bundesdatenpool ergeben, und wie hoch der dadurch ausgelöste Analyseaufwand für die betroffenen Leistungserbringer ist.

4.2. Stichprobenprüfung

Die Stichprobenprüfung ist ein Verfahren zur externen Validierung. Sie zielt darauf ab, quantitative Aussagen darüber treffen zu können, inwiefern die Angaben in den Datenfeldern der QS-Dokumentation die tatsächliche Versorgung abbilden. Um dies zu ermöglichen, wird ein externes Kriterium verwendet, von dem angenommen wird, dass es die reale Versorgungsqualität besser abbildet als die QS-Dokumentation. Sowohl in der externen stationären Qualitätssicherung als auch in der ambulanten Qualitätssicherung ist es üblich, dafür Angaben aus der Patientenakte heranzuziehen. Insbesondere können hier auch objektive Messwerte, wie sie z. B. aus Laborbefunden und bildgebenden Verfahren generiert werden, genutzt werden. Diesbezüglich kann das Stichprobenverfahren der KVen ein Vorbild sein. Für eine Stichprobe von Leistungserbringern und Patienten, zu denen es eine QS-Dokumentation gibt, soll dann eine Zweiterfassung ausgewählter⁶ Datenfelder der QS-Dokumentation anhand der Patientenakte erfolgen. Anschließend werden die fallbezogenen Angaben aus der Patientenakte mit denen aus der QS-Dokumentation abgeglichen.

Prüfkriterien der Stichprobenprüfung

Zentrales Prüfkriterium der Stichprobenprüfung ist die *Konkordanz* (siehe Kapitel 2.3.3), das einzige Prüfkriterium, das aufgrund seiner Beschaffenheit *nur* in der Stichprobenprüfung validiert werden kann und gleichzeitig auch das wichtigste Kriterium der Akkuratess ist. Außerdem können an dieser Stelle ebenfalls mögliche Dubletten überprüft und fehlende Werte nachgetragen werden.

Bildung der Stichproben

In den Vorgaben zum Konzept der Sektorenübergreifenden Sicherung der Dokumentationsqualität hat der G-BA festgelegt, dass 5% der Leistungserbringer in die Stichprobenprüfung einbezogen werden. Dabei soll das Konzept einer *einrichtungsbezogenen Klumpenstichprobe* angewendet werden. Der Hauptvorteil einer solchen Vorgehensweise ist die Kostenersparnis gegenüber einer patientenbezogenen einfachen Zufallsstichprobe, da wesentlich weniger Einrichtungen besucht werden müssen (Singh 2003: 765). Auf dieser Ebene können bereits für die Qualitätsdokumentation relevante Strukturangaben der Einrichtungen überprüft werden.

In einer ersten Stufe könnten je 5% der Leistungserbringer aus den an dem QS-Verfahren beteiligten Sektoren gezogen werden. Dabei würden wir die Wahrscheinlichkeiten in die Zufallsstichprobe zu gelangen gemäß der Sollfallzahl der Einrichtungen gewichten, weil andernfalls kleinere Einrichtungen in der Stichprobe überrepräsentiert wären (Lee et al. 2006: 6f). In einer zweiten Stufe würden dann 20 Fälle und 10 Reservefälle, für die dann die Angaben in den QS-Dokumentationen mit korrespondierenden Angaben in der QS verglichen werden, gezogen werden. Zusätzlich würden Quotenstichprobenkriterien angewandt werden, beispielsweise, dass mindestens ein Leistungserbringer aus jedem Bundesland gezogen werden muss.

⁶ Erfahrungen mit dem Stichprobenverfahren in der externen stationären QS haben gezeigt, dass aus Effizienzgründen eine Auswahl an Datenfeldern getroffen werden sollte, die sich besonders gut für eine Validierung anhand der Patientenakte eignen (z. B. Verfügbarkeit der Information in der Patientenakte, Relevanz des Datenfeldes für die Berechnung von Qualitätsindikatoren).

Des Weiteren könnten die Informationen der Auffälligkeitsprüfung als zusätzliches Schichtungskriterium berücksichtigt werden. Die Stichprobe könnte sich beispielsweise so zusammensetzen, dass diese 3% aller Leistungserbringer umfasst, die in der Prüfung unauffällig waren, während die restlichen 2% aus Einrichtungen generiert werden, die in der Prüfung auffällig geworden sind.

Gibt es für einen QS-Verfahren weniger als 2% Leistungserbringer mit rechnerischen Auffälligkeiten, könnten alle Leistungserbringer mit Auffälligkeiten gezogen werden. Anschließend könnten weitere Leistungserbringer ohne Auffälligkeiten gezogen werden, bis insgesamt 5% aller Leistungserbringer in der Stichprobe repräsentiert sind. Indem gezielt Leistungserbringer und Patienten in die Stichprobe aufgenommen werden, bei denen sich im Rahmen der Prüfung rechnerische Auffälligkeiten gezeigt haben, wäre es möglich, quantitative Aussagen über die Dokumentationsqualität der Daten insgesamt sowie über die den Auffälligkeitskriterien zugrunde liegenden Datenfelder zu erhalten. Zusätzlich könnten die die Auffälligkeitsprüfung begründenden theoretischen Annahmen empirisch überprüft werden. Falls die grundsätzlich plausible Annahme zutrifft, dass die in der Basisprüfung auffälligen Leistungserbringer auch häufiger Datenvalidierungsprobleme aufweisen, wäre damit eine präzisere Schätzung des Ausmaßes von Datenvalidierungsproblemen möglich (Schnell et al. 2008). Die Stichprobe würde also effektiver sein als eine einfache mehrstufige Stichprobe, wie sie im bisherigen Verfahren implementiert ist.

Diese so erhaltenen Informationen könnten dann ebenfalls im Rahmen des Stellungnahmeverfahrens, wie es für rechnerisch auffällige Leistungserbringer eingeleitet wird, berücksichtigt werden.

Andererseits hat dieses Vorgehen eine Reihe von bedeutsamen Nachteilen. So hängen die Anzahl der zu ziehenden Einrichtungen und damit die Größe der Stichprobe sehr stark von der Anzahl der Leistungserbringer ab. Dies hätte zur Konsequenz, dass für Verfahren (oder auch Sektoren) mit besonders vielen Leistungserbringern auch eine sehr umfangreiche Stichprobenprüfung durchgeführt werden müsste, sodass die notwendige Fallzahl für eine quantitative Ermittlung der Dokumentationsqualität übererfüllt oder zumindest nicht sonderlich effektiv wäre. Umgekehrt konnte in der Vergangenheit in der externen stationären QS auch die Erfahrung gemacht werden, dass eine Orientierung der Stichprobenziehung an einer fixen Prozentzahl der Leistungserbringer – bei Verfahren, die in wenigen Abteilungen durchgeführt werden – zu Ergebnissen führt, die eine seriöse Einschätzung der Dokumentationsqualität auf Basis dieser Stichprobe nicht mehr zulässt; etwa wenn nur 1 oder 2 Einrichtungen in der Stichprobe repräsentiert sind.

