

Anhang 2 – Zusammenfassung von Literatursichtung und Expertenbefragung

Version 1.1, 03.03.2006

Fortschreibung von Version 1.0 vom 27.10.2005

1 Einleitung

Der erste Schritt bei der Entwicklung der Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern war die Identifikation und Analyse der bestverfügbaren Evidenz über eine Sichtung der Literatur und eine Befragung von Experten. Empfehlungen zum Vorgehen bei der Entwicklung von Leitlinien wurden national mehrfach veröffentlicht [AWMF/ÄZQ 2001] und wissenschaftlich begründet [Bundesärztekammer/Kassenärztliche Bundesvereinigung 2005]. Das Vorgehen bei der Literatursichtung entspricht den Schritten zur Quellensuche und Quellenbewertung, wie es die BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH für die externe vergleichende Qualitätssicherung nach § 137 SGB V beschreibt [Mohr et al. 2005].

Unabhängig von den hier gewonnenen Ergebnissen wurde festgelegt, durchgehend das Glossar [GMDS-AG Qualitätsmanagement in der Medizin 2003] der Arbeitsgruppe Qualitätsmanagement in der Medizin der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) als Referenz für die Definition qualitätsrelevanter Begriffe zu Grunde zu legen. Dieses Glossar findet sowohl in der Fachöffentlichkeit als auch in der Praxis durchgehend Anerkennung und kann derzeit in Deutschland als De-facto-Standard gelten.

Das 1996 erschienene Glossar zu Begriffen und Konzepten des Qualitätsmanagements wurde 2003 grundlegend überarbeitet und aktualisiert. Es versteht sich als Zusammenstellung grundlegender Begriffe mit Verweisen auf weiterführende Literatur, Institutionen, Fachgesellschaften und Internet-Adressen vor dem Hintergrund einer zunehmenden Bedeutung des Themas Qualitätsmanagement im Gesundheitswesen und soll ein einheitliches

Verständnis der Qualitätsbegriffe fördern. Die im Glossar enthaltenen Begriffe und Konzepte gliedern sich inhaltlich in die Teile:

- Qualitätsbezogene Begriffe: Qualität, Merkmal, Qualitätsmerkmal, Anforderung, Fähigkeit, Anspruchsklasse, Kunde, Lieferant, Prozess, Produkt, Struktur-/Prozess-/Ergebnisqualität, Standard, Richtlinie, Leitlinie, Empfehlung
- (Qualitäts)bewertungsbezogene Begriffe: Bewertung, Wirksamkeit, Effizienz, Prüfung/Inspektion, Qualitätsindikator, Konformität, Fehler/Mangel, Audit, Peer review, Zertifizierung, Akkreditierung
- Qualitätsmanagementbezogene Begriffe: Qualitätsmanagement, Qualitätspolitik, Qualitätsziele, Qualitätsplanung, Qualitätslenkung, Qualitätskontrolle, Qualitätssicherung, Qualitätsverbesserung, Ständige Verbesserung, Qualitätszirkel, Bewertung des Qualitätsmanagementsystems, Qualitätsmanagementhandbuch, Umfassendes Qualitätsmanagement, Benchmarking

2 Literatursichtung

2.1 Einführung

Eine systematische Sichtung der Literatur steht am Anfang der Leitlinienentwicklung, um die verfügbare Evidenz aufzufinden und zu analysieren.

2.2 Material und Methoden

Als relevante Bereiche zur Literatursichtung wurden Qualitätsmessung in Medizin und Industrie, Qualitätsindikatoren, Statistische Me-

thoden in der Qualitätssicherung, Qualitätslenkung, Ringversuche sowie Richtlinien/Empfehlungen für die Erstellung und den Inhalt medizinischer Leitlinien identifiziert. Folgende Quellen wurden in die Recherche aufgenommen.

Suchmaschinen:

- AltaVista
- Google
- Google Scholar
- Google Print
- Lycos
- Yahoo

Datenbanken:

- Medline über Pubmed (Literatur seit 1950), www.pubmed.org
- SynnX mediator (Testaccount)
- Cochrane Library
- Scopus

Institutionen, die mit Qualitätsmanagement oder Leitlinienerstellung zu tun haben:

- Agency for Healthcare Research and Quality (AHRQ)
- American Society for Quality
- American National Standards Institute (ANSI)
- Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)
- BQS Bundesgeschäftsstelle Qualitätssicherung gGmbH
- CATWEB (Online-Katalog Europäischer Standards)
- Deutsche Bibliothek (Online-Kataloge)
- Deutsche Gesellschaft für Qualität e. V.
- Deutsches Netzwerk Evidenz basierter Medizin
- DIN Deutsches Institut für Normung e. V.
- Forschungsgemeinschaft Qualität e. V.
- ISO International Organization for Standardization
- Leitlinien.de
- Q-M-A.de (Qualitätsmanagement in der Ambulanten Versorgung)
- Quality-Management-Infosource
- Zentralstelle der Deutschen Ärzteschaft zur Qualitätssicherung in der Medizin

Für die Recherche wurden folgende Schritte festgelegt:

- 1) Identifizierung relevanter Schlagworte der Medical Subject Headings (MeSH) (s. unten)
- 2) Direkte Suche mit Schlagworten in Suchmaschinen und Datenbanken
- 3) Suche auf der Homepage von Institutionen nach Quellen und relevanten Links
- 4) Sichtung von Katalogen der ISO- und DIN-Normen: Normen mit Bezug zu Qualitätsmanagement
- 5) Suche in Literaturverzeichnissen gefundener Quellen
- 6) Suche in verwandten Artikeln zu gefundenen Quellen („related articles“ in Pubmed)
- 7) Suche in allen Arbeiten von Autoren, die mit mehreren relevanten Arbeiten gefunden wurden
- 8) Literaturempfehlungen von Experten (IMI-BE und extern)

2.2.1 Verwendete Schlagworte/MeSH-Terms

- MeSH Headings: „Quality Control“, „Data Collection“, „Feedback“
- MeSH Entry Terms:
 - „Data Quality“ (unter MeSH Heading „Research Design“)
 - „Quality Assessment“ (unter MeSH Heading „Quality Assurance, Health Care“),
 - „monitoring“ als Bestandteil von „Data and Safety Monitoring Boards“, „Safety Monitoring Boards“, „Data Monitoring Committees“ (unter MeSH Heading „Clinical Trials Data Monitoring Committees“)
 - „Quality Indicators“ (unter MeSH Heading „Quality Indicators, Health Care“)
- Sonstige Worte: „quality improvement“, „quality progress“, „adaptive monitoring“, „Qualitätsindikator“, „Qualitätslenkung“, „Qualitätsmessung Medizin“, „Ringversuch Qualität“, „data accuracy“, „source data verification“
- Zur Einschränkung der Suche Kombinationen mit: „registry“ (MeSH Heading „Registries“), „medical registry“, „cohort“ (MeSH Heading „Cohort Studies“), „clinical trials“ (MeSH Heading

2.3 Suchanfragen

2.3.1 AltaVista

- Suche mit allen Wörtern data quality, exakter Phrase medical registry, Zugriff am 08.07.2005, 3.190 Treffer

2.3.2 Google

- Suche mit allen Wörtern data quality, Zugriff am 08.07.2005, ca. 129 Mio. Treffer
- Suche mit genauer Wortgruppe data quality, Zugriff am 08.07.2005, ca. 2,7 Mio. Treffer
- Suche mit allen Wörtern data quality, genauer Wortgruppe medical registry, Zugriff am 08.07.2005, 1.790 Treffer

2.3.3 Google Scholar

- Suche mit Qualitätsindikator, all words, anywhere in the article, Zugriff am 7.7.2005, 279 Treffer
- Suche mit adaptives Monitoring, all words, anywhere in the article, Zugriff am 8.7.2005, 135 Treffer
- Suche mit Qualitätslenkung, all words, anywhere in the article, Zugriff am 8.7.2005, 159 Treffer
- Suche mit Ringversuch Qualität, all words, anywhere in the article, Zugriff am 8.7.2005, 159 Treffer
- Suche mit Qualitätsmessung Medizin, all words, anywhere in the article, Zugriff am 8.7.2005, 102 Treffer
- Suche mit Qualitätsmessung Industrie, all words, anywhere in the article, Zugriff am 14.7.2005, 80 Treffer
- Suche mit round robin test data quality, all of the words data quality AND exact phrase round robin test, anywhere in the article, Zugriff am 14.7.2005, 241 Treffer
- Suche mit all of the words data quality AND exact phrase medical registry anywhere in the article, Zugriff am 27.07.2005, 61 Treffer

2.3.4 Google Print

- Suche mit allen Wörtern data quality, Zugriff am 24.08.2005, 288 Treffer

2.3.5 Lycos

- Suche mit allen Wörtern data quality, dem genauen Ausdruck medical registry, Zugriff am 08.07.2005, 117 Treffer

2.3.6 Yahoo

- Suche mit all of the words data quality, the exact phrase medical registry, Zugriff am 08.07.2005, 3.070 Treffer

2.3.7 Pubmed

- data quality[Text Word]; Zugriff am 08.07.2005, 1.275 Treffer
- ((„registries“[TIAB] NOT Medline[SB]) OR „registries“[MeSH Terms] OR registry[Text Word]) AND „data quality“[All Fields]; Zugriff am 12.07.2005, 82 Treffer
- ((„registries“[TIAB] NOT Medline[SB]) OR „registries“[MeSH Terms] OR registry[Text Word]) AND („quality assessment“[All Fields] OR „quality control“[All Fields]); Zugriff am 12.07.2005, 332 Treffer
- cohort[All Fields] AND („quality assessment“[All Fields] OR „quality control“[All Fields]); Zugriff am 12.07.2005, 225 Treffer
- Related articles zu Gissler M, Teperi J, Hemminki E, Merilainen J. Data quality after restructuring a national medical registry. *Scand J Soc Med.* 1995 Mar; 23 (1): 75–80. PMID: 7784857 [PubMed – indexed for MEDLINE]; Zugriff am 26.07.2005, 137 Treffer
- Related articles zu Welzer T, Brumen B, Golob I, Sanchez JL, Druzovec M. Diagnostic process from the data quality point of view. *J Med Syst.* 2005 Feb; 29 (1): 59–63. PMID: 15839332 [PubMed – indexed for MEDLINE]; Zugriff am 26.07.2005, 99 Treffer

7. Related articles zu Crerand WJ, Lamb J, Rulon V, Karal B, Mardekian J. Building data quality into clinical trials. *J AHIMA*. 2002 Nov–Dec; 73 (10): 44–6, 48–53, 2; quiz 55–6. PMID: 12432815 [PubMed – indexed for MEDLINE]; Zugriff am 26.07.2005, 528 Treffer
8. Related articles zu Brown PJ, Warming-ton V. *Data quality probes-exploiting and improving the quality of electronic patient record data and patient care*. *Int J Med Inform*. 2002 Dec 18; 68 (1–3): 91–8. PMID: 12467794 [PubMed – indexed for MEDLINE]; Zugriff am 26.07.2005, 402 Treffer
9. („fraud“[MeSH Terms] OR fraud[Text Word]) AND („clinical trial“[Publication Type] OR „clinical trials“[MeSH Terms] OR „clinical trial“[Text Word]); Zugriff am 24.08.2005, 20 Treffer
10. „data collection“[MeSH Terms] AND „medical registry“[All Fields]; Zugriff am 24.08.2005, 16 Treffer
11. „data accuracy“[All Fields]; Zugriff am 24.08.2005; 79 Treffer
12. Arts D[Author]; Zugriff am 09.09.2005, 9 Treffer
13. Knatterud GL[Author]; Zugriff am 09.09.2005, 90 Treffer
14. Ollenschlager G[Author]; Zugriff am 09.09.2005, 102 Treffer
15. „source data verification“[All Fields]; Zugriff am 20.09.2005, 6 Treffer
16. „feedback“[MeSH Terms] AND („medical registry“[All Fields] OR cohort[All Fields]); Zugriff am 20.09.2005, 28 Treffer

2.3.8 SyynX mediator

SyynX mediator ist ein Rechercheportal zu Medline mit folgenden Eigenschaften:

- Publikationen in Medline von 1996 bis 2005
- Link zu Pubmed für Zugriff auf Abstracts und Volltexte
- Vorschlag von Fachtermini zur Sucheinsgrenzung (aus der Menge der Suchergebnisse extrahiert)
- Expertenprofile der Autoren
- sprachunabhängige Analyse ganzer Texte (können als Suchbegriff eingegeben werden)

- Zeitreihenanalyse (Suchergebnisse)
- Direkte Artikelbestellung über subito
- Abspeicherung von Suchanfragen und automatische Benachrichtigung des Nutzers, wenn neue passende Artikel in Medline publiziert werden

Der Zugang erfolgte über einen 30-tägigen Testaccount. Eine Speicherung von Suchergebnissen war unter dem Testaccount nicht möglich. Eine Textsuche mit dem vollständigem Abstract eines Artikels von Arts (*Arts DGT, de Keizer N, Scheffer G-J. Defining and improving data quality in medical registries: A literature review, case study, and generic framework*. *J Am Med Inform Assoc* 2002; 9: 600–611) lieferte nur Quellen, die auch bei anderen Recherchen gefunden wurden. Das Suchergebnis hing stark von der gewählten Gewichtung ab: 0 (der Suchartikel selbst wurde nicht gefunden!) bis 7.378 Treffer. Bei Beibehaltung der voreingestellten Gewichtung ergaben sich 906 Treffer (einschließlich des Suchartikels).

2.3.9 Cochrane Library

1. Suche mit Search all Text data quality; Zugriff am 26.08.2005, Treffer: 3.915 Systematic Reviews, 20 Methodology Reviews, 598 Methodology Register
2. Suche mit Search all Text „data quality“; Zugriff am 26.08.2005, Treffer: 21 Systematic Reviews, 2 Methodology Reviews, 26 Methodology Register

2.3.10 Scopus

1. (ALL(„medical registry“) AND ALL (data quality)); Zugriff am 08.07.2005, 179 Treffer

2.4 Ergebnisse

Die Suchergebnisse wurden anhand ihres Titels und ggf. eines Abstracts von einem der Autoren auf Relevanz geprüft. Hierbei wurde nach Konkretisierung der angestrebten Projektergebnisse mit den Projektpartnern auf drei Themen fokussiert: Indikatoren für Da-

tenqualität, Source Data Verification sowie Feedback zur Datenqualität. Die anderen in Abschnitt 2.4 beschriebenen Bereiche wurden in Folge nur noch cursorisch betrachtet. Die als interessant erscheinenden Beiträge wurden im Volltext beschafft, so weit möglich. Die Volltexte sind Grundlage der folgenden Darstellung, die die bis zum 03.03.2006 betrachtete relevante Literatur enthält. Die Gliederung erfolgt dabei in Fortschreibung der Bereiche, die zu Beginn des Projektes als relevant für die Literatursichtung identifiziert wurden.

2.4.1 Erstellung und Inhalt medizinischer Leitlinien

AWMF 2000

Medizinische Leitlinien werden in einem 3-Stufen-Prozess entwickelt:

1. Expertengruppe (S1): Eine repräsentativ zusammengesetzte Expertengruppe der Wissenschaftlichen Medizinischen Fachgesellschaften erarbeitet im Konsens eine Empfehlung, die vom Vorstand der Fachgesellschaft verabschiedet wird.
2. Formale Evidenz-Recherche (S2e) oder formale Konsensusfindung (S2k): Die Leitlinien werden aus formal bewerteten Aussagen aus der Literatur abgeleitet oder in einem formalen Konsensusverfahren (nominaler Gruppenprozess, Konsensuskonferenz oder Delphikonferenz) beraten und verabschiedet.
3. Leitlinie mit allen Elementen systematischer Entwicklung (S3)

Leitlinien können in Textform, als Tabellen, klinische Algorithmen und als Kombination dieser Elemente dargestellt werden.

AWMF o. J.

Folgende Qualitätskriterien sollten Leitlinien der Entwicklungsstufe S3 erfüllen:

- Validität/Gültigkeit: Die verfügbaren wissenschaftlichen Erkenntnisse und Erfahrungen müssen korrekt interpretiert werden, so dass die Nutzung der Leitlinie als Hilfe zur Entscheidungsfindung zu der beabsichtigten Verbesserung in Diagnostik und/oder Therapie führt.

- Reliabilität/Zuverlässigkeit: Unter gleichen klinischen Umständen sollte jeder Arzt die Leitlinie gleich oder sehr ähnlich als Hilfe zur Entscheidungsfindung nutzen.
- Reproduzierbarkeit: Bei gleichen verfügbaren wissenschaftlichen Erkenntnissen und Erfahrungen sollte eine weitere unabhängige Expertengruppe zu gleichen Empfehlungen kommen.
- Repräsentative Entwicklung: Alle Schlüsseldisziplinen sollten ihren Beitrag zur Entwicklung der Leitlinie geleistet haben.
- Klinische Anwendbarkeit: Die Zielgruppe, für die wissenschaftliche Erkenntnisse und Erfahrungen verfügbar sind, ist definiert.
- Klinische Flexibilität: Die Leitlinie nennt Ausnahmefälle und zeigt auf, wie die Bedürfnisse der Patienten in die Entscheidungsfindung einzubeziehen sind.
- Klarheit: Leitlinien benutzen präzise Definitionen, eine eindeutige Sprache und benutzerfreundliche Formate.
- Genaue Dokumentation: Die Leitlinien machen Angaben über die Teilnehmer an der Entwicklung, die Annahmen und Methoden und verknüpfen die ausgesprochenen Empfehlungen mit den verfügbaren wissenschaftlichen Erkenntnissen und Erfahrungen.
- Planmäßige Überprüfung: Die Leitlinien enthalten Angaben darüber, wann und wie sie überprüft werden.
- Überprüfung der Anwendung: Die Leitlinien zeigen Verfahren auf, mit denen die Akzeptanz der Empfehlungen in der Praxis ermittelt werden kann.
- Kosten-Nutzen-Verhältnis: Leitlinien sollen zur Verbesserung der medizinischen Versorgung bei akzeptablen Kosten führen.

AWMF/ÄZQ 2001

AWMF und ÄZQ haben ein gemeinsames Leitlinien-Manual herausgegeben. Es beschreibt die Methodik für die Entwicklung und Implementierung ärztlicher Leitlinien. Die folgenden Themen werden mit Beispielen von deutschen Leitlinienprojekten behandelt:

- Organisation der Leitlinienentwicklung
 - Klare Nennung der für die Leitlinienentwicklung verantwortlichen Institution

- Detaillierte Angabe über finanzielle oder andere Formen der Unterstützung durch Dritte
- Diskussion möglicher Einflussnahme durch Sponsoren
- Auswahl des Leitlinien-Themas
 - Explizite Nennung der Gründe für die Leitlinienentwicklung
 - Eindeutige Definition der Ziele (u.a. Outcome), der Zielgruppen und Anwender
 - Angabe über Situationen, in denen spezielle Empfehlungen der Leitlinie nicht berücksichtigt werden können oder sollen
 - Berücksichtigung von Ansichten, Präferenzen und möglichen Reaktionen der Patienten-Zielgruppe(n)
- Autoren und Herausgeber der Leitlinie
 - Klare Nennung der an der Erstellung der Leitlinie Beteiligten (Fach-, Interessen-, Patientengruppen) hinsichtlich ihrer Funktion und der Art ihrer Beteiligung
 - Beteiligung der von der Leitlinie im wesentlichen Betroffenen (die mit der Thematik befassten Fachdisziplinen und Patienten)
- Systematische Evidenz-Recherche
 - Beschreibung der Quellen und Methoden, mit deren Hilfe die den Empfehlungen zugrunde liegenden Belege (Evidenz) gesucht, identifiziert und ausgewählt wurden (Formulierung von Schlüsselfragen durch die Autoren der Leitlinie)
 - Nachvollziehbare Dokumentation der Informationsquellen und Suchstrategien
 - Nennung der Methoden zur Interpretation und Bewertung der Evidenzstärke (z. B. 4-Stufen-Modell der Arzneimittelkommission der deutschen Ärzteschaft (AkdÄ), Einteilung der Evidenzstärke von Therapiestudien in „Levels of Evidence“, Bewertungssystem der British Hypertension Society für Studien und Empfehlungen u. a.)
- Formulierung der Empfehlungen
 - Eine ausschließlich auf theoretischer Evidenzbewertung basierte Strategie ohne praxisbezogene Formulierung der Leitlinienempfehlungen im Rahmen eines systematischen und transparenten Konsensverfahrens ist nicht zulässig
 - Nennung der zur Auswahl der Leitlinien-Empfehlungen eingesetzten Methoden sowie der zur Konsentierung der Leitlinien-Empfehlungen eingesetzten Verfahren
 - Angabe, wie die Vorstellungen interessierter aber nicht beteiligter Gruppen berücksichtigt wurden
 - Exakte Dokumentation der Verknüpfung der wichtigsten Empfehlungen mit der zugrunde liegenden Evidenz
 - Nennung von Nutzen, Nebenwirkungen, Kosten, Ergebnissen von Leitlinien-Empfehlungen und Abwägung möglicher Vorteile, Risiken und Kosten
- Begutachtung und Pilotversuch
 - Begutachtung der Leitlinie durch unabhängige Reviewer vor der Veröffentlichung
 - Darstellung der Methoden, Kommentierungen und Konsequenzen der Begutachtung
 - Pilotversuch vor der Veröffentlichung
 - Darstellung der Methoden, Ergebnisse und Konsequenzen des Pilotversuchs
 - Vergleich der Leitlinie mit anderen thematisch vergleichbaren Leitlinien
- Präsentation, Disseminierung, Implementierung
 - Klarheit und Eindeutigkeit der Versorgungsprobleme und Empfehlungen
 - differenzierte Empfehlungen zur Entscheidung hinsichtlich ambulanter oder stationärer Versorgung
 - Angaben über notwendige, unzureichende, überflüssige oder obsolezte Maßnahmen
 - Zusammenfassende Darstellung des Inhalts
 - Existenz von Maßnahmen/Instrumenten zur Anwendungsunterstützung
 - Berücksichtigung möglicher Anwendungsprobleme (Einstellungs-, Verhaltensänderungen, organisatorische Hindernisse)
 - Angabe messbarer Kriterien, anhand

derer der Effekt der Leitlinienanwendung überprüft werden kann

- Planmäßige Überarbeitung
 - Angabe eines Datums für Überprüfung/Aktualisierung
 - Angabe der Zuständigkeit und Verfahrensweise für die Überprüfung/Aktualisierung

AWMF/ÄZQ 2005

AWMF und ÄZQ haben im Jahre 2005 die Fassung 2005/2006 des Deutschen Instruments zur methodischen Leitlinien-Bewertung (DELBI) herausgegeben. Die aktuell gültige Version ist unter www.delbi.de zu finden. Die aktuelle Version wurde am 07.06.2005 verabschiedet. Verantwortlich für die Weiterentwicklung ist das Ärztliche Zentrum für Qualität in der Medizin (ÄZQ) gemeinsam mit der Leitlinienkommission der AWMF. DELBI umfasst folgende sieben Domänen:

- Domäne 1: Geltungsbereich und Zweck
- Domäne 2: Beteiligung von Interessengruppen
- Domäne 3: Methodologische Exaktheit der Leitlinien-Entwicklung
- Domäne 4: Klarheit und Gestaltung
- Domäne 5: Generelle Anwendbarkeit
- Domäne 6: Redaktionelle Unabhängigkeit
- Domäne 7: Anwendbarkeit im deutschen Gesundheitssystem

Zu jeder Domäne gehören mehrere Fragen, die jeweils auf einer Skala von 1 (trifft überhaupt nicht zu) bis 4 (trifft uneingeschränkt zu) zu beantworten sind. Nach einer allgemeinen Einführung wird der Gebrauch des Instruments detailliert beschrieben.

Bundesärztekammer/Kassenärztliche Bundesvereinigung 1997

Die Vorstände der Bundesärztekammer und der Kassenärztlichen Bundesvereinigung haben Beurteilungskriterien für Leitlinien in der medizinischen Versorgung beschlossen. *Leitlinien* sind systematisch entwickelte Entscheidungshilfen über die angemessene ärztliche Vorgehensweise bei speziellen gesundheitlichen Problemen. Sie sind Orientierungshilfen, von denen in begründeten Fällen abgewichen werden kann oder sogar

muss. Im Gegensatz dazu werden *Richtlinien* als Regelungen des Handelns oder Unterlassens verstanden, die von einer rechtlich legitimierten Institution konsentiert, schriftlich fixiert und veröffentlicht wurden, für den Rechtsraum dieser Institution verbindlich sind und deren Nichtbeachtung definierte Sanktionen nach sich zieht.

Ziele der Leitlinien sind:

- Sicherung und Verbesserung der gesundheitlichen Versorgung der Bevölkerung
- Berücksichtigung systematisch entwickelter Entscheidungshilfen in der ärztlichen Berufspraxis
- Motivation zu wissenschaftlich begründeter und ökonomisch angemessener ärztlicher Vorgehensweise
- Vermeidung unnötiger und überholter medizinischer Maßnahmen und unnötiger Kosten
- Verminderung unerwünschter Qualitätsschwankungen im Bereich der ärztlichen Versorgung
- Information der Öffentlichkeit

Als Qualitätskriterien für Leitlinien werden genannt:

- Transparenz
- Gültigkeit
- Zuverlässigkeit und Reproduzierbarkeit
- Multidisziplinäre Entwicklung
- Anwendbarkeit
- Flexibilität
- Klarheit, Eindeutigkeit
- Dokumentation der Leitlinienentwicklung
- Planmäßige Überprüfung
- Überprüfung der Anwendung
- Kosten-Nutzen-Verhältnis
- Verfügbarkeit der Leitlinie

Die folgende Gliederung von Leitlinien wird vorgeschlagen:

1. Ziele
2. Vorgehensweise
3. Ergebnisse
4. Beweise und Auswahlkriterien
5. Nutzen, Nebenwirkungen, Kosten
6. Zusammenfassende Empfehlungen der Leitlinie
7. Implementierungsplan
8. Belege für die Berücksichtigung und Wirksamkeit der Leitlinie

9. Gültigkeitsdauer und Verantwortlichkeit für die Fortschreibung
10. Angabe von Autoren, Kooperationspartnern, Konsensusverfahren, Sponsoren

Bundesärztekammer/Kassenärztliche Bundesvereinigung 2005

Bundesärztekammer, Kassenärztliche Bundesvereinigung und die Arbeitsgemeinschaft der Medizinischen Wissenschaftlichen Fachgesellschaften haben sich im Jahr 2000 auf ein gemeinsames Programm zur Förderung der Qualität von Leitlinien in der Medizin geeinigt. Grundlage waren neben nationalen Anforderungen die Empfehlungen des Europarates zur Leitlinien-Methodik. Ergebnis der Bemühungen ist eine Leitlinien-Checkliste („Deutsches Instrument zur methodischen Leitlinien-Bewertung (DELBI)“). DELBI wurde von einer multidisziplinären Expertengruppe erarbeitet und am 07.06.2005 verabschiedet. Verantwortlich für die Weiterentwicklung ist das Ärztliche Zentrum für Qualität in der Medizin (ÄZQ) gemeinsam mit der Leitlinienkommission der AWMF (s. Beschreibung von DELBI unter AWMF/ÄZQ 2005).