Aus rein methodischer Sicht wären absolute Ober- wie Untergrenzen von Stichproben sinnvoll. Da jedoch gemäß Anlage 1.1 (Kapitel 2.6.3) mindestens 5% der Leistungserbringer bzw. Patienten überprüft werden müssen, wäre die oben skizzierte Regelung dahingehend zu ergänzen, dass mindestens 200 Patienten bzw. 10 Einrichtungen in der Stichprobe vertreten sein sollten.

Grundsätzlich könnte angedacht werden, eine Stichprobenziehung auf Basis von Patienten in einem QS-Verfahren durchzuführen. Dies würde in einem Paneldesign allerdings ggf. zu einer deutlichen Vermehrung der in die Stichprobe einbezogenen Leistungserbringer führen und damit wenig effektiv sein. Da darüber hinaus, eine sinnvolle Überprüfung der richtigen Verknüpfung der Patienten nur über die Vertrauensstelle möglich ist, sollte dieser Weg eher nicht beschritten werden.

5. Praktische Umsetzung des Datenvalidierungsverfahrens

Nachdem im vorherigen Kapitel das Konzept zur Überprüfung der Dokumentationsqualität (Datenvalidierungsverfahren) beschrieben wurde, werden im Folgenden zentrale Aspekte der praktischen Umsetzung benannt. Abbildung 4 visualisiert die verschiedenen Schritte, auf die es auf die praktische Umsetzung ankommt.

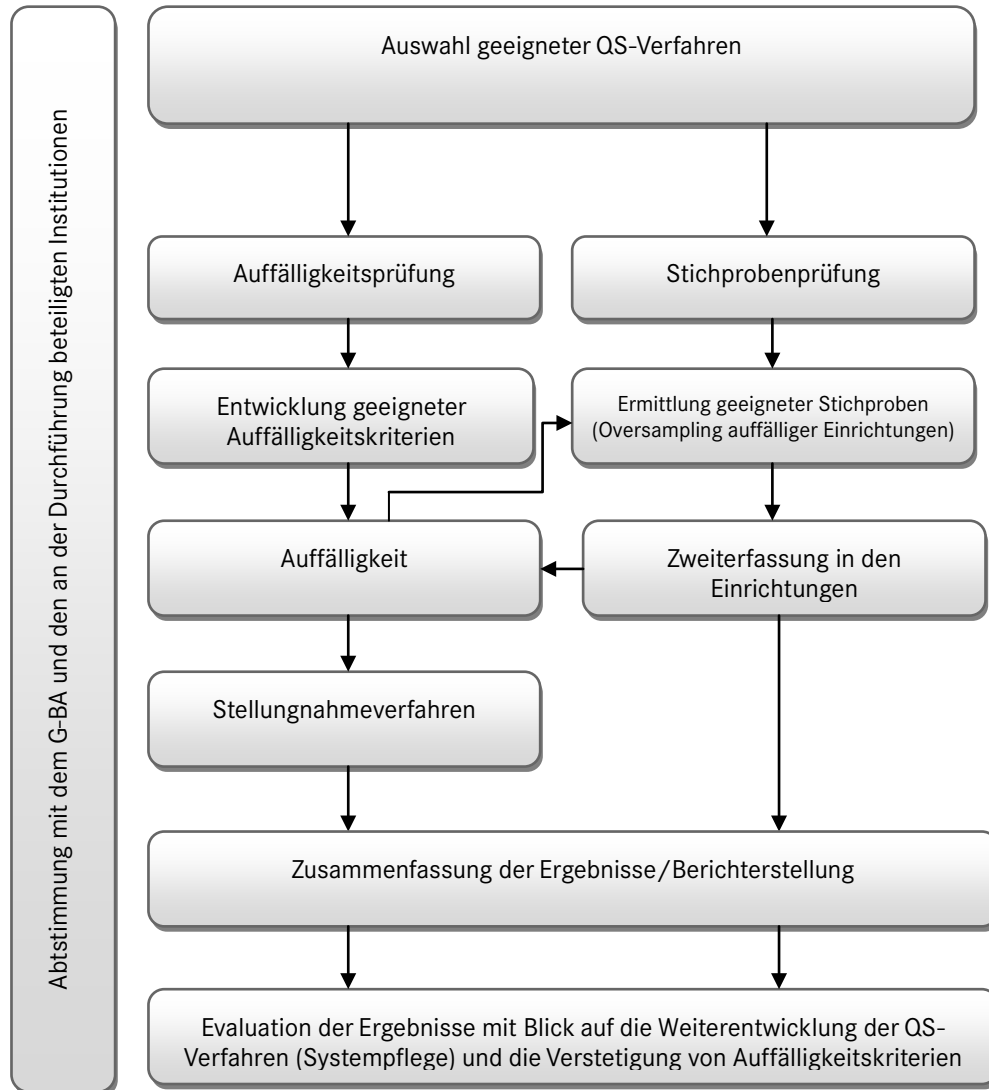


Abbildung 4: Datenvalidierungsverfahren in der sektorenübergreifenden Qualitätssicherung

Grundsätzlich empfehlen wir, die Prozesse zur Bewertung der Auffälligkeiten und zur Durchführung von Qualitätssicherungsmaßnahmen (vgl. § 17 der Qesü-RL) bezogen auf das Datenvalidierungsverfahren und die Ergebnisse der Qualitätsindikatoren, aufeinander abzustimmen. Hiervon ausgehend wird im Weiteren die praktische Umsetzung des Datenvalidierungsverfahrens beschrieben. Zur Systematisierung des praktischen Umsetzungsprozesses wird dieser in drei Phasen untergliedert:

- Vorbereitende Maßnahmen
- Durchführung
- Evaluation des Verfahrens und Verfassen des Berichts (siehe Kapitel 5.4)

In den nächsten Abschnitten werden zu jeder Phase die zentralen Akteure benannt, die bei der praktischen Umsetzung des Verfahrens von Bedeutung sind. Eine tabellarische Zusammenfassung mit Zeitplan findet sich jeweils am Ende des Abschnitts.

An dieser Stelle sei darauf hingewiesen, dass der Zeitplan des Umsetzungsprozesses sich an dem bisherigen Zeitplan des Datenvalidierungsverfahrens in der externen stationären Qualitätssicherung orientiert. Grund dafür ist, dass die Zeitpläne kompatibel sein sollten, um für die stationäre und sektorenübergreifende Qualitätssicherung keine parallelen zeitlichen Abläufe entstehen zu lassen. Allerdings könnten die themenspezifischen Bestimmungen es erforderlich machen, im Einzelfall von den hier benannten Zeitvorgaben abzuweichen.

5.1. Vorbereitende Maßnahmen

Um eine effektive Umsetzung des Datenvalidierungsverfahrens zu gewährleisten, sind bereits in dem Jahr, in dem die Erfassung der Qualitätsdaten erfolgt, vorbereitende Maßnahmen zu treffen. Wie weiter oben beschrieben, ist es nicht empfehlenswert, jährlich eine umfassende Datenvalidierung in jedem QS-Verfahren durchzuführen. Folglich ist es erforderlich, jedes Jahr eine Auswahl an geeigneten QS-Verfahren zu treffen, in denen eine Datenvalidierung umgesetzt werden soll. Im Rahmen der gesetzlichen Vorgaben ist in der stationären Qualitätssicherung (bisher) das folgende Verfahren vorgesehen, das sich auch in der sektorenübergreifenden Qualitätssicherung anbieten würde: Die Institution nach § 137a SGB V schlägt geeignete QS-Verfahren vor, der Unterausschuss Qualitätssicherung des G-BA wählt dann die endgültigen QS-Verfahren aus. Für einen transparenten Auswahlprozess sollen im Rahmen einer Projektgruppe zur Datenvalidierung in der sektorenübergreifenden Qualitätssicherung Kriterien entwickelt werden, die für eine Auswahl geeigneter QS-Verfahren hilfreich sind. Kriterien zur Auswahl geeigneter QS-Verfahren sind:

- QS-Verfahren, die bisher nicht in die Datenvalidierung einbezogen waren
Begründung: Die Datenvalidierung liefert wichtige Hinweise zur Interpretation und Weiterentwicklung der Indikatoren und der Dokumentation.
- QS-Verfahren, deren Datensätze neu entwickelt oder überarbeitet wurden
Begründung: Änderungen und Neuentwicklungen sollten einerseits früh validiert werden. Da das erste Jahr aber meist eine Lernkurve zeigt, sollte eine Überprüfung andererseits erst nach einer einjährigen Übergangsphase erfolgen.
- QS-Verfahren mit Qualitätsindikatoren, die verpflichtend im Qualitätsbericht der Krankenhäuser zu dokumentieren waren
Begründung: Die Verpflichtung zur Veröffentlichung stellt hohe Anforderungen an die Richtigkeit der Daten, die der Berechnung eines Indikators zugrunde liegen.
- QS-Verfahren, deren Qualitätsindikatoren eng an hochwertige Leitlinien gekoppelt sind
Begründung: Leitliniengestützte Indikatoren genießen die größte Akzeptanz aufgrund ihrer Basierung auf Evidenz und Konsens. Sie zeigen im Strukturierten Dialog und in der Verlaufsdarstellung über mehrere Jahre die größten Verbesserungsraten.
- QS-Verfahren mit Qualitätsindikatoren mit besonderem Handlungsbedarf
Begründung: Für diese Indikatoren werden besondere Maßnahmen zur Verbesserung eingeleitet. Hier ist es wichtig zu wissen, in welchem Ausmaß Fehldokumentationen zu den Indikatorergebnissen beitragen (Datenabgleich). Gleichzeitig können Auffälligkeiten in den Daten statistisch identifiziert und prospektiv korrigiert werden (Basisprüfung).

Unabhängig von den vorgenannten Kriterien, sollte jedes neu etablierte QS-Verfahren spätestens vier Jahre nach Einführung das Datenvalidierungsverfahren durchlaufen haben.

Nach dem Beschluss über die ausgewählten QS-Verfahren werden von der Institution nach § 137a SGB V Datenfelder für die Zweiterfassung und Auffälligkeitskriterien für die statistische Basisprüfung separat für jedes QS-

Verfahren vorgeschlagen und dem G-BA vorgelegt, der eine Implementierung verabschiedet. Die Auswahl der Datenfelder und die Entwicklung der Auffälligkeitskriterien sollen im Rahmen einer Projektgruppe zur Datenvalidierung in einem interdisziplinären Team aus Biometrikern und Medizinerinnen erfolgen, da hierfür sowohl medizinische als auch methodische Fachkenntnis erforderlich ist. Wie in Kapitel 5.2 beschrieben wird, sollen für die Entwicklung der unterschiedlichen Auffälligkeitskriterien verschiedene Datenbanken genutzt werden. Da zum Zeitpunkt der Entwicklung der Auffälligkeitskriterien (Oktober des Erfassungsjahres) die Daten, für die eine Validierung erfolgen soll, noch nicht vorliegen, muss die Entwicklung der Auffälligkeitskriterien anhand der Daten des Vorjahres erfolgen. Für die Entwicklung der unterschiedlichen Auffälligkeitskriterien sollen die folgenden Daten (des Vorjahres) zur Anwendung kommen. Damit Auffälligkeitskriterien zur Vollzähligkeit entwickelt werden können, ist es erforderlich, dass die Daten zur Soll-Ist-Statistik pseudonymisiert für jeden Leistungserbringer von den Datenannahmestellen an die Institution nach § 137a SGB V übermittelt werden.

Nach dem Beschluss des G-BA bezüglich der Auffälligkeitskriterien für die statistische Basisprüfung und der Datenfelder für die Stichprobenprüfung, führt die Institution nach § 137a SGB V auf Basis der Daten, für die auch tatsächlich eine Datenvalidierung erfolgen soll, eine Stichprobenziehung für die Zweiterfassung durch und ermittelt die rechnerischen Auffälligkeiten anhand der Auffälligkeitskriterien für die statistische Basisprüfung.

Im Falle eines landesspezifischen Verfahrens teilt die Institution nach § 137a SGB V den LAGen mit, welche Leistungserbringer im Rahmen der Auffälligkeitsprüfung auffällig geworden sind und welche Leistungserbringer und Patienten im Rahmen der Zufallsstichprobe gezogen wurden. Zudem stellt sie Spezifikationen zum Übermittlungsformat der Daten aus der Zweiterfassung und den Ergebnissen des Stellungnahmeverfahrens, das im Rahmen der Auffälligkeitsprüfung geführt wurde, bereit.

Um ein transparentes und einheitliches Vorgehen zu garantieren, empfehlen wir zum einen die Nutzung eines **Erfassungstools**, das durch die Institution nach § 137a SGB V entwickelt wird. Zum anderen sind **Handbücher für das Stellungnahmeverfahren** in Planung. Wir würden die Parameter dieses Verfahrens – mit geringen Modifikationen – auf den Ausführungen im Handbuch zum Datenvalidierungsverfahren (BQS, 2009; AQUA, 2010: 14-19) basieren. Analoges ist auch für die sektorenübergreifende QS im ambulanten Sektor geplant und wird im Zuge der konkreten Planung der ersten sektorenübergreifenden Datenvalidierungsverfahren entwickelt werden.