EMA 2005a

Dieses Papier beschreibt verschiedene Guidelines, die das europäische gesetzliche Rahmenwerk für Arzneimittel und ein harmonisiertes Vorgehen für ihre Entwicklung unterstützen. Die Gesetzgebung der EU auf dem pharmazeutischen Sektor wird begleitet von einer Reihe von Guidelines. Hier ist Guideline definiert als ein EU-Dokument, das entweder direkt der Erfüllung einer gesetzlichen Verpflichtung dienen soll oder eine Hilfestellung zur bestmöglichen Erfüllung gesetzlicher Vorschriften anbietet. Verschiedene Typen von Guidelines existieren im Pharmabereich:

- Regulatory Guidelines
- Scientific Guidelines related to Quality, Safety and Efficacy
- Good Manufacturing Practice (GMP) Guidelines
- Maximum Residue Limits Guidelines
- Pharmacovigilance Guidelines
- Good Clinical Practice (GCP) and conduct of clinical trials Guidelines

- Orphan Medicinal Products Designation Guidelines
- Herbal medicinal products Guidelines
- Good Distribution Practice (GDP) Guidelines
- Good Laboratory Practice (GLP) Guidelines
- European Directorate for the Quality of Medicines (EDQM) and the European Pharmacopoeia
- Other technical and procedural EMA guidelines
- Other related Community documents prepared/published by EMA

Die Entwicklung einer Guideline geschieht normalerweise in folgenden Schritten:

1. Auswahl des Themas und Einschluss in das/die relevante(n) Arbeitsprogramm(e)
 2. Berufung des Berichterstatters und falls notwendig Ko-Berichterstatters
 3. Entwicklung eines Konzeptpapiers
 4. Annahme und Freigabe des Konzeptpapiers für die Beratung
 5. Vorbereitung der initialen Draftversion der Guideline
 6. Freigabe der Draftversion der Guideline für die Beratung
 7. Sammlung von Kommentaren
 8. Vorbereitung der finalen Version der Guideline
 9. Annahme der finalen Version der Guideline für die Veröffentlichung
 10. Implementierung
- Daran schließen sich an:
11. Training
 12. Pflege und Revision der Guidelines

Europarat 2002

Der Europarat hat Empfehlungen zur Entwicklung einer Methodik für die Ausarbeitung von Leitlinien für die optimale medizinische Praxis herausgegeben. Die Europäische Union hat im Rahmen ihres Biomed-Programms eine internationale Arbeitsgruppe finanziert, die ein Instrument zur Leitlinienbewertung entwickelt hat (AGREE [Appraisal of Guidelines for Research and Evaluation in Europe]-Instrument, Tabelle 4). Folgende Eigenschaften klinischer Leitlinien sind erstrebenswert:

- Validität
- Evidenzstärke

- Wahrscheinliche Therapieergebnisse
- Zuverlässigkeit/Reproduzierbarkeit
- Klinische Anwendbarkeit
- Klinische Flexibilität
- Klarheit
- Multidisziplinärer Prozess
- Planmäßige Überprüfung
- Dokumentation

Derzeit bevorzugtes Verfahren ist die systematische Suche nach einschlägiger Evidenz vor Formulierung der Leitlinie. Grundlage sollen die besten verfügbaren wissenschaftlichen Belege und ggf. eine explizite Ergebnisschätzung für alternative (nicht-empfohlene) Vorgehensweisen sein. Leitlinien können nicht nur Evidenz basiert, sondern auch ergebnisorientiert sein. In diesem Fall wird die Größe der erwarteten Effekte von einer Expertengruppe geschätzt sowie Vor- und Nachteile abgewogen. Fehlt es an zuverlässiger Evidenz, kann die Expertengruppe Feststellungen zu wichtigen Versorgungsentscheidungen nach dem Konsensprinzip treffen. Vor der Freigabe einer Leitlinie sollte sie begutachtet oder in einer Pilotstudie getestet werden. Werden Leitlinien aus anderen Ländern oder Regionen übernommen, müssen sie überarbeitet oder auf ihre Anwendbarkeit im neuen Umfeld geprüft werden. Die Verbreitung der Leitlinie sollte als Teil des Erstellungsprozesses geplant werden. Generell sollte eine Leitlinie über möglichst viele Kanäle verbreitet werden, da unterschiedliche Formen sich nicht nur gegenseitig verstärken, sondern auch die inhaltliche Relevanz der Leitlinie hervorheben. Strategien zur Implementierung können sich auf verschiedene Ebenen beziehen (Tabelle 7):

- Edukative Interventionen
- Finanzielle Interventionen (gegenüber Leistungserbringern oder Patienten)
- Organisatorische Interventionen
- Regulatorische Interventionen

Zur Überprüfung der Wirksamkeit einer Leitlinie werden Indikatoren definiert. Diese sollten folgende Eigenschaften aufweisen. Sie sollten

- leicht messbar,
- aus der Routedokumentation zugänglich,
- vergleichbar zwischen verschiedenen Einheiten und

- Indikatoren für definitive gesundheitliche Outcomes sein

Da das medizinische Fachvokabular zahlreiche nur vage definierte Konzepte enthält und ein und derselbe Begriff in verschiedenen Ländern völlig unterschiedliche Bedeutungen haben kann, ist es bei der Erstellung und Implementierung von Leitlinien wichtig, den Gebrauch der verschiedenen Begriffe in ihrem jeweiligen sozialen und sprachlichen Kontext zu prüfen. Im Anhang des Dokuments werden Leitlinienprogramme und Clearingstellen in verschiedenen europäischen Ländern sowie verschiedene Beispiele für die Einteilung von Leitlinien in Evidenzklassen genannt.

Helou/Ollenschläger 1998

Leitlinien dienen als Instrument zur Qualitätssicherung. Sie sollten bestimmten Qualitätsanforderungen genügen und werden damit selbst zum Gegenstand der Qualitätssicherung. Da die Qualität deutscher Leitlinien in der Vergangenheit wiederholt in Zweifel gezogen wurde, hat die Planungsgruppe der Ärztlichen Zentralstelle Qualitätssicherung auf der Basis der von der BÄK und KBV festgelegten „Beurteilungskriterien für Leitlinien in der medizinischen Versorgung“ ein deutschsprachiges Instrument zur standardisierten Erfassung und Dokumentation der Qualitätskriterien von Leitlinien herausgegeben. Zu der Checkliste wurde ein Nutzermanual erstellt, mit dem sich die Autoren beschäftigen.

Zentrales Qualitätskriterium einer Leitlinie ist ihre Validität. Validität bedeutet hier, dass die Anwendung der Leitlinie die erwarteten medizinischen und ökonomischen Ergebnisse liefert. Ideal wäre eine Effektevaluation, die aber aufgrund methodischer Probleme und eines beträchtlichen finanziellen und personellen Aufwands nicht für alle existierenden medizinischen Leitlinien leistbar ist. An die Stelle der ergebnisorientierten Evaluation kann eine vorausschauende Bewertung der Leitlinienqualität treten. Aus der Literatur lassen sich drei zentrale Prädiktoren für die Validität einer Leitlinie identifizieren:

- die systematische Suche, Identifizierung, Auswahl und Synthese der wissenschaftlichen Evidenz

- die repräsentative Zusammensetzung des Leitlinienpanels
- die explizite Kenntlichmachung des Zusammenhangs zwischen den Empfehlungen und der zugrundeliegenden Evidenz

Die Qualitätsbewertung von Leitlinien wirft einige methodische Probleme auf:

- unzureichende Interrater-Übereinstimmung
- unzureichende Dokumentation der Leitlinienerstellung

Um diesen Problemen zu begegnen, wurde das Nutzermanual entwickelt.

Stausberg et al. 2003a

Klinische Leitlinien sind ein wichtiger Ansatz zur Qualitätsverbesserung. Am Beispiel einer Leitlinie für das akute klinische Management polytraumatisierter Patienten sollten die folgenden Fragen beantwortet werden:

1. Wird eine formale Analyse mit einem Werkzeug zur Modellierung von Geschäftsprozessen formale Inkonsistenzen und Fehler einer papierbasierten Leitlinie reduzieren?
2. Führt der Einsatz eines Modellierungswerkzeugs zu einer Verbesserung der medizinischen Qualität einer Leitlinie?

Die untersuchte Leitlinie liegt in Papierform vor und besteht aus einer Reihe von Flussdiagrammen, ergänzt um Checklisten und Fußnoten. Für die Validation wurde ein für die Industrie entwickeltes Geschäftsprozessmodell (BPM nach Scheer) benutzt, das verschiedene Sichten (Daten, Ressourcen, Funktionen) auf einen Prozess erlaubt. Diese Sichten werden auf einer Steuerungsebene zusammengeführt. Die papierbasierte Leitlinie wurde zuerst 1:1 in das Modell übertragen. Das entstandene Modell wurde zuerst auf formale Fehler und logische Inkonsistenzen überprüft und diese korrigiert. Dabei wurden die folgenden Fehlertypen gefunden:

- Sackgassen, d. h. Endpunkte, die weder ein Ergebnis der Leitlinie repräsentieren, noch eine Verbindung zum Hauptprozess besitzen.
- Entscheidungsknoten, die mehrere Fragen enthalten, aber nur Ausgänge für ein globales „ja“ oder „nein“.
- Zusammengesetzte Funktionen mit meh-

ren Aktionen. Die Reihenfolge der Aktionen ist nicht ersichtlich.

- Schleifen führen u. U. zu Situationen, die nicht automatisch gelöst werden können (Endlosschleifen).
- Kriterien außerhalb der Flussdiagramme (Klartext) müssen formal beschrieben werden.
- Der Gebrauch von Synonymen kann zu Missverständnissen führen.
- Unspezifizierte Funktionen können nicht eindeutig interpretiert werden.

In einem zweiten Schritt wurde das Modell unter medizinischen Gesichtspunkten überarbeitet. Insgesamt lässt sich die Frage 1 eindeutig mit „ja“ beantworten. Ein Quasi-Standard für die generische Modellierung von medizinischen Leitlinien ist das Guideline Interchange Format (GLIF). Eine GLIF-Repräsentation kann in UML abgebildet werden.

2.4.2 Qualitätsindikatoren

Altenhofen et al. 2005

Seit den 70er Jahren des 20. Jahrhunderts existieren in Deutschland einrichtungsübergreifende vergleichende Qualitätsbewertungen im medizinischen Bereich. Die zunächst freiwilligen Maßnahmen mit dem Charakter einer freiwilligen Selbstverpflichtung sind mittlerweile durch gesetzliche Anforderungen (§ 135a SGB V) verpflichtend geworden. Alle Leistungserbringer müssen Qualitätsmanagement betreiben. Da die Ergebnisse der Qualitätsvergleiche für die Versorgungslandschaft sowie für die einzelnen Leistungserbringer wirtschaftlich relevant sind, erhöhen sich die methodischen Anforderungen an die Qualitätsbewertungen.

Basis der Qualitätsbewertung sind Qualitätsindikatoren, anhand derer die aktuelle Versorgungssituation quantifiziert („Ist-Wert“) und mit einem vorab festgelegten Soll-Wert verglichen wird. Die Qualitätsindikatoren beziehen sich auf Strukturen, Prozesse und Ergebnisse der medizinischen Versorgung. Mit ihrer Hilfe wird bewertet, wie sich die Versorgungsergebnisse gestalten (Evaluationsfunktion), und ob die Versorgung dem allgemeinen Stand der wissenschaftlichen Erkenntnisse entspricht

(Monitoringfunktion). Darüber hinaus haben die Qualitätsindikatoren eine Steuerungsfunktion, indem sie eine besondere Aufmerksamkeit auf die bewerteten Strukturen, Prozesse und Ergebnisse richten und eine Auswahl von Leistungserbringern aufgrund von Qualitätsvergleichen ermöglichen. Daraus ergibt sich die Forderung nach einer guten Qualität der Qualitätsindikatoren selbst. Die aktuelle Situation ist durch zwei Probleme gekennzeichnet:

- Qualitätsindikatoren zu gleichen Krankheitsentitäten werden von verschiedenen Gruppen entwickelt und führen zu einer kaum noch handhabbaren Dokumentationsflut. Die existierenden Klinik- und Praxisinformationssysteme umfassen nicht alle erforderlichen Daten und unterstützen nicht alle unterschiedlichen Datenformate für die Erfassung der Qualitätsindikatoren.
- Es ist unklar, inwieweit die eingesetzten Qualitätsindikatoren die allgemein diskutierten methodischen Gütekriterien für solche Indikatoren erfüllen. Oft fehlt die Prüfung der Evidenzbasierung.

Die „Expertengruppe Qualitätsindikatoren“ beim Ärztlichen Zentrum für Qualität in der Medizin (ÄZQ) empfiehlt die Einrichtung eines Programms zur Entwicklung und Implementierung Evidenz basierter Qualitätsindikatoren. Als methodische Basis kann die „Checkliste zur Überprüfung der methodischen Qualität von Qualitätsindikatoren“ der ÄZQ dienen. Zusätzlich sollten Expertise aus dem Rehabilitations- und ambulanten Sektor sowie die Perspektive der Patienten in die weitere Entwicklung der Checkliste einfließen.

ÄZQ 2002

Klinische Messgrößen messen Eigenschaften der medizinischen Versorgung, die im Rahmen des Qualitätsmanagements bewertet werden sollen. Der Begriff „klinische Messgröße“ wird hier als Oberbegriff zu den Begriffen „Kriterien“ und „Indikatoren“ benutzt. Die untersuchten Eigenschaften der medizinischen Versorgung werden auch als Versorgungsaspekte bezeichnet. Die Joint Commission on Accreditation of Healthcare Organi-

zations (JCAHO) schlägt Kriterien für die Leistungserbringung bei einzelnen Versorgungsaspekten vor:

- Zugänglichkeit der Versorgung
- Angemessenheit der Versorgung
- Stetigkeit/Koordination der Versorgung
- Wirksamkeit der Versorgungspraxis
- Wirksamkeit unter Idealbedingungen
- Wirtschaftlichkeit in der Versorgung
- Patientenorientierung der Versorgung
- Sicherheit der Versorgungsumgebung
- Rechtzeitigkeit der Versorgung

Qualitätsindikatoren sind qualitätsbezogene Kennzahlen, die die Qualität einer Einheit durch Zahlen oder Zahlenverhältnisse (Raten) indirekt abbilden und eine Unterscheidung zwischen guter und schlechter Qualität erlauben. Der Referenzbereich ist derjenige Wertebereich eines Qualitätsindikators, innerhalb dessen die Qualität als „gut“ oder „unauffällig“ definiert wird. Fallen Ober- und Untergrenze des Referenzbereichs zusammen, wird er als Referenzwert bezeichnet.

Klinische Messgrößen können aus Leitlinien abgeleitet werden. Sie bilden u. a. die Basis für die Sicherung und Verbesserung der gesundheitlichen Versorgung der Bevölkerung und können im Rahmen des internen Qualitätsmanagements und der externen Qualitätssicherung eingesetzt werden. Klinische Messgrößen müssen selbst eine Reihe von Qualitätskriterien erfüllen:

- Relevanz von klinischen Messgrößen für die Gesundheitsversorgung
 - Konsens über die Bedeutung der Versorgungsaspekte
 - Beeinflussbarkeit des Versorgungsaspektes
 - Relevanz
 - Ausgewogenheit
 - Kosten-Nutzen-Verhältnis
 - Aktualität/planmäßige Überprüfung
- Qualität einer klinischen Messgröße
 - Verständlichkeit
 - Validität
 - Reproduzierbarkeit
 - Reliabilität
 - Diskriminationsfähigkeit
 - Adjustierbarkeit
- Qualitätsdarlegung für klinische Messgrößen

- Dokumentation der Entwicklung eines Programms für klinische Messgrößen
- Transparenz
- Beteiligung Betroffener
- Verwendung der Messergebnisse

Kriterien und Indikatoren sollten durch ein systematisches Programm entwickelt, implementiert, evaluiert und gepflegt werden. Als Beispiel hierfür kann der mit ungefähr 120.000 Diabetikern durchgeführte deutsche Diabetes-Vertrag zwischen den Krankenkassen in Nordrhein-Westfalen und der Kassenärztlichen Vereinigung Nordrhein genannt werden. Analog zum Umgang mit Leitlinien empfiehlt sich die systematische kritische Bewertung klinischer Messgrößen. Hierzu geben die Autoren eine Checkliste zur Bewertung klinischer Messgrößen des Qualitätsmanagements an.

Hölzer et al. 2001

Für Qualitätsindikatoren im medizinischen Bereich gelten primär die Grundsätze der Validität und Reliabilität. Zusätzlich gefordert werden ein vertretbarer Aufwand für die Erhebung und Transparenz. Zwei Arten von Indikatoren lassen sich unterscheiden:

1. populationsbasiert: Anteil Personen mit einem bestimmten Ereignis bezogen auf die Gruppe insgesamt
2. fallbasiert: Definition einer Gruppe und Prüfung, ob ein bestimmtes Ereignis in der Gruppe auftritt, danach Vergleich der beobachteten Häufigkeit mit derjenigen eines definierten Standards

Die Probleme bei der Anwendung von Indikatoren werden beschrieben und Ergebnisse aus dem Bereich der Onkologie vorgestellt.

JCAHO 1990

Im Gesundheitswesen gibt es einen wachsenden Bedarf für objektive Qualitätsnachweise. Standardisierte Evaluation und Fallbasierter Review sind zwei geläufige Verfahren, um die Qualität der medizinischen Pflege zu beurteilen. Die standardisierte Evaluation benutzt zuvor festgelegte struktur- und prozessorientierte Standards, um den Fokus auf die Faktoren zu lenken, die für eine gute Qualität der Pflege notwendig sind. Der Bedarf an detaillierter Information über die Leistung einer

Organisation ist mit diesem Ansatz nicht mehr erfüllbar. Die Ergebnisse eines Fallbasierten Reviews sind trotz Rückgriff auf Standards oft subjektiv und nicht aggregierbar, so dass Trends und Muster nicht untersucht werden können.

Die Joint Commission und andere Segmente der Gesundheitsindustrie befassten sich mit der Erstellung eines Systems für Monitoring und Evaluation der Leistung einer Organisation im Gesundheitswesen. Leistung meint hier die Fähigkeit zur Ausführung und Steuerung aller Aktivitäten, die am meisten zu den gewünschten Ergebnissen für die Patienten beitragen. Das komplette System besteht aus drei Teilen:

- Leistungsindikatoren für Monitoring und Evaluation wichtiger Funktionen der Organisation. Sie sollen nicht als direkte Indikatoren für Qualität dienen, sondern Hinweise für Verbesserungsmöglichkeiten liefern.
- Leitlinien oder Standards
- Leistungsdatenbank als Basis für Aussagen über die Qualität der Pflege. Spezielle Maßnahmen zur Sicherstellung einer guten Datenqualität (Vollständigkeit, Richtigkeit) sind hier notwendig.

Durch Leistungsindikatoren messbare Komponenten der Pflegequalität sind:

- Zugänglichkeit der Pflege
- Angemessenheit der Pflege
- Kontinuität der Pflege
- Effektivität der Pflege
- Wirksamkeit der Pflege
- Effizienz der Pflege
- Themen aus der Patientenperspektive (Einbindung in Entscheidungsprozesse, Zufriedenheit mit der Pflege)
- Sicherheit der Pflegeumgebung
- Rechtzeitigkeit der Pflege

Indikatoren lassen sich unter verschiedenen Gesichtspunkten klassifizieren:

- Ergebnisindikator/Prozessindikator
- Raten basierter Indikator (Anzahl Ereignisse pro Gesamtheit)/Wächterindikator für Auftreten ernster Ereignisse, die eine umgehende Reaktion erfordern
- Indikator für erwünschte Ereignisse/Indikator für unerwünschte Ereignisse

Um einen Indikator richtig einsetzen zu können, sollten folgende Informationen über ihn

bereitgestellt werden (nach Table 5 auf Seite 15):

- Indikatoraussage
- Begriffsdefinitionen
- Angabe des Indikatortyps
- Begründung für die Wahl des Indikators
- Beschreibung der dem Indikator zugrunde liegenden Patientenpopulation (Zähler, Nenner)
- Beschreibung der Datensammlung und – aggregierung zur Berechnung des Indikators
- Zugrundeliegende Faktoren, die den Indikator beeinflussen können (Patient, Arzt, Organisation)

Der Prozess der Indikatorentwicklung hat drei Hauptziele:

- Identifikation der Schlüsselfunktionen, die durch den Indikator gemessen werden sollen
- Organisation des Entwicklungsprozesses in einer Weise, die eine effiziente und effektive Nutzung der Ressourcen der Organisation und der Experten erlaubt
- Erreichung eines Expertenkonsensus über die entwickelten Indikatoren

Vor der formalen Implementierung sollte die Reliabilität und Validität der Indikatoren durch die Anwendung in ausgewählten Testzentren (Krankenhäuser) getestet werden. Die Reliabilität ist direkt mit der Qualität der zugrunde liegenden Daten verknüpft.

Mainz 2003a

Qualität im Gesundheitswesen kann definiert werden als „der Grad, zu dem Gesundheitsdienstleistungen für den Einzelnen und die Bevölkerung die Wahrscheinlichkeit eines gewünschten Gesundheitsergebnisses erhöhen und konsistent mit dem aktuellen Wissensstand sind“. Ein idealer klinischer Qualitätsindikator erfüllt folgende Voraussetzungen:

- Er basiert auf festgelegten Definitionen und ist erschöpfend beschrieben.
- Er ist sehr spezifisch und sensitiv.
- Er ist valide und reliabel.
- Er unterscheidet gut.
- Er bezieht sich auf für den Benutzer klar definierte Ereignisse.
- Er erlaubt sinnvolle Vergleiche.
- Er ist Evidenz basiert.

Ein klinischer Indikator lässt sich auf verschiedene Weise klassifizieren:

- Raten basierter Indikator (Anzahl Ereignisse pro Gesamtheit)/Wächterindikator für Auftreten ernster Ereignisse, die eine umgehende Reaktion erfordern
- Struktur/Prozess/Ergebnis
- Allgemein/krankheitsspezifisch
- Pflgetyp: präventiv/akut/chronisch
- Funktion: Screening/Diagnose/Behandlung/Nachverfolgung
- Modalität: Vorgeschichte/körperliche Untersuchung/Laboruntersuchung/Medikation/andere Intervention

Mainz 2003b

Dieses Papier beschreibt das methodische Vorgehen bei der Entwicklung Evidenz basierter klinischer Indikatoren für das Gesundheitswesen. Die Methodik basiert auf publizierten Erkenntnissen und Erfahrungen aus dem Dänischen Nationalen Indikator Projekt und besteht aus einer Planungsphase und der Entwicklungsphase. In der Planungsphase werden folgende Schritte durchlaufen:

1. Auswahl des klinischen Bereichs für die Evaluation (Wichtigkeit, Möglichkeit der klinischen Intervention)
2. Organisation der Arbeitsgruppe

Die folgenden Schritte sind Bestandteil der Entwicklungsphase:

1. Gewinnung des Überblicks über vorhandene Evidenz und Praxis
2. Auswahl klinischer Indikatoren und Standards
 - Auswahl von Prozessindikatoren
 - Auswahl von Ergebnisindikatoren
 - Identifizierung prognostischer Faktoren (Risikoadjustierung)
 - Etablierung von Konsensus- und Bewertungsverfahren
3. Entwurf der Messspezifikation
 - Definition von Indikatoren und Standards
 - Identifizierung der Zielpopulation
 - Bestimmung von Einschluss- und Ausschlusskriterien
 - Ausarbeitung der Strategie zur Risikoadjustierung
 - Identifikation der Datenquellen

- Beschreibung des Vorgehens bei der Datensammlung
 - Planentwurf
4. Test (Reliabilität, Validität)

Rubin et al. 2001

Dieses Papier beschreibt den Prozess für die Entwicklung und Implementierung von Qualitätsindikatoren für die Untersuchung von Prozessqualität im Gesundheitssektor. Der Entwicklungsprozess besteht aus folgenden Schritten:

1. Definition der Zielgruppe (z. B. Ärzte, Patienten, Kunden) und der Nutzung des Indikators (z. B. Qualitätsverbesserung, Auswahl von Anbietern)
2. Auswahl des klinischen Bereichs für die Evaluation (Auswirkungen auf Morbidität, Mortalität und Kosten)
3. Organisation der Arbeitsgruppe
4. Auswahl der zu messenden Pflegeaspekte oder Prozesskriterien (Einfluss auf wichtige Ergebnisse, vorhandene Indikatoren)
5. Niederlegung der Messspezifikation (Indikatoren, Einschluss- und Ausschlusskriterien, Risikoadjustierung, Datenquellen, Datensammlung)
6. Durchführung von Tests (Durchführbarkeit, Reliabilität, Validität)
7. Festlegung der Spezifikation für Auswertung der Indikatoren (Bewertung der Messergebnisse, Definition akzeptabler Qualität)

Schneider et al. 2003

Qualität wird üblicherweise differenziert nach Struktur, Prozess und Ergebnis. Hinzugenommen werden sollten Gerechtigkeit und Effizienz. Die Entwicklung von Qualitätsindikatoren kann durch Expertendiskussion, Evidenz basiert (randomisierte kontrollierte Studien) oder durch Ableitung aus Leitlinien, die auf Evidenz und Konsens basieren, erfolgen. In den USA und Großbritannien wird häufig die in den USA entwickelte RAND-Methode eingesetzt.