Tabelle 11: Zeitplan vorbereitende Maßnahmen

Zeitpunkt	Aktion	Akteure
August des Vorjahres	Vorschlag der QS-Verfahren für die Datenvalidierung im Folgejahr	Institution nach § 137a SGB V in Abstimmung mit den an der Durchführung beteiligten Institutionen
Oktober des Vorjahres	Beschluss der QS-Verfahren für die Datenvalidierung im Folgejahr	G-BA
von Oktober bis Dezember des Vorjahres	Entwicklung von Auffälligkeitskriterien für die statistische Basisprüfung anhand der Daten des Vorjahres Auswahl von Datenfeldern für die Stichprobenprüfung	Institution nach § 137a SGB V in Abstimmung mit den an der Durchführung beteiligten Institutionen
Dezember des Vorjahres	Vorschlag der Auffälligkeitskriterien und Datenfelder	Institution nach § 137a SGB V in Abstimmung mit den an der Durchführung beteiligten Institutionen
Januar/Februar	Beschluss der Auffälligkeitskriterien und Datenfelder für die Datenvalidierung	G-BA
bis April	Übermittlung der rechnerischen Auffälligkeiten an die LAGen Bereitstellung der Spezifikationen zum Übermittlungsformat der Daten aus dem Datenvalidierungsverfahren	Institution nach § 137a SGB V

5.2. Durchführung

Die praktische Durchführung des Datenvalidierungsverfahrens erfolgt für die länderbezogenen Verfahren durch die LAGen und für die bundesbezogene Verfahren durch die Bundesauswertungsstelle. Weitere einzubeziehende Akteure sind die Leistungserbringer selbst, aber auch die in § 1 (5) genannten Akteure, also KVen, KZVen, LQS, LKGen und LAGen.

Auffälligkeitsprüfung: Im Rahmen der Auffälligkeitsprüfung werden die Leistungserbringer, bei denen sich bei der Anwendung der Auffälligkeitskriterien rechnerische Auffälligkeiten gezeigt haben, um eine Stellungnahme gebeten. Aus der Stellungnahme sollte hervorgehen, inwieweit es sich bei den ermittelten Auffälligkeiten tatsächlich um fehlerhafte Dokumentationen handelt und welche Verbesserungsmaßnahmen gegebenenfalls ergriffen wurden oder geplant sind. Wenn ein Leistungserbringer, bei dem die statistische Basisprüfung rechnerische Auffälligkeiten in der Dokumentationsqualität gezeigt hat, auch im Rahmen der Zufallsstichprobe gezogen worden ist, sollen die Ergebnisse der Zweiterfassung in das Stellungsverfahren integriert werden.

Stichprobenprüfung: Bei den Leistungserbringern, die im Rahmen der Zufallsstichprobe gezogen worden sind, erfolgt eine Zweiterfassung der ausgewählten Datenfelder. Wenn sich im Rahmen der Zweiterfassung Auffälligkeiten in der Dokumentationsqualität zeigen, kann auch hier ein Stellungsverfahren eingeleitet werden.

Nach Abschluss des Datenvalidierungsverfahrens müssen die Ergebnisse des Stellungsverfahrens und die Ergebnisse des Stichprobenverfahrens entsprechend der Vorgaben in den Spezifikationen an die Institution nach § 137a SGB V übersendet werden.

5.3. Umgang mit Leistungserbringern mit statistischen Auffälligkeiten

Ein zentrales Element des Datenvalidierungsverfahrens besteht in der Aufklärung von statistischen Auffälligkeiten, die bei einzelnen Leistungserbringern in Bezug auf die Dokumentationsqualität ermittelt werden. Wie bereits in Kapitel 4 dargelegt, können im Rahmen des Datenvalidierungsverfahrens grundsätzlich zwei unterschiedliche Arten von Auffälligkeiten auftreten:

- Auffälligkeit des Leistungserbringers in der Auffälligkeitsprüfung;
- Auffälligkeit des Leistungserbringers in der Stichprobenprüfung.

Eine Auffälligkeit bei der Auffälligkeitsprüfung ergibt sich durch die Abweichung eines Leistungserbringers vom Referenzbereich eines Auffälligkeitskriteriums. Mit der Abweichung eines Leistungserbringers vom Referenzbereich ist die Hypothese verbunden, dass der Grund für die Abweichung eine fehlerhafte Dokumentation ist. Die Auffälligkeit eines Leistungserbringers im Rahmen der Stichprobenprüfung kann sich durch hohe Abweichungsraten zwischen den Angaben in der Patientenakte und der QS-Dokumentation in einem oder mehreren Datenfeldern ergeben.

Unabhängig von der Art der Auffälligkeit, besteht auf Seiten der Leistungserbringer gemäß § 16 Abs. 3 der Qesü-RL die folgende Verpflichtung zur Mitwirkung an der Aufklärung der statistischen Auffälligkeit:

„[...] Bei Auffälligkeiten haben sie [die Leistungserbringer; Anm. d. Red.] auf Nachfrage der datenvalidierenden Stelle an der Aufklärung und Beseitigung der Ursachen mitzuwirken. Soweit dies nicht erfolgt, kommen die Maßnahmen nach § 17 entsprechend zur Anwendung.“

Vor diesem Hintergrund wird ein Vorgehen erörtert, wie mit den Leistungserbringern, die statistische Auffälligkeiten in Bezug auf die Dokumentationsqualität aufweisen, umgegangen werden soll. Hierbei stellt ein zielgerichteter Dialog zwischen den datenvalidierenden Stellen und den Leistungserbringern ein zentrales Element dar, mit dem konkrete Ziele verfolgt werden:

1. Sensibilisierung der Leistungserbringer über ein Feedback zu ermittelten Auffälligkeiten;
2. Recherche der zugrunde liegenden Ursachen für Auffälligkeiten;

3. Eintreten in einen Verbesserungsprozess durch Beratung der überprüften Einrichtungen und ggf. Vereinbarung von Maßnahmen.

Zur Umsetzung der Ziele werden im Folgenden konkrete Maßnahmen beschrieben. Um den Dialog zwischen den datenvalidierenden Stellen und den Leistungserbringern zu systematisieren und für Dritte nachvollziehbar zu machen, basieren die Maßnahmen auf der Prämisse, dass dieser in schriftlicher Form (Stellungnahmeverfahren) geführt wird; so wie es in § 17 Abs. 2 der Qesü-RL beschrieben wird.

Zu (1.) Sensibilisierung der Leistungserbringer über ein Feedback zu ermittelten Auffälligkeiten

Die Leistungserbringer werden schriftlich über die ermittelten Auffälligkeiten informiert. Das Anschreiben sollte neben der Benennung auch eine Erläuterung der Relevanz der jeweiligen rechnerischen Auffälligkeit enthalten. Im Falle einer Auffälligkeit als Abweichung vom Referenzbereich eines Auffälligkeitskriteriums wird zudem die Hypothese mitgeteilt, warum eine fehlerhafte Dokumentation vermutet wird. Zusätzlich werden die Leistungserbringer darüber in Kenntnis gesetzt, welche Auswirkung eine etwaige Fehldokumentation auf die Berechnung der Qualitätsindikatoren haben könnte.

Zu (2.) Recherche der Ursachen für Auffälligkeiten

Im Falle einer Auffälligkeit, die sich aus der Abweichung vom Referenzbereich eines Auffälligkeitskriteriums ergibt, werden die Leistungserbringer im Anschreiben darum gebeten, die Dokumentation der auffälligen Fälle zu überprüfen und das Ergebnis der Überprüfung der datenvalidierenden Stelle schriftlich mitzuteilen. Zu diesem Zweck wird den Leistungserbringern eine Liste mit den Fällen übermittelt, die sich bei der Anwendung des Auffälligkeitskriteriums als auffällig erwiesen haben und bei denen eine fehlerhafte Dokumentation zu vermuten ist. Das Ziel der Überprüfung der Dokumentation besteht darin, dass die Leistungserbringer für die entsprechenden Datenfelder (z. B. anhand der Patientenakte) überprüfen, ob die Dokumentation in der QS-Dokumentation tatsächlich korrekt ist. Falls sich bei der Überprüfung der Dokumentation zu den auffälligen Fällen fehlerhafte Dokumentationen zeigen, werden die Leistungserbringer zudem dazu aufgefordert, die Ursachen hierfür zu eruieren und das Ergebnis auch den datenvalidierenden Stellen mitzuteilen.

Wenn sich im Rahmen des Stichprobenverfahrens eine Auffälligkeit aufgrund einer hohen Abweichungsrate zwischen der QS-Dokumentation und den Angaben in der Patientenakte ergibt, dann werden die Leistungserbringer ebenfalls im Anschreiben darum gebeten, die jeweiligen Ursachen zu eruieren und den datenvalidierenden Stellen schriftlich mitzuteilen.

Zu (3.) Eintreten in einen Verbesserungsprozess durch Beratung der überprüften Einrichtungen und ggf. Vereinbarung von Maßnahmen

Letztlich werden die Leistungserbringer im Rahmen des Anschreibens auch darum gebeten, eigene Vorschläge zur Optimierung der Dokumentation zu machen bzw. Maßnahmen zu benennen, die sie bereits realisiert haben. Das umfassende Antwortschreiben der Leistungserbringer wird zuständigen Fachexperten zur Diskussion und Bewertung vorgelegt. Sind die Erläuterungen zu den statistischen Auffälligkeiten für die Fachexperten nachvollziehbar, werden die Leistungserbringer darüber informiert und der Dialog kann als abgeschlossen betrachtet werden. Falls das Antwortschreiben jedoch Fragen offen lässt bzw. weiterer Klärungsbedarf besteht, werden die entsprechenden Leistungserbringer um Präzisierung gebeten, es wird das persönliche Gespräch gesucht oder es erfolgt eine Begehung der entsprechenden Einrichtung (vgl. § 17 Abs. 2 der Qesü-RL). Das grundsätzliche Ziel des intensivierten Dialogs mit den Leistungserbringern besteht darin, bei einem nachweislichen Mangel in der Dokumentationsqualität entsprechende Optimierungsmaßnahmen zu initiieren. Können die Auffälligkeiten im Dialog nicht geklärt werden (Handelt es sich bei der Auffälligkeit um eine Fehldokumentation?), können gem. § 17 Abs. 3 der Qesü-RL konkrete Maßnahmen mit dem Leistungserbringer vereinbart werden, deren Ziel darin besteht, die Unklarheiten zu beseitigen (Maßnahmestufe 1). Wenn sich Belege für schwerwiegende Missstände in Bezug auf die Dokumentation finden oder ein Leistungserbringer sich verweigert, bestimmte Zielvereinbarungen in der vorgegebenen Zeit zu erfüllen, ist ihm die Möglichkeit zur schriftlichen Stellungnahme mit einer 4-monatigen Frist einzuräumen. Bei Ausbleiben der schriftlichen Stellungnahme wird gem. § 17 Abs. 4 der Qesü-RL die Maßnahmestufe 2 eingeleitet, die die Korrektur der Vereinbarung oder die Information der für Vergütungsabschläge oder Entziehung der Abrechnungsmöglichkeit der jeweiligen Leistung zuständigen Stellen, mit entsprechenden Empfehlungen, zur Folge haben kann.

Nicht unerwähnt bleiben soll, dass es notwendig erscheint, länderübergreifend einheitliche Klassifikationen der behandelten Fälle zu etablieren. Hier besteht Abstimmungsbedarf zwischen den Ländern und dem AQUA-Institut. Insgesamt stellt das hier dargelegte Vorgehen zum Umgang mit Leistungserbringern mit statistischen Auffälligkeiten in Bezug auf die Dokumentationsqualität ein effektives Vorgehen dar, um die Ursachen von Auffälligkeiten zu eruieren und entsprechende Optimierungsmaßnahmen anzustoßen. Zudem bieten die unterschiedlichen Eskalationsstufen (Maßnahmestufen 1 und 2) die Möglichkeit, dem Dialog zur Dokumentationsqualität den notwendigen Nachdruck zu verleihen.

5.4. Evaluationen

An die Durchführung des Datenvalidierungskonzeptes schließt sich eine doppelte Evaluation an. Zum einen sollen die Ergebnisse der Prüfungen mit den QS-Verfahren rückgekoppelt werden. Zum anderen soll das Datenvalidierungskonzept selbstreferentiell optimiert werden.

5.4.1. Evaluation der QS-Verfahren

Zur Evaluation der Dokumentationsqualität innerhalb der QS-Verfahren, sowie um die Erkenntnisse, die im Rahmen des Verfahrens auf Landesebene gesammelt worden sind, systematisch zu erfassen und transparent zu machen, soll nach der Durchführung eines Datenvalidierungsverfahrens eine Befragung der LAGen durchgeführt werden.

Bei den bundesbezogenen Verfahren soll neben den nach §1 Abs. 5 der Qesü-RL mandatierten Akteuren, also der Vertrauensstelle und der Bundesdatenauswertungsstelle, auch auf die Fachkompetenz der relevanten Organisationen (Bundes(zahn)ärztekammer, Vertreter der Versicherungsträger, Patientenvertretungen und relevante Landes- und Wissenschaftsorganisationen) hinzugezogen werden.

Hierbei sollen insbesondere die folgenden Fragen behandelt werden:

- Welche Konsequenzen lassen sich aus den Ergebnissen des Datenvalidierungsverfahrens für die Aussagekraft der Qualitätsindikatoren ableiten?
- Wurden aufgrund der Ergebnisse des Datenvalidierungsverfahrens auf Ebene der Leistungserbringer, der LAGen oder der Bundesebene Maßnahmen zur Optimierung der Dokumentationsqualität eingeleitet (z. B. Anpassungen der Ausfüllhinweise oder der Plausibilitätsregeln)?