Wünschenswerte Eigenschaften von Qualitätsindikatoren in der Medizin sind:

- valide
- reliabel
- sensitiv

- praktikabel
- Evidenz gestützt
- Reduzierung der Krankheitslast
- kosteneffektiv
- Indikatortauglichkeit
- beeinflussbar

Eine beispielhafte Liste mit Indikatorsystemen aus der ganzen Welt wird vorgestellt. Anwendungsmöglichkeiten für Qualitätsindikatoren im deutschen Gesundheitswesen werden aufgezeigt: Qualitätszirkel, Disease Management Programme, Arztnetze, Praxisassessment.

Seufert et al. 2000

Evaluierte Qualitätsindikatoren werden in bestehende Dokumentationssysteme als Ergänzung zur „regionalen“/lokalen Qualitätssicherung integriert. Gegenseitige Abhängigkeit zwischen Qualitätssicherung und valider medizinischer Dokumentation erfordert gemeinsame Weiterentwicklung.

Spyra et al. 2002

Vorgestellt wird die Entwicklung von Qualitätsindikatoren für Prozess- und Ergebnisqualität in der kardiologischen Rehabilitation. Verschiedene Instrumentarien für das Qualitätsmanagement in der Kardiologie sind in den letzten Jahren mit Fokus auf Prozessqualität in der kardiologischen Rehabilitation entstanden. In dieser Arbeit wird die Entwicklung risikoadjustierter Outcome-Indikatoren mit praktischer Nutzung für den Klinikvergleich erläutert. Die Klinikvergleiche sollen aus dem Vergleich der Ergebnisqualität Schlussfolgerungen für die Verbesserung der Struktur- und Prozessqualität ermöglichen. Dabei interessiert die Rückführbarkeit der Ergebnisqualität auf wirksame Prozesskomponenten.

Zorn/Ollenschläger 1999

Thema ist die Darstellung eines Entwicklungsschemas für Qualitätsindikatoren. Bewertung von Qualität erfolgt durch Indikatoren und ist grundsätzlich ein Vergleich zwischen Soll- und Ist-Zustand. Leitlinien und Standards liegt eine Vorstellung von Sollwerten zugrunde. Die Definition der Sollwerte erfolgt durch wissenschaftliche Evidenz, Meinung einzelner Experten, Konsensus einer größeren Gruppe oder eine Kombination dieser Möglichkeiten.

Qualität ist lediglich ein theoretisches, vom subjektiven Standpunkt des jeweiligen Betrachters abhängiges Konstrukt und daher nicht direkt messbar. Qualitätsindikatoren sind Hilfsgrößen, die die Qualität einer Einheit durch Zahlen bzw. Zahlenverhältnisse indirekt abbilden, und dienen hauptsächlich der gezielten Lenkung qualitätssichernder Aktivitäten auf auffällige Bereiche. Ein Beispiel aus den USA für Indikatoren aus dem stationären Bereich wird genannt. Eine exemplarische Entwicklung von Indikatoren für den Bereich der ambulanten Versorgung in ausgewählten Anwendungsfeldern erfolgt durch die Zentralstelle der Deutschen Ärzteschaft zur Qualitätssicherung in der Medizin und Begleitung durch wissenschaftliche Fachgesellschaften. Indikatoren müssen u. a. valide und reliabel sein. Ein Ablaufschema für die Indikatorenentwicklung wird am Beispiel des Diabetes mellitus vorgestellt.

2.4.3 Definitionen und Normen

Bak et al. 2004

Die gesetzliche Verpflichtung zur Qualitätssicherung ist in §135 ff SGB V geregelt. Alle Vertragsärzte und zugelassenen Krankenhäuser müssen sich an einrichtungsübergreifenden Maßnahmen zur Qualitätssicherung beteiligen. Das Gesundheitsmodernisierungsgesetz (§§ 135a, 136) verlangt internes Qualitätsmanagement. Die verschiedenen Träger der Rehabilitation haben in der Vergangenheit unterschiedliche Qualitätsmanagement-Systeme implementiert. § 20 Abs. 1 SGB IX verpflichtet Rehabilitationseinrichtungen zur Qualitätssicherung.

Die Autoren suchten in der wissenschaftlichen Literatur im Zeitraum von 1980 bis 2004 nach Zertifizierungs- und Akkreditierungssystemen (als Teil des Qualitätsmanagements). Die Suchstrategie ist angegeben. Fünf der am häufigsten genannten und weltweit verbreiteten, ein europäisches, ein nationales und ein am häufigsten mit der Rehabilitation genanntes System wurden selektiert:

- DIN EN ISO
- JCIA
- JCAHO

- EFQM
- Visitationsmodelle (visitatje, VIP, TOPAS)
- CARF
- KTQ
- DEGEMED

Der Einsatz solcher Systeme im Gesundheitswesen nimmt rapide zu. Eine Teilnahme wird als Nachweis der Qualität der erbrachten Leistung angesehen. Für den Zusammenhang zwischen der Zertifizierung und Akkreditierung von Einrichtungen und den Ergebnissen der Behandlung besteht jedoch eine unzureichende Evidenz.

Bundeszentrale für gesundheitliche Aufklärung 2001

Im Auftrag der Bundeszentrale für gesundheitliche Aufklärung (BZgA) hat die Abteilung Epidemiologie, Sozialmedizin und Gesundheitsforschung der Medizinischen Hochschule Hannover eine Bestandsaufnahme des Qualitätsmanagements in Gesundheitsförderung und Prävention mit Anwendungsbeispielen konzipiert und koordiniert.

Die zwei wichtigsten „Qualitätsphilosophien“ sind folgende:

- Erfüllungsparadigma: Primäres Ziel ist die Erfüllung festgelegter Kriterien und die Einhaltung von Standards. Dieser Ansatz stammt aus der Industrie und wird dazu benutzt, Produktionsprozesse so zu gestalten, dass das gewünschte Produkt herauskommt.
- Optimierungsparadigma: Primäres Ziel ist hier neben der Kriterienerfüllung die ständige Weiterentwicklung von Strukturen, Prozessen und Ergebnissen (Total-Quality-Management). Das Paradigma zeichnet sich im wesentlichen durch folgendes aus:
 - Kontinuierliche Verbesserung
 - Nutzer-/Kundenorientierung
 - Umfassende Konzeption

Für die Bereiche Gesundheitsförderung und Prävention ist das Optimierungsparadigma vorzuziehen.

Die bekannten Qualitätsdimensionen werden hier durch die Assessmentqualität ergänzt, da die im Rahmen einer Prävention vorgenommenen Interventionen mit komplexen sozialen Systemen zu tun haben, die bei der Pla-

nung der Interventionen bekannt sein müssen. Die folgenden Qualitätsparameter sind für den Bereich der Gesundheitsförderung und Prävention relevant:

- **Assessmentqualität:**
 - Abklärung von Bedarf und Bedürfnis
 - Theoretische Grundlagen
 - Erfahrungswerte aus anderen Projekten
 - Kenntnisse über die rechtlichen, kulturellen und politischen Rahmenbedingungen
- **Strukturqualität:**
 - Trägerschaft
 - Organisation
 - Verankerung/Vernetzung
 - Kommunikationsstruktur
 - Qualifikationen der Mitarbeiterinnen und Mitarbeiter
 - Klare Vertragslage/Leistungsauftrag
 - Finanzielle und personelle Ressourcen
- **Prozessqualität:**
 - Projektplanung
 - Feedbackschleifen
 - Qualitätsziele
 - Dokumentation
 - Wohlbefinden aller im Projekt involvierten Personen
 - Umgang mit Konflikten
- **Ergebnisqualität:**
 - Zielerreichung und Wirkung
 - Wirkungen bei der Zielgruppe
 - Günstiges Kosten-Nutzen-Verhältnis
 - Nachhaltigkeit

EN ISO 9001 2000

Die Norm trifft Festlegungen für Anforderungen an Qualitätsmanagementsysteme, welche durch Organisationen für interne Zwecke oder für Zertifizierungs- oder Vertragszwecke verwendet werden können. Diese Norm ist auf die Wirksamkeit des Qualitätsmanagementsystems hinsichtlich der Erfüllung von Kundenanforderungen gerichtet.

EN ISO 9004 2000

Diese Norm gibt Anleitungen für einen im Vergleich zu ISO 9001 erweiterten Bereich von Zielen eines Qualitätsmanagementsystems, um insbesondere die Gesamtleistung, Effizienz und Wirksamkeit einer Organisation

ständig zu verbessern. Die Norm wird als Leitfaden für Organisationen empfohlen, deren oberste Leitung beim Streben nach ständiger Leistungsverbesserung über die Anforderungen von ISO 9001 hinausgehen will. Diese Norm ist jedoch nicht für Zertifizierungs- und Vertragszwecke vorgesehen.

Igl 1999

Der rechtliche Rahmen der Qualitätssicherung in der Medizin wird im Sozialgesetzbuch SGB V vorgegeben. Dort wird Qualitätssicherung als Qualitätssicherung bei der Produktion medizinischer Leistungen definiert. Eine umfassende und systematische Regelung der Qualitätssicherung erfolgt im SGB V nicht und ist wohl auch an dieser Stelle nicht leistbar. Die Leistungserbringer werden in die Pflicht genommen, in eigener Verantwortung qualitätssichernde Maßnahmen durchzuführen. Primär geht es um die interne Qualitätssicherung. Bei der stationären Patientenversorgung wird durch die Beteiligungsrechte von Krankenkassen, Krankenhausgesellschaft u. a. auch eine externe Qualitätssicherung gefordert.

2.4.4 Labor/Analytik

Hüfers 2004

Es handelt sich hier um Vorlesungsfolien zur analytischen Qualitätssicherung. Qualität ist transzendent: universell erkennbar, nicht präzise zu definieren und durch Erwartung empfunden. Qualität ist produktbezogen: präzise messbar, Qualitätsunterschiede werden durch Eigenschaften/Bestandteile eines Produktes quantitativ widerspiegelt. Qualität ist anwenderbezogen: im Auge des Betrachters, verschiedene Konsumenten haben verschiedene Wünsche. Qualität ist prozessbezogen: Einhaltung von Spezifikationen und gute Arbeit führen zu guter Qualität. Qualität ist Preis Nutzen bezogen: Qualität wird durch Kosten und Preise ausgedrückt. Es werden die wichtigsten Begriffe und Normen vorgestellt: ISO 9000 ff (umfassender Ansatz, Prozesse, Kundenorientierung, ständige Verbesserung), ISO 45000 ff (Genauigkeit, Vergleichbarkeit, Kalibrierung und Validierung von Messverfahren), DIN 55350 (Norm mit Definition von Qualität und

Qualitätssicherung, Good Laboratory Practice (GLP) (Organisation, Dokumentation, Archivierung).

GLP ist ein Qualitätssicherungssystem, das sich mit dem organisatorischen Ablauf und den Rahmenbedingungen befasst, unter denen nicht-klinische gesundheits- und umwelt-relevante Sicherheitsprüfungen geplant, durchgeführt und überwacht werden. (Arzneimittel, Kosmetika etc.).

Qualitätsindikatoren sind:

- Richtigkeit (Übereinstimmung des Wertes/Mittelwertes mit einem als richtig akzeptierten Wert)
- Präzision (Ausmaß der Übereinstimmung von Messwerten, qualitatives Maß für Streuung der Ergebnisse)
- Genauigkeit = Richtigkeit + Präzision

Der Autor stellt ein 4-Phasenmodell für die Einführung neuer Verfahren vor: Erstellung und Einführung mit vorbereitender Qualitätssicherung (Qualitätsmerkmale ermitteln, Kenndaten ermitteln, SOP erstellen), Routine-Qualitätssicherung und externe Qualitätssicherung (zur Aufrechterhaltung von Zuverlässigkeit und Vergleichbarkeit – Probenahme, Plausibilitätsprüfungen, Prüfung auf systematische/zufällige Fehler, Referenzproben).

An Ringversuchen sollten mindestens acht Labore teilnehmen und vier Parallelbestimmungen je Probe (z. B. Standardlösungen oder natürliche Proben) durchgeführt werden.

Kühn-Stoffers 2004

Dieses Vorlesungsskript befasst sich mit der Qualitätssicherung in der Analytik. Der Qualitätsbegriff wird nach der Norm DIN 55350 (Begriffe zu Qualitätsmanagement und Statistik) definiert. Ziel der Qualitätssicherung sind richtige, vergleichbare, justiziable Messergebnisse. Qualitätssicherung wird für Analyseverfahren (Validierung) sowie für Personal, Ausstattung und Organisation (Zertifizierung nach DIN EN ISO 17025 oder Anerkennung als GLP-Labor) durchgeführt. Qualitätsindikatoren sind:

- Präzision (Maß für Streuung, Standardabweichung oder Variationskoeffizient) als Wiederholpräzision/Wiederholbarkeit (mit identischen Bedingungen) bzw. Vergleichs-

präzision/Reproduzierbarkeit (mit verschiedenen Labors, Geräten und Personen) – Streuungsparameter

- Richtigkeit (Übereinstimmung des Wertes mit der wahren Konzentration bzw. mit einem als richtig bekannten Wert – Lageparameter)
- Genauigkeit = Summe aus Präzision und Richtigkeit

Die Validierung umfasst die Beschreibung der Anforderungen, die Bestimmung von Verfahrenskenngrößen, den Vergleich Anforderungen/Kenngrößen und Aussagen über die Methodeneignung für die Fragestellung.

Die Qualitätssicherung erfolgt in der Routine durch:

- Kontrollkarten (Ziel mit maximal zulässigen Abweichungen, Auftragen der Messwerte) als MW-Kontrollkarte (Präzision prüfen), Blindwert-Kontrollkarte (Verunreinigungen finden), Wiederfindungsraten-Kontrollkarte (Richtigkeit prüfen), Spannweiten-Kontrollkarte (Präzision prüfen)
- Ringversuche (dienen nicht nur der Qualitätssicherung, sondern auch der Verfahrensentwicklung und -normung sowie der Zertifizierung von Referenzmaterialien) – Infos in IRIS und EPTIS im WWW
- Teilnahme an Audits (intern/extern, horizontal/vertikal)

Die Dokumentation wird durch SOPs (detailliert, klar verständlich, eindeutig, relevante Störungen und Maßnahmen zur Beseitigung angeben) und Qualitätsmanagement-Handbuch (stellt Qualitätssicherungs-Maßnahmen geordnet und systematisch dar, dient der Übersicht und Orientierung für Mitarbeiter und Kunden, erfüllt Anforderungen der einschlägigen Norm, muss regelmäßig aktualisiert werden) geregelt und geschieht durch Archivierung von Qualitätssicherungs-Dokumenten, Rohdaten, Prüfberichten etc.

Normen geben anerkannte Regeln der Technik vor, setzen klare und transparente Qualitätsstandards und harmonisieren. Der Autor stellt die für die Analytik-Qualität relevanten Normen vor (DIN EN-ISO 17025 und GLP). DIN EN-ISO 17025 beschreibt Akkreditierung von Qualitätsmanagement-System.

GLP regelt Anforderungen an nicht-klinische gesundheits- und umweltrelevante Prüfungen aus insgesamt neun Kategorien. Ziele der GLP sind: Daten angemessener Qualität erzeugen, Untersuchungen durch lückenlose Dokumentation nachvollziehbar machen, Organisation und Verantwortlichkeiten transparent machen, Daten international vergleichbar und anerkennbar machen. GLP umfasst die Bereiche Organisation und Personal (Leitung, Prüfleitung, sonstiges Personal), das Qualitätssicherungs-Programm, alle Materialien und Geräte, Referenzsubstanzen, SOPs, Abläufe, Berichte und Archivierung. Es gibt ein GLP-Anerkennungsverfahren nach Durchführung einer Inspektion; alle vier Jahre erfolgt für anerkannte Organisationen eine Folgeinspektion. GLP bringt aber viel Papier, kostet Zeit und Geld und bewertet nicht, ob eine Methode richtig ist, sondern nur die Nachvollziehbarkeit und Lückenlosigkeit der Dokumentation!

2.4.5 Industrie

2.4.5.1 Statistische Methoden

American Society for Quality o. J.

Verschiedene statistische Methoden können im Rahmen einer Qualitätsverbesserung eingesetzt werden:

- Hypothesentestung (Durchführung von Experimenten)
- Regressionsanalyse
- Statistical Process Control (SPC): Entdeckung ungewöhnlicher Variation als Hinweis auf außer Kontrolle geratenen Prozess
- ANOVA (Fisher's analysis of variance)
Faktoren, die die Streuung von Messsystemen beeinflussen, können mittels der GR & R-Technik (gage repeatability and reproducibility) untersucht werden. Sie ist auf die meisten produktionsbezogenen Messsysteme anwendbar und kann zu folgenden Zwecken genutzt werden:
 - Beurteilung neuer Messgeräte
 - Vergleich zwischen Messgeräten
 - Mittel zur Leistungsverbesserung bei Messinstrumenten

- Vergleich von reparierten Messgeräten mit dem Zustand vor der Reparatur
- Hilfsmittel für die Berechnung von Prozessabweichungen und die Festlegung einer Akzeptanzebene für einen Produktionsprozess
- Messinstrument zur Ermittlung des Trainingsbedarfs für die Nutzung von Messinstrumenten

DeVor et al. 1992

In diesem Buch werden die Konzepte und statistischen Methoden zur Qualitätslenkung detailliert beschrieben.

EMA 2005b

Die ICH Q9 „Quality Risk Management“ beschreibt Prinzipien des Qualitätsrisiko-Managements, die auf alle Aspekte pharmazeutischer Qualität angewandt werden können, und wird nach Verabschiedung als Anhang in den GMP (Good Manufacturing Practice) Guide eingefügt. Das Management von Qualitätsrisiken ist ein systematischer Prozess, der über den gesamten Lebenszyklus eines Medikamentes wirksam ist. Er besteht aus folgenden Komponenten:

- Risikoabschätzung
 - Risikoidentifikation: was könnte schief gehen?
 - Risikoanalyse: wie hoch sind die Wahrscheinlichkeit, dass es schief geht, und die Chance, dies zu entdecken? Was sind die Konsequenzen (Schweregrad)?
 - Risikoevaluation: Vergleich der identifizierten und analysierten Risiken mit vorgegeben Risikokriterien; Ermittlung der Evidenzstärke
- Risikosteuerung
 - Risikoreduktion: Prozesse zur Verminderung oder Vermeidung von Risiken, die ein akzeptables Niveau überschreiten
 - Risikoakzeptanz: Implizite oder explizite Entscheidung, bestimmte Risiken zu tolerieren
- Risikokommunikation: Informationsaustausch über die Risiken und ihr Management zwischen Entscheidungsträgern und anderen (z. B. Zulassungsbehörden, In-

dustrie, Patienten, Mitarbeiter des eigenen Unternehmens)

- Risikoreview: Review der Ergebnisse des Risikomanagement-Prozesses, Berücksichtigung neuer Erkenntnisse und Erfahrungen

Verschiedene Verfahren können das Risikomanagement unterstützen, z. B.:

- Hazard Analysis and Critical Control Points (HACCP)
- Hazard Operability Analysis (HAZOP)
- Failure Mode Effects Analysis (FMEA)
- Failure Mode, Effects and Criticality Analysis (FMECA)
- Fault Tree Analysis (FTA)
- Preliminary Hazard Analysis (PHA)
- Risikogewichtung und –ranglisten
- Statistische Werkzeuge (z. B. Pareto-Charts)

Kaomea 1994

Der Autor beschreibt ein statistisches Modell zur Bewertung der Datenqualität in einem vorgegebenen Entscheidungsszenario. Das Modell basiert auf der Prämisse, dass der Wert von Daten eine Funktion ihres Inhaltes, Kontextes und Qualität ist, und analysiert Entscheidungsbäume. Ein bestimmter Dateninhalt, der Einfluss auf ein unsicheres Ereignis (Knoten im Entscheidungsbaum) hat, beeinflusst die Wahrscheinlichkeit, dass dieses Ereignis eintritt. Dies kann sich auf die optimale Entscheidung bzw. das optimale Ergebnis des Entscheidungsprozesses auswirken. Wenn der Kontext korrekt modelliert wird, kann die Analyse den Effekt der Daten auf Entscheidungen und erwartete Ergebnisse zeigen.

Montgomery 1991

Die statistischen Methoden zur Qualitätssicherung, statistische Verfahren zur Prozesslenkung sowie Stichprobenverfahren werden in diesem Buch detailliert vorgestellt.

Weih/Jessenberger 1999

Dieses Buch stellt statistische Methoden zur Qualitätssicherung und Qualitätsoptimierung in der Industrie vor.

2.4.5.2 Six Sigma

Benedetto 2003

Der Autor berichtet über ein Projekt aus der Radiologie. Es wird zwischen evolutionären und revolutionären Methoden zur Leistungssteigerung unterschieden. Evolutionäre Methoden sind z. B. Qualitätszirkel, CQI und TQM. Sie sind geeignet, wenn man mäßige Verbesserungen sucht und das Problem nicht abteilungsübergreifend ist. Revolutionäre Methoden wie z. B. Six Sigma und Reengineering sind für drastische Veränderungen und abteilungsübergreifende Probleme geeignet.

Six Sigma setzt voraus, dass die Unternehmenskultur datengetriebene (statt politisch motivierte) Entscheidungen zulässt. Als Grundlage für das Six Sigma-Projekt sollte ein ausführlicher Problembeschreibung dienen. Wo es um Daten geht, sollte ein Experte aus der IT-Abteilung hinzugezogen werden, insbesondere in der Phase der Messungen und Analysen. Die Datenbasis sollte die Antworten auf relevante Fragen ermöglichen. Es ist sinnvoll, dass das leitende Personal in Six Sigma geschult wird. Dabei sollte der obersten Hierarchie-Ebene sowie den Führungskräften auf Abteilungs- und Bereichsleiter-Ebene klar werden, wie wichtig ihr Engagement und ihre Unterstützung (auch für unpopuläre Entscheidungen) sind. Eine wichtige Voraussetzung für Six Sigma-Projekte ist eine Mentalität, die die Bereitschaft zu Veränderungen umfasst. Six Sigma-Projekte bringen immer Veränderung, meist ist diese umfassend und gravierend und häufig widersetzt sich das Personal den Veränderungen.

Die fünf Phasen des Six Sigma-Prozesses lassen sich mit dem Akronym DMAIC beschreiben: Define (Definition von Schlüsselproblemen), Measure (quantitative Daten sammeln, um die aktuelle Leistung zu beschreiben), Analyze (herausfinden, welche Prozess-Schritte geändert werden müssen), Improve (Veränderungen planen und implementieren), Control (Prozess- und Qualitätskontrolle implementieren). Der Artikel beschreibt in erster Linie den Projektbeginn bis zur Define-Phase.

Zur Define-Phase gehört die Erstellung einer Projektbeschreibung. Diese sollte enthalten:

operationalisierte Projektziele, sowohl quantifiziert als auch laienverständlich beschrieben; geschätzte Zeitplanung; verfügbare Ressourcen; informed consent der Abteilungsleiter. Die zweite wichtige Aufgabe der Define-Phase ist die Auswahl von Messgrößen und Beschaffung von Basisdaten für diese Messgrößen. Danach werden Mini-Projektteams für die jeweiligen Einzelprojekte gebildet. Der Artikel beschreibt die notwendigen Schritte zur Zusammensetzung, Gründung und Arbeit dieser Teams. Zu Beginn des Projekts sollte die Kultur der Veränderung auch den Angestellten und ggf. Betroffenen außerhalb der Abteilung vermittelt werden.

Als mögliche Fallstricke benennt der Autor Schwierigkeiten bei der Erhebung der Daten (z. B. durch nicht standardisierte Datendefinitionen oder unzureichende Dokumentation des Datenerhebungs- und Auswertungsprozesses), bei der Durchführung von Befragungen zur Zufriedenheit und der Archivierung der Rohdaten. Desweiteren kann es zu Problemen führen, wenn die Projektteams mit unzureichend qualifizierten Mitarbeitern besetzt sind oder keine Strategien zum Auffangen von Personalfuktuationen vorgesehen sind. Die Team-Mitglieder sollten in Statistik geschult und zur Anwesenheit bei den Treffen motiviert werden. Der Projektmanager muss eine ausreichend hohe Position in der Hierarchie haben. Mitarbeiter müssen auch über die längere Projektlaufzeit hinweg ausreichend motiviert werden. Bzgl. unangenehmer Entscheidungen sollte die Wahrheit gesagt werden. Kommunikation ist einer der wichtigsten Aspekte eines Six Sigma-Projektes. „Rivalitäten“ zwischen Abteilungen bzw. Unterabteilungen sollten durch entsprechende Teamarbeit beseitigt werden. Die Zusammenarbeit mit der IT-Abteilung ist von besonderer Bedeutung. Ggf. müssen Fehler im IT-System korrigiert werden. Es ist zu erwarten, dass relativ viel Anstrengung in die Überarbeitung von SOPs investiert werden muss. Das Six Sigma-Projekt sollte zum strategischen Plan der Abteilung passen. Alle Besprechungen, Entscheidungen etc. sollten ausführlich dokumentiert und alle Betroffenen, internen und externen Kunden regelmäßig informiert werden.

Lucas 2002

Six Sigma wurde ursprünglich in den 80er Jahren des 20. Jahrhunderts bei Motorola als Methodologie, die der disziplinierten Qualitätsverbesserung dient, und Teil von TQM entwickelt. Six Sigma setzt eine modifizierte Version des Shewhart-Zyklus (plan-do-check-act) ein und ist stark projektbezogen, ist eher ein Geschäftssystem und auf der Management-Ebene und nicht in der Technik angesiedelt.

Ziel ist es, praktisch alle Fehler zu eliminieren, d. h. eine Reduktion auf 3,4 Fehler pro Million Fehlergelegenheiten (DPMO) = Six Sigma. Zur Zielsetzung wird ein metrisches System verwendet, das die Variablen umfasst, die CTQ (critical to quality) sind, unabhängig von ihrer Verteilung.