Neben den Ergebnissen des Datenvalidierungsverfahrens selbst soll hierzu eine Befragung der LAGen als Grundlage dienen.

5.4.2. Evaluation des Datenvalidierungsverfahrens

Das Datenvalidierungsverfahren sollte einem iterativen PDCA-Zyklus (Deming 2000) unterworfen werden. Im Rahmen dieser Evaluationsplanung muss eine Abwägung zwischen Aufwand und Nutzen besondere Beachtung finden. Dazu schlagen wir Schwerpunktsetzungen nach Verfahren sowie einen verstärkten Rückgriff auf Routinedaten vor.

Das Gesamtergebnis der Evaluation muss sich an dem Erfüllungsgrad der eingangs formulierten Ziele orientieren. Im Vordergrund stehen dabei folgende Fragen:

- Welche Auffälligkeitskriterien, die im Rahmen der Auffälligkeitsprüfung zur Anwendung kamen, haben sich in der praktischen Umsetzung als besonders geeignet herausgestellt (Verstetigung)?
- Hat sich das Datenvalidierungsverfahren als geeignet erwiesen, die Dokumentationsqualität der QS-Daten zu überprüfen? Gibt es Optimierungsvorschläge?

Zur Evaluation der Eignung des Datenvalidierungsverfahrens insgesamt bieten sich grundsätzlich sehr verschiedene Methoden an:

- Vergleich der QS-Dokumentationen mit Routinedaten als Referenzquelle;
- eine Longitudinalbeobachtung der Ergebnisse einzelner QS-Verfahren;
- statistische Überprüfungen;
- externe Überprüfungen durch eine stichprobenartige Befragung der Studienteilnehmer vor Ort;
- externe Überprüfungen durch Befragungen der LAGen.

Vergleich der QS-Dokumentationen mit Routinedaten als Referenzquelle

Neben der Patientenakte (siehe Stichprobenprüfung) sollen auch Routinedaten als Referenzquelle für Datenprüfungen angewendet. Ein Beispiel sind Prüfungen der Vollzähligkeit der QS-Daten anhand von Routinedaten aus den jeweiligen Sektoren. Zur Prüfung der Daten im stationären Bereich könnte ein Abgleich der QS-Daten mit den §21-Daten gemäß KHEntgG des InEK erfolgen. Zur Prüfung der Daten im ambulanten Bereich könnten Abrechnungsdaten aus dem kassenärztlichen und kassenzahnärztlichen Bereich genutzt werden. Allerdings besteht derzeit nur hinsichtlich der §21-Daten eine gesetzliche Regelung der Nutzung für Zwecke der Qualitätssicherung. Hier besteht unseres Erachtens ein gesetzlicher Regelungsbedarf. Gesetzlicher Regelungsbedarf besteht auch für den selektivvertraglichen Kontext. Um dort eine Evaluation auf Basis von Routinedaten zu ermöglichen, müsste für die Verträge einheitlich festgelegt werden, welche Informationen im Rahmen der Abrechnung mindestens bereitzustellen sind.

Die Routinedaten werden zunächst nur als Grundlage für eine Evaluation des Datenvalidierungsverfahrens verwendet, weil in einzelnen Fällen merkliche Abweichungen zwischen Routinedaten und QS-Dokumentationen bekannt sind. Anschließend soll entschieden werden, ob und wie sich diese in den Prozess des Datenvalidierungsverfahrens (z. B. durch Kreuzvalidierungen) integrieren lassen.

Longitudinalbeobachtungen und statistische Prüfungen

Da Datenvalidierungsverfahren kein Selbstzweck sind, wird der Erfolg des Programms vor allem an nachweisbaren Verbesserungen der Dokumentationsqualität zu messen sein. Dazu gibt es eine Reihe gut erprobter Analysemethoden für Zeitreihenanalysen – teilweise sogar für Untersuchungen, die nicht beständig wiederholt werden (Wagner et al. 2002) –, die in langjährig laufenden Studien solche Entwicklungen untersuchen können. Zudem kann man mithilfe des Vergleichs der Prüfkriteriendaten in den Auffälligkeitsfällen mit den in anderen Prüfungsphasen gewonnenen Erkenntnissen abschätzen, welche Indikatoren besonders brauchbar für die Entdeckung von Problemen sind. Schließlich kann auch mit einfachen statistischen Verfahren die Diskriminationsfähigkeit der verschiedenen Prüfkriterien abgeschätzt werden.

Befragungen

Eine zweite Evaluationsmethode stellt die Befragung relevanter Akteure dar. Hierzu sollen die LAGen schriftlich befragt werden und ausgewählte Leistungserbringer vor Ort gehört werden.

Da die Leistungserbringer über das größte empirische Wissen in Hinblick auf den tatsächlichen Verfahrensablauf verfügen, bietet es sich grundsätzlich an, ausgewählte Leistungserbringer in Fokusgruppen (Kitzinger 1995) in die Evaluation einzubeziehen. Zunächst soll dieses Vorgehen jedoch nicht generell in die routinemäßige Evaluation einbezogen werden, weil der notwendige Zeitbedarf noch geklärt werden muss und die Integration in die übrigen Elemente der Evaluation noch offen ist.

Die Ergebnisse der Befragung sollen dann gemeinsam mit den Ergebnissen zur Auffälligkeitsprüfung im Stellanahmeverfahren und den Ergebnissen des Stichprobenverfahrens mit Datenabgleich an die Institution nach § 137a SGB V übersandt werden. Die Institution nach § 137a SGB V führt die Ergebnisse zusammen, analysiert und bewertet sie und erstellt daraus einen zusammenfassenden Bericht.

Prozessplan

Evaluationen sollen jährlich durchgeführt werden. Sie setzen sich zusammen aus einem mindestens durchzuführendem Teil und spezifischen, zusätzlichen Evaluationen. Die zusätzlichen Evaluationen werden jeweils vorab, im Zuge der Empfehlungen zu den auszuwählenden Leistungsbereichen der Datenvalidierung, beschreiben. Jede jährliche Evaluation umfasst mindestens eine Analyse der Aussagekraft der Qualitätsindikatoren, Vorschläge für zu verstetigende Auffälligkeitskriterien und Aussagen dazu, welche Verbesserungen im QS-Verfahren eingeleitet werden sollen. Ferner soll eine schriftliche Befragung der LAGen ein fester Bestandteil der Evaluation werden. Tabelle 12 fasst einen möglichen Zeitplan für die regelmäßigen Evaluationen zusammen.