Die Umsetzung von Six Sigma kann auf verschiedene Weise geschehen:

- Breakthrough-Strategie: Ein vierschrüttiger Verbesserungsprozess (messen, analysieren, verbessern, kontrollieren) wird durchgeführt. Häufig wird vorher noch ein Definitionsschritt (der CTQ-Variablen) eingefügt. Die Verbesserungsstrategien müssen zum allgemeinen Unternehmensziel passen.
- Top-Down-Strategie: Das Management überwacht alle Prozesse. Sie werden nach dem Teile-und-Herrsche-Prinzip in Projekte aufgebrochen. Der CEO (Chief Executive Officer) ist oberste Instanz und für jedes Projekt wird ein Champion als Erfolgsverantwortlicher benannt (und finanziell entsprechend entlohnt). Der Projektleiter wird als „Black Belt“ bezeichnet und leitet typischerweise innerhalb von zwei Jahren acht bis zwölf Projekte aus verschiedenen Geschäftsbereichen. Die Projektteam-Mitglieder heißen „Green Belts“. Ausbildung und Unterstützung erfolgt durch erfahrene „Master Black Belts“. (Faustregel: 1 Black Belt pro 100 Angestellte, 1 Master Black Belt pro 100 Black Belts). In der Ausbildung lernen sie statistische Grundlagen, Design von Experimenten, Qualitätskontrolle und Projektmanagement. Nach ca. zwei erfolgreichen Projekten werden sie zertifiziert, Master Black Belts nach 20 erfolgreichen Projekten.

Alle Six Sigma-Projekte werden rigoros auf ihre finanzielle Wirksamkeit geprüft. Jedes Projekt soll eine Ersparnis von ca. 175.000 Dollar bringen.

2.4.5.3 TQM

Favre 2003

Bei dieser Arbeit handelt es sich um eine Dissertation zum Total Quality Management (TQM) im Bankwesen. Hier gilt folgende Definition: Qualität ist die permanente Erfüllung aller vereinbarten Kundenanforderungen. Das Qualitätsmanagement muss ganzheitlich die unternehmensinternen Voraussetzungen für Qualität schaffen. Nach Seghezzi hat es 5 Funktionen: Qualitätsplanung, Qualitätslenkung, Qualitätssicherung, Qualitätsförderung, Qualitätsführung. Operationalisierbare Ansatzpunkte finden sich in allen Unternehmensbereichen. Nach dem EFQM (TQM-Modell der European Foundation for Quality Management) sind Qualitätsdimensionen: Qualität der Leistungen, der Prozesse, der Ressourcen, der Mitarbeiter, der Führung.

TQM strebt gleichzeitige Optimierung von Qualität, Kosten und Zeit an und ist eine Führungsphilosophie, die Qualität in den Mittelpunkt stellt. Prinzipien: Quality first, über Qualität entscheidet nur der Kunde, Kundenorientierung, Einbeziehung und Mitwirkung aller Mitarbeiter, Prozessorientierung, systematische und permanente Qualitätsverbesserungen. An diese Prinzipien müssen Unternehmenskultur, -organisation und -prozesse angepasst werden.

Das EFQM-Modell umfasst neun Bewertungskriterien (Führung, Personal, Politik und Strategie, Partnerschaften und Ressourcen, Prozesse, Personalergebnisse, Kundenergebnisse, Gesellschafts-Ergebnisse, Schlüsselergebnisse in der Performanz). Die Resultate sollen eindeutig auf die ergriffenen TQM-Maßnahmen zurückführbar sein, anhand des Modells wird der Realisierungsgrad im Unternehmen beurteilt (RADAR-Modell: Results, Approach, Deployment, Assessment, Review). D. h., dass ein Unternehmen die zu erzielenden Ergebnisse (Results) bestimmt, Vorgehensweisen (Approach) zu ihrer Erreichung entwickelt, diese systematisch umsetzt (Deployment) und

ihre Überprüfung und Bewertung (Assessment und Review) vornimmt. Das Modell ist eher theoretisch und wenig detailliert/operationalisiert.

Im Dienstleistungsbereich ist TQM nicht so weit verbreitet wie in der produzierenden Industrie. Besonderheiten von Dienstleistungen sind ihre Immaterialität, die Notwendigkeit der Leistungsfähigkeit und der interaktive Charakter. Sie können potential-, prozess- oder ergebnisorientiert betrachtet werden. Die Beurteilung der Servicequalität erfolgt in fünf Dimensionen:

- Annehmlichkeit des konkreten Umfelds (tangibles)
- Zuverlässigkeit (reliability)
- Reaktionsfähigkeit (responsiveness)
- Leistungskompetenz (assurance)
- Einfühlungsvermögen (empathy)

Zur Messung der Dienstleistungsqualität werden vor allem Befragung und Beobachtung des Kunden eingesetzt. Der Autor schlägt vor, das TQM-Prinzip der Prozessorientierung um ein Element der Interaktionsorientierung zu ergänzen. Das Humankapital ist die entscheidende Ressource, daher müssen die Mitarbeiterzufriedenheit gestärkt, die Mitarbeiter motiviert und ihre Kompetenzen gefördert werden. Kommunikation und Information sind von großer Bedeutung. Die Kommunikation mit dem Kunden soll informieren und profilieren, motivieren, Leistungserwartungen in Erfahrung bringen und die Corporate Identity vermitteln.

Die Arbeit stellt die Umsetzung des TQM in Banken dar und die Ergebnisse einer empirischen Studie mit Befragung von ausgewählten Mitarbeitern in Schweizer Banken. Die untersuchten Banken gehen von einer kundenorientierten Qualitätsdefinition aus, die Meinungen zur ISO-Zertifizierung sind unterschiedlich. Die TQM-Anwendung folgt dort, wo sie betrieben wird, den Anforderungen der Theorie. TQM ist jedoch schwierig umzusetzen, es fehlt oft das nötige Know-How und Engagement. Die Anforderungen werden unterschätzt und das Top Management steht nicht dahinter.

2.4.5.4 Failure Mode and Effects Analysis (FMEA)

Helbling/Weinbeck 2004

Zur Bewertung von technischen Risiken wird FMEA (Failure Mode and Effects Analysis = Fehlermöglichkeits- und Einflussanalyse, in DIN 25448 = Ausfalleffektanalyse) häufig verwendet. Aufgrund der Eintrittswahrscheinlichkeit und der Tragweite wird einem Fehler (Abweichung vom Soll oder Handeln mit negativen Konsequenzen) eine

- Risikoprioritätszahl (Auftrittswahrscheinlichkeit * Bedeutung * Entdeckungswahrscheinlichkeit = RPZ Faktoren als natürliche Zahl zwischen 1 und 10, RPZ zwischen 1 und 1.000, je höher der Wert, desto höher die Priorität)

zugeordnet, die sich durch präventive Maßnahmen verändern lässt. Aus der Lebensmittelindustrie stammt die HACCP (Hazard Analysis and Critical Control Points), die die kritischen Punkte eines Prozesses ermittelt.

Die Autoren prüfen die Realisierbarkeit einer Human-FMEA, da sich viele Fehler auf menschliches Versagen zurückführen lassen, aber die Ursachen nicht so einfach bewertet werden können, wie dies bei technischen Fehlern der Fall ist.

Als Qualitätsmanagement-Werkzeuge lassen sich IRM (Incident Risk Management = Lernen aus vergangenen Fehlern, indem bereits kleine Fehler erkannt, gemeldet und behoben werden) und Demingkreis (kontinuierlicher Verbesserungsprozess begleitet die gesamte Produktion) einsetzen.

Relevante Fragen sind: lässt sich der menschliche Faktor in die technisch orientierte FMEA transferieren, welches sind die täglichen Einflüsse, bis wo sind Maßnahmen des Unternehmens ohne Eingriff in die Privatsphäre möglich.

Handlungsfehler werden als Nichterreichung eines Ziels definiert. Es werden verschiedene Grundlagen menschlicher Handlungsfehler vorgestellt. Aus der Handlungsregulationstheorie, die zwischen verschiedenen Ebenen unterschiedlicher Bewusstheit unterscheidet, wird abgeleitet, dass eine bewusste Fehlerwahrnehmung nur auf der obersten (intellektuellen) Ebene möglich ist. Fehler werden in Planungs-, Ausführungs- und Kontrollfehler

kategorisiert und diese weiter unterteilt. Hinzu kommen Fehler, die durch den zwischenmenschlichen Umgang entstehen. Desweiteren werden Fehler bestimmten Phasen des Arbeitsprozesses zugeordnet. Als Einflussfaktoren werden interne wie externe Faktoren aufgeführt. Das Rollenverhalten wird je nach verschiedenen Persönlichkeiten differenziert und menschliche Leistungsmerkmale (z. B. Konzentration, Motivation) werden mit fördernden und hinderlichen Einflussfaktoren angegeben. Mit Hilfe von Ishikawa-Diagrammen werden Ursachen-Wirkungsanalysen für individuelle und unternehmensspezifische Faktoren dargestellt.

Weitere Methoden zur Fehlerbewertung sind:

- Risikomatrix: Eintrittswahrscheinlichkeit und Schadenpotential werden multipliziert und können in einer Risikomatrix grafisch dargestellt werden (grün-gelb-rot je nach angemessenen Reaktionen).
- Failure Tree Analysis (FTA): Der definitive Systemzustand ist top event. Ist die Hauptgefahr eruiert, werden mögliche Ursachen von oben nach unten zu ihren Basisereignissen zurückverfolgt. Qualitativ sollen alle möglichen Ausfall- und Ursachenkombinationen identifiziert werden, quantitativ ist die Ermittlung von Eintrittswahrscheinlichkeiten differenzierter Ausfallkombinationen von Bedeutung.
- Ishikawa-Diagramm (Fischdiagramm): Das Problem wird am Kopf des Fisches eingetragen, an den Enden der „Gräten“ die Haupteinflussgrößen, an den „Gräten“ selbst die zugehörigen Ursachen.
- Root Cause Analysis (RCA): RCA geht von einem eingetroffenen Ereignis aus und sucht die Ursachen. Auf eine Problemdefinition und Datensammlung folgt eine Task-Analyse (Aufteilung in Einzelschritte), die Differenz-Analyse (Ist vs. Soll), die Hindernis- oder Barrierenanalyse, Ereignis- und Ursachenfaktorgrafik (als Flussdiagramm), Interviews (mit beteiligten Personen) und Ursachenanalyse, Korrekturmaßnahmen und Abschluss (durch einen Bericht).

Die FMEA ist nicht auf reaktive Fehlererkennung und -korrektur ausgelegt, sondern auf

Prävention. Daher müssen so früh wie möglich Fehlerursachen definiert werden. Die Analyse sollte von einem multifunktionalen Team aus verschiedenen Unternehmensbereichen durchgeführt werden. Zu Beginn wird das betrachtete System eingegrenzt und exakt strukturiert. Alle Elemente werden mit ihren Funktionen dargelegt und auf potentielle Fehlerursachen, -arten und -folgen untersucht. Dann wird die Risikobewertung durchgeführt und die RPZ ermittelt. Liegt dieser Wert über 125, so werden Verbesserungsmaßnahmen eingeleitet.

Die RPZ wird aus den einzelnen Komponenten ermittelt:

- Bedeutung: Sehr gering bis Sehr hoch, 5 Stufen, Wert 1–10
- Auftretenswahrscheinlichkeit: Sehr gering bis Sehr hoch, 5 Stufen, Wert 1–10
- Entdeckungswahrscheinlichkeit: Sehr gering bis Sehr hoch, 5 Stufen, Wert 1–10

Die Bedeutung kann nicht beeinflusst werden, und von den beiden Wahrscheinlichkeiten ist die Auftretenswahrscheinlichkeit interessanter (weil eine Verbesserung der Entdeckungswahrscheinlichkeit zwar mehr Fehler aufdeckt, diese jedoch nicht korrigiert).

Es gibt vier Arten von FMEA: System-, Konstruktions-, Prozess- und Gefährdungs-FMEA. Als Tools werden vor allem Formulare und Software (z. B. APIS) eingesetzt.

Die Autoren stellen ihren Ansatz anhand von je einem Beispiel aus der Industrie (Wartungsfehler Maschine) und dem Dienstleistungsbereich (Zustellungsfehler E-Mail) anhand von Fehlerbäumen und Ishikawa-Diagrammen dar. Das klassische FMEA-Formblatt, das Fehlerfolge, Fehler und Fehlerursache umfasst, wird um mögliche menschliche Einflussfaktoren erweitert (involvierter Arbeitsprozess, Handlungsphase). Die RPZ werden vor und nach Verbesserungsmaßnahmen bestimmt. Maßnahmen sollten praxisnah sein und eine messbare Verbesserung bringen. Die Umsetzung der FMEA kann proaktiv wie reaktiv erfolgen. Die Autoren stellen den möglichen Nutzen vor. Durch Schaffung einer Vertrauenskultur im Unternehmen wird FMEA ermöglicht.

2.4.5.5 Softwarequalität

American Society for Quality o. J.

Zwei Ansätze zur Handhabung von Software-Qualität sind weit verbreitet:

- Defect management approach (Fehlermanagement): Fehler werden gezählt und nach Schweregrad klassifiziert. Die Klassenbelegung wird als Planungsgrundlage verwendet oder es werden defect linkage matrices (Auftragen der Anzahl der Entwicklungsphasen, in denen der Fehler bis zu seiner Entdeckung übersehen wurde) analysiert.
- Quality attributes approach: Dieser Ansatz wird durch sechs Qualitätscharakteristika beschrieben:
 - Functionality
 - Reliability
 - Usability
 - Efficiency
 - Maintainability
 - Portability

Bobrowski et al. 1999

Im Bereich des Software Engineerings wurden die in einem System gespeicherten Daten lange ignoriert. Validations- und Verifikations-techniken fokussierten auf die Software. Die Autoren vertreten die Meinung, dass viele Probleme der Datenqualität auf schlechte Praktiken des Software Engineerings zurückzuführen sind und fordern die Einbeziehung der Datenqualitätsaspekte in das Software Engineering.