Tabelle 12: Zeitplan Evaluation und Bericht

Zeitpunkt	Aktion	Akteur
Januar/Februar	Evaluation des Verfahrens: schriftliche Befragung der LAG	LAGen
bis 15. März	Übersendung der Daten zum Stellungnahmeverfahren, zur Zweiterfassung und zur Befragung (Evaluation) an die Institution nach § 137a SGB V	LAGen
von Mitte März bis Mitte Mai	Zusammenführung, Analyse und Bewertung der Ergebnisse	Institution nach § 137a SGB V
15. Mai	Berichterstattung Ergebnisse des Datenvalidierungsverfahrens Ergebnisse der Evaluation des Datenvalidierungsverfahrens	Institution nach § 137a SGB V

Die oben genannten weiteren Evaluationselemente müssen jeweils differenziert vor dem Hintergrund des jeweiligen QS-Verfahrens gesehen werden. Beispielsweise sind Longitudinalanalysen erst ab dem 2. Jahr möglich.

5.4.3. Berichtswesen

Kernstück der öffentlichkeitswirksamen Umsetzung der Ergebnisse der Evaluation ist das Berichtswesen. Dabei schlagen wir vor, dass die Bundesauswertungsstelle je einen Bericht zur Datenvalidierung selbst und zu den Empfehlungen zur Optimierung der betroffenen QS-Verfahren vornimmt. Dabei soll konkret auf die in den zwei vorangegangenen Unterkapiteln aufgeworfenen Fragen eingegangen werden. Zusätzlich empfehlen wir den LQS in den länderbezogenen Verfahren eigene Berichte dieser Art zu produzieren. Tabelle 12 zeigt einen grundlegenden möglichen Zeitplan. Zu beachten ist dabei, dass einige besondere Evaluationsinstrumente (z. B. Vergleich mit Routinedaten) zwangsläufig in separate Berichtsteile ausgegliedert werden müssen, weil die entsprechenden Daten erst später verfügbar sind.

Die Berichte selber sollen die in unserem Methodenpapier (AQUA 2010a) ausgeführten Zielgruppen – G-BA, Fachpublikum und interessierte Öffentlichkeit – erreichen. Um letztere zu informieren, bietet sich unter Umständen ein *Digest* der Berichte an, der besonders einfach interpretierbare, aber handlungsrelevante Informationen auswählt.

Diese Berichte werden dem G-BA übermittelt, der geeignete Handlungsdirektiven entwickeln sollte, um die gewonnenen Erkenntnisse umzusetzen.

6. Empfehlungen und Ausblick

Auf Basis des vorliegenden Berichts empfiehlt das AQUA-Institut ein breiteres, ganzheitliches Konzept zur Datenvalidierung als dieses vom G-BA ursprünglich angedacht war. Dabei bleiben notwendigerweise einige Fragen offen, da die gesetzlichen Grundlagen in Teilen des Prozesses noch zu unbestimmt sind und auch noch keine konkreten QS-Verfahren benannt worden sind. Bereits sehr konkret sind jedoch unsere Vorschläge zur Gestaltung der Prüfinstrumente. Dabei empfiehlt das AQUA-Institut, die in Kapitel 2 enumerierten Prüfkriterien in einem mehrstufigen Prozess anzuwenden. In einer ersten Stufe werden die gepoolten Daten einer Basisprüfung unterworfen. Teilweise darauf aufbauend (nämlich in der Konstruktion der Stichprobe) sollen in einem Stichprobenverfahren weitere Erkenntnisse über die Dokumentationsqualität gewonnen werden. Auffällig gewordene Leistungserbringer werden zur Qualitätssicherung in einen systematischen Stellungnahmeprozess integriert. Schließlich wird das Verfahren sowohl formal als auch inhaltlich evaluiert, und die Erkenntnisse werden in zukünftige QS- und Datenvalidierungsverfahren eingebunden.

Aufgrund der Neuheit dieses Konzepts besteht ein erhöhter Erprobungsbedarf vor allem mit Blick auf einige Unwägbarkeiten, die sich erst im Laufe der Anwendung lassen werden. So sind u. a. folgende Aspekte offen:

- Das praktische zur Verfügung stellen fehlender Routinedaten.
- Die verfahrensabhängigen Datenpools konnten noch nicht gesichtet werden, weil die Auswahl der Verfahren noch nicht getroffen ist.
- Das Zusammenspiel der verschiedenen Akteure, vor allem im logistischen Bereich, muss noch erprobt werden.
- Möglichkeiten, die sich durch die Einführung der elektronischen Gesundheitskarte ergeben könnten, wurden noch nicht diskutiert.

Trotz der genannten Unsicherheiten sind wir der Überzeugung, dass wir mit dem diesem Konzept zwar Neuland betreten, aber eine solide Grundlage für die Überprüfung der Dokumentationsqualität in der sektorenübergreifenden Qualitätssicherung geschaffen haben.

Literaturverzeichnis

ANZICS. Proposed Outlier Management Plan. Australian and New Zealand Intensive Care Society: Centre for Outcome and Resource Evaluation (ANZICS CORE). 2008.

AQUA. Allgemeine Methoden im Rahmen der sektorenübergreifenden Qualitätssicherung im Gesundheitswesen nach § 137a SGB V. Göttingen: AQUA 2010a.

AQUA. Datenvalidierung 2009 zum Erfassungsjahr 2008. Zusammenfassender Bericht. Dekubitusprophylaxe, Herzschrittmacher-Implantation, Mammachirurgie. Stand: 14.05.2010. Göttingen: AQUA-Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen 2010b.

AQUA. Qualitätsreport 2009. Göttingen: AQUA-Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen 2010c.

AQUA. Technische Dokumentation zur QS-Spezifikation für QS-Dokumentationssoftware 14.0 zur Datenerfassung, Plausibilitätsprüfung und Datenübermittlung im Rahmen der externen vergleichenden Qualitätssicherung. Stand 30. Juni 2010. Göttingen: AQUA-Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen 2010d.

Aronsky D, Haug PJ. Assessing the quality of clinical data in a computer-based record for calculating the pneumonia severity index

4. J Am Med Inform Assoc. 2000; 7(1):55-65.

Arts DG, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc. 2002; 9(6):600-611.

Asch SM, McGlynn EA, Hogan MM, Hayward RA, Shekelle P, Rubenstein L, Keesey J, Adams J, Kerr EA. Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. Ann Intern Med. 2004; 141(12):938-945.

Batini, C, Scannapieca M. Data Quality: Concepts, Methodologies and Techniques. Berlin: Springer 2006.

BQS. Handbuch zum Datenvalidierungsverfahren, Version 6. Düsseldorf: BQS 2009.

Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. Qual Saf Health Care. 2002; 11(4):358-364.

Chin R, Lee BY. Section V. Analysis of Results. In: Chin R et al. (eds). Principles and Practice of Clinical Trial Medicine. Academic Press 2008:302-323.

David PA. Path Dependence and the Quest for Historical Economics: One More Chorus of the Ballad of Qwerty. (=Discussion Papers in Economic and Social History). University of Oxford 1997.

Deming, WE. Out of the crisis. 1st MIT Press ed, Cambridge, Mass: MIT Press 2000.

Djelic M-L, Quack S. Overcoming path dependency: path generation in open systems. Theor Soc. 2007; 36:161-186.

Farquhar M. AHRQ Quality Indicators. In: Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Agency for Healthcare Research and Quality 2008:

Fink AS, Itani KM, Campbell DC, Jr. Assessing the quality of surgical care. Surg Clin North Am. 2007; 87(4):837-52, vi.

Goldstone JA. Initial conditions, general laws, path dependence, and explanation in historical sociology. The American Journal of Sociology. 1998; 104(3):829-845.

Gribben B, Coster G, Pringle M, Simon J. Non-invasive methods for measuring data quality in general practice. The New Zealand Medical Journal. 2001; 114(1125):30-32.

Hasford, J, Staib AH. Arzneimittelprüfungen und Good Clinical Practice. MVV. Medizin Verlag München. München: MMV Medizin Verlag 1994.

- Helsedirektoratet. Rammeverk for et kvalitetsindikatorsystem i helsetjenesten. Oslo: Helsedirektoratet 2010.
- IOM. Future Directions for the National Healthcare. Quality and Disparities Reports. Washington, DC: Institute of Medicine, The National Academies Press 2010.
- Kitzinger J. Qualitative Research: Introducing focus groups. *BMJ*. 1995; 311(7000):299-302.
- Klein J. Report on the Use of Medi-Cal Managed Care Encounter Data for Research Purposes. Sacramento: California Department of Health Services 2002.
- Körner T, Schmidt J. Qualitätsberichte - ein neues Informationsangebot über Methoden, Definitionen und Datenqualität der Bundesstatistiken. *Wirtschaft und Statistik*. 2006; 2:109-117.
- Lee, ES, Forthofer RN. Analyzing complex survey data. 2nd ed, Thousand Oaks, Calif: Sage Publications 2006.
- Leidl R. Der Effizienz auf der Spur: Eine Einführung in die ökonomische Evaluation. In: Schwartz FW et al. (eds). *Das Public Health Buch: Gesundheit und Gesundheitswesen*. 2, München: Urban & Fischer 2003:461-484.
- Maydanchik, A. Data Quality Assessment. Technics Publications, LLC 2007.
- Mc Colm D. Grant Final Report [HS 17094]: Standardization and Automatic Extraction of Quality Measures in an Ambulatory EHR. Rockville: The Agency for Healthcare Research and Quality (AHRQ) 2009.
- McKenzie K, Walker S, Besenyei A, Aitken LM, Allison B. Assessing the concordance of trauma registry data and hospital records. *HIM J*. 2005; 34(1):3-7.
- Mosbech J, Jorgensen J, Madsen M, Rostgaard K, Thornberg K, Poulsen T. Landspatientregisteret. Evaluering af datakvaliteten. *Ugeskrift for Læger*. 1995; 157(26):3741-3745.
- NHS. DoCDat Data Collection Manual. London: The Information Centre for Health and Social Care, National Health Service 2007.
- NHS. Data Quality Report for Independent Sector NHS funded treatment Q4 2008/09. Version 1.0. London: The Information Centre, National Health Service 2009.
- Nonnemacher, M, Weiland D, Stausberg J. Datenqualität in der medizinischen Forschung: Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft 2007.
- Olsen, JE. Data quality: the accuracy dimension. San Francisco: Morgan Kaufman Publishers 2003.
- Rand Health. Evaluation of the Use of AHRQ and Other Quality Indicators. Agency for Healthcare Research and Quality 2007.
- Ratzel, R, Luxenburger B. *Handbuch Medizinrecht*. Bonn: Deutscher Anwaltverlag 2008.
- Schnell, R, Hill P, Esser E. *Methoden der empirischen Sozialforschung*. 8, München: Oldenbourg-Verlag 2008.
- Shortliffe, EH, Perreault LE, Wiederhold G, Fagan LM. *Medical informatics computer applications in health care*. Reading, Mass: Addison-Wesley Pub. Co 1990.
- Singh, S. *Advanced sampling theory with applications: how Michael 'selected' Amy*. Dordrecht: Kluwer Academic Publishers 2003.
- Socialstyrelsen. Bilaga 2. Täckningsgradsjämförelser mellan hälsodataregistren vid Socialstyrelsen och Nationella Kvalitetsregister. Stockholm: Sveriges Kommuner och Landsting 2010.
- Sørensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol*. 1996; 25(2):435-442.
- Stange KC, Zyzanski SJ, Smith TF, Kelly R, Langa DM, Flocke SA, Jaen CR. How valid are medical records and patient questionnaires for physician profiling and health services research? A comparison with direct observation of patients visits. *Med Care*. 1998; 36(6):851-867.
- Stausberg J. Die Kodierqualität in der stationären Versorgung. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2007; 50(8):1039-1046.

Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing paper-based with electronic patient records: lessons learned during a study on diagnosis and procedure codes. *J Am Med Inform Assoc.* 2003; 10(5):470-477.

Stausberg J, Nonnemacher M, Weiland D, Antony G, Neuhäuser M. Management of Data Quality - Development of a Computer-Mediated Guideline. In: Hasman A et al. (eds). *Ubiquity: Technologies for Better Health in Aging Societies. Proceedings of MIE2006.* Amsterdam: IOS Press 2006:

Supekar K, Marwadi A, Lee Y, Medhi D. Fuzzy Rule-Based Framework for Medical Record Validation. In: Yin H et al. (eds). *Intelligent Data Engineering and Automated Learning - IDEAL 2002.* 2412, Berlin/Heidelberg: Springer 2002:1-27.

Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics.* 2002; 27(4):299-309.

Webster AC, Supramaniam R, O'Connell DL, Chapman JR, Craig JC. Validity of registry data: agreement between cancer records in an end-stage kidney disease registry (voluntary reporting) and a cancer register (statutory reporting). *Nephrology (Carlton).* 2010; 15(4):491-501.

Willis CD, Jolley DJ, McNeil JJ, Cameron PA, Phillips LE. Identifying and improving unreliable items in registries through data auditing. *Int J Qual Health Care.* 2011.

Winkler WE. Methods for evaluating and creating data quality. *Information Systems.* 2004; 29(7):531-550.

Zaiß A, Graubner B, Ingenerf J, Leiner F, Lochmann U, Schopen M, Schrader U, Schulz S. Medizinische Dokumentation, Terminologie und Linguistik. In: Lehmann T et al. (eds). *Handbuch der Medizinischen Informatik.* München, Wien: Carl Hanser Verlag 2002:45-102.