Es gibt keine universelle Definition von Qualität. Datenqualität umfasst die folgenden Dimensionen:

- completeness
- relevance
- reliability
- consistency
- correctness
- timeliness
- precision
- conciseness

Ein gutes Ergebnis in einer Dimension allein ist kein Nachweis guter Datenqualität. Die Dimensionen lassen sich danach klassifizieren, ob sie direkt messbar sind (completeness, relevance, consistency, correctness, timeliness,

precision) oder ob ihr Wert von anderen Dimensionen abhängt (conciseness, reliability). Der aus dem Bereich der Software-Metrik stammende Ansatz der Goal-Question-Metrik (GQM) kann zur Messung der Datenqualität eingesetzt werden. GQM ist ein Rahmenwerk für die Definition von Metriken und besteht aus folgenden Teilen:

- Festlegung von Zielen
- Formulierung von Fragen, die die relevanten Attribute der Ziele adressieren
- Festlegung von Messmethoden zur Beantwortung der Fragen

Ein Ziel ist präzise definiert durch das Objekt, auf das es sich richtet, mit einem Zweck, von einer bestimmten Perspektive aus in einer festgelegten Umgebung betrachtet. Eine Metrik besteht aus einer Ansammlung von (subjektiven oder objektiven) Daten, die eine Frage quantitativ beantworten.

Mathieu/Khalil 1998

Schlechte Datenqualität ist pervasiv und hat einen signifikanten negativen Einfluss auf den Geschäftserfolg. Die Experten, die für die Informationssysteme in Unternehmen verantwortlich sind, haben nur selten eine formale Ausbildung in Techniken zur Verbesserung der Datenqualität erhalten. Daraus ergibt sich die Notwendigkeit, die Thematik der Datenqualität in Kurse über Datenbanksysteme zu integrieren. Aus der Literatur leiten die Autoren Dimensionen der Datenqualität ab:

- Lebenszyklus der Daten: Datenakquisition
 - Definition der Sicht (logisches Datenbankdesign)
 - Implementierung der Sicht (physikalisches Datenbankdesign)
 - Beschaffung der Daten
 - Aktualisierung der Datensätze
- Lebenszyklus der Daten: Datennutzung
 - Definition von Sichten
 - Datenabfrage
 - Datenmanipulation
 - Präsentation der Ergebnisse
- Prozessmanagement
 - Geschäftsprozesse
 - Design und Modellierung von Prozessen

- Managementaufgaben (z. B. Verantwortlichkeiten)
- Transaktionsverarbeitung
- Qualitätskontrolle der Daten
 - Datenqualitätsmetriken
 - Managementaufgaben (Planung und Administration der Datenqualitätsfunktionen)
 - Qualitätskontrolle der Daten
 - Datenbanksicherheit

Aus der Analyse von fünf weit verbreiteten Fachbüchern zum Thema Datenbanksysteme leiten die Autoren die Empfehlung ab, folgende Module in Kurse über Datenbanksysteme zu integrieren:

- die Wichtigkeit der Datenqualität und die Rolle von Geschäftsprozessen für die Datenqualität
- die Rolle der Datenqualität im Lebenszyklus der Daten
- Messung, Bewertung, Verfolgung und Verbesserung von Datenqualität
- Tipps für die Verbesserung der Datenqualität im Datenbankdesign und Testen

Storey/Wang 1998

Qualitätsaspekte von Daten in einer Datenbank sind aus der Perspektive des Nutzers wichtig. In der Literatur, die sich mit Datenbank-Design beschäftigt, wurde dieser Aspekt bisher vernachlässigt. Die Autoren beschreiben einen Ansatz, der Datenqualitätsanforderungen in das logische Design einer Datenbank integriert und zeigen seine Anwendung beim Design einer kleinen Beispieldatenbank. Zur formalen Darstellung wird hier das Entity-Relationship-Modell (ER-Modell) benutzt.

Die Anforderungen der Nutzer lassen sich in drei Kategorien aufteilen:

- Anforderungen an die Anwendung (z. B. Definition des Datenobjekts „Kunde“)
- Anforderungen an die Qualität der Anwendung (verbunden mit der Erstellung eines Produktes)
- Anforderungen an die Datenqualität der zu speichernden Daten

Die Qualitätsanforderungen werden durch den Datenbankdesigner in Zusammenarbeit mit den Nutzern identifiziert und modelliert. Die Anforderungen an die Datenqualität werden separat modelliert, da sich die Datenqua-

lilität auf die Qualität von Attributwerten bezieht. Dazu führen die Autoren zwei Typen von Datenqualitäts-Entitäten ein:

- Datenqualitätsdimensions-Entität: aktueller Wert einer Qualitätsdimension
- Datenqualitätsmaß-Entität: Interpretation aller möglichen Werte einer Qualitätsdimension

Da das ER-Modell keinen direkten Mechanismus für die Zuordnung eines Attributs einer Entität zu einer anderen Entität enthält, wird eine Attributverlaufsrepräsentation eingeführt.

2.4.6 Sonstige Bereiche

Mettin 2002

Es handelt sich hier um einen Bericht zum anuba-Modellversuch (Aufbau und Nutzung von Bildungsnetzwerken zur Entwicklung und Erprobung von Ausbildungsmodulen in IT- und Medienberufen) Bildungsnetzwerke bewerten: Qualitäts- und Risikomanagement in Bildungsnetzwerken. Der Ruf eines Unternehmens und sein wirtschaftliches Überleben hängen vom Faktor Qualität ab. Der Qualitätsbegriff wird in der Norm ISO 8402 festgelegt. Durch Zeitverzögerung von Entwicklung und Produktion treten Entwicklungsfehler mit zeitlichem Abstand auf. Durch TQM ist Qualität Teil einer Unternehmensstrategie und -kultur. Qualitätsmanagement-Ansätze fußen in der Regel auf dem Deming-Kreis (Plan-Do-Check-Act). Plan = Qualitätsplanung, Do = Qualitätslenkung, Check = Qualitätssicherung, Act = Qualitätsverbesserung. Das Modell von Seghezzi und Hansen misst Defizite zwischen idealer (vom Kunden erwarteter und wahrgenommener Qualität) und produzierter Qualität mit Hilfe von zehn Delta-Variablen, die alle möglichst Null sein sollten.

Zur Qualitätsplanung wird in einem Lastenheft beschrieben, was der Kunde will. Die Anforderungen an das Produkt gehen ins Pflichtenheft ein. Qualitätsanforderungen werden bei Gestaltung und Entwicklung festgelegt, für Prozesse gibt es ein Prozesspflichtenheft. Dann erfolgt die Prozessplanung und -entwicklung, die Produkte müssen verifiziert und validiert werden. Es gibt zwei Arten von

Nichtkonformität zwischen Soll und Ist: Qualitätsdefizit, wenn Erwartungen nicht erfüllt werden und wenn der Ist-Zustand Merkmale aufweist, die der Kunde gar nicht erwartet. Ziele der Qualitätsplanung sind Treffsicherheit (Übereinstimmung mit Kundenerwartungen), Differenzierung (Wettbewerbsvergleich), Robustheit (gegenüber absehbaren Störungen) und Fehlerfreiheit. Als Methoden der Qualitätsplanung gibt es QFD (Quality Function Deployment, Teamorientierung und Einsatz von Qualitätstabellen, streng kundenorientiert), FMEA (Fehlermöglichkeits- und Einflussanalyse, systematische Vorgehensweise) und Poka Yoke (technische Vorkehrungen, die Fehler am Produkt durch menschliche Fehlhandlungen möglichst vermeiden sollen).

Die Qualitätslenkung kann in mittelbar und unmittelbar eingeteilt werden, Ziele sind die Einhaltung von Standards/Spezifikationen und die Beherrschung der Prozessqualität. Die Aufgaben umfassen die Strukturierung der entscheidenden Prozesse, Maßnahmen zum Erreichen der Konformität, Messen der Qualität, Festlegung der Verantwortung für das Messen (Fremd- und/oder Selbstprüfung) und Arbeiten in Regelkreisen.

Qualitätssicherung umfasst u. a. strukturelle und verhaltensbezogene Maßnahmen. Ziele sind die Schaffung von Vertrauen beim Kunden und die Sicherstellung, dass Fehler erkannt, behoben und in ihrer Wirkung beherrscht werden. Zu den Methoden zählen Bilanz für Fehler, Design Reviews, Audits und Frühwarnsysteme. Bei der Planung von Qualitätssicherungs-Maßnahmen sollte der Umfang an der Risiko-Höhe festgemacht werden (Risikomanagement). Diese setzt sich zusammen aus der Wahrscheinlichkeit des Auftretens eines Fehlers, der Wahrscheinlichkeit, dass dieser Fehler nicht rechtzeitig entdeckt wird und der wahrscheinlichen Tragweite des Fehlers. Risikomanagement nach Tiemeyer umfasst die Identifikation, Bewertung, Klassifizierung und Behandlung von Risiken. Zur Identifikation dienen Checklisten (Risiko bei Planung, Terminen, Kosten, Personal, Akzeptanz, Projektorganisation, Vorgehen und Methodik). Zur Bewertung und Klassifizierung dienen z. B. Matrizen. Die Behandlung um-

fasst Vorsorge, Eventualmaßnahmen, Verringerung und Beherrschung von Risiken.

Naumann/Rolker 2000

Eine schlechte Informationsqualität ist ein großes Problem für die Nutzer von Informationen, die in Internet oder Intranet bereitgestellt werden. Die Autoren bemängeln das Fehlen geeigneter Konzepte und Methoden zur Bestimmung der Informationsqualität. Deshalb schlagen sie nach einer kurzen Vorstellung bekannter Ansätze zur Klassifikation von Merkmalen der Informationsqualität eine eigene Klassifikation vor:

- **Subjekt-Kriterien:** Diese Kriterien werden durch den Nutzer bewertet (z. B. Verständlichkeit).
- **Prozess-Kriterien:** Diese Kriterien leiten sich aus dem Prozess der Informationsgewinnung durch den Nutzer ab (z. B. Antwortzeit) und können zum Teil automatisch gemessen werden.
- **Objekt-Kriterien:** Diese Kriterien beziehen sich auf die Informationsquelle selbst (z. B. Vollständigkeit).

Die Autoren definieren mögliche Qualitätskriterien und geben zusätzlich die in der Literatur vorkommenden Synonyme für die einzelnen Kriterien an (s. Tabelle auf S. 148).

2.4.7 Datenqualität

2.4.7.1 Datenqualität in der Medizin (allgemein)

Abate et al. 1998

Datenqualität muss immer im Zusammenhang gesehen werden mit den Prozessen, die die Daten erzeugen, und dem beabsichtigten Gebrauch der Daten. Die Norm ISO 8402 definiert Qualität als „The totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs.“ Daten haben somit die erforderliche Qualität, wenn sie die Anforderungen einer entsprechenden Spezifikation erfüllen und die Spezifikation den intendierten Gebrauch korrekt wiedergibt. Diese Eigenschaften bezeichnen die Autoren als Übereinstimmung („conformance“) und Nützlichkeit („utility“). Sie schlagen vor, die festgelegten und impliziten Erfordernisse in einen

Satz von Charakteristika zu übersetzen, die im Rahmen einer Qualitätsuntersuchung strukturiert analysiert werden können.

Verwandte Attribute der Datenqualität werden in Qualitätsdimensionen zusammengefasst, die jeweils ein einzelnes Qualitätskonstrukt repräsentieren. Eine solche Betrachtungsweise soll die Entdeckung systematischer Fehler unterstützen. Eine von den Autoren zitierte Quelle nennt die folgenden 15 Dimensionen der Datenqualität:

- **Access Security:** Der Datenzugriff muss beschränkt werden.
- **Accessibility:** Die Daten müssen verfügbar oder einfach und schnell zu bekommen sein.
- **Accuracy:** Die Daten müssen korrekt, reliabel und frei von Fehlern sein.
- **Appropriate Amount of Data:** Menge oder Volumen der Daten müssen angemessen sein.
- **Believability:** Die Daten müssen als wahr, real und glaubwürdig angesehen werden.
- **Completeness:** Die Daten müssen in ausreichender Breite und Tiefe für die geplante Nutzung zur Verfügung stehen.
- **Concise Representation:** Die Daten müssen kompakt sein.
- **Ease of Understanding:** Die Daten müssen klar und eindeutig sein.
- **Interpretability:** Die Daten müssen in angemessenen Formulierungen und Einheiten vorliegen, und die Datendefinitionen müssen klar sein.
- **Objectivity:** Die Daten dürfen keinen Bias enthalten und müssen unvoreingenommen sein.
- **Relevancy:** Die Daten müssen nützlich und sinnvoll für den geplanten Gebrauch sein.
- **Representational Consistency:** Die Daten müssen immer in demselben Format vorliegen und vergleichbar mit früheren Daten sein.
- **Reputation:** Die Daten müssen aufgrund ihrer Herkunft oder Quelle als vertrauenswürdig gelten.
- **Timeliness:** Das Alter der Daten muss dem Verwendungszweck angemessen sein.
- **Value-Added:** Die Nutzung der Daten muss einen Vorteil bieten.

Qualitätskriterium	Synonym	Beschreibung
availability	accessibility, reliability, retrievability, performability	Prozentsatz der Zeit, in der die Informationsquelle zur Verfügung steht
accuracy	data quality, error rate, correctness, integrity, precision	Quotient aus der Anzahl korrekter Werte und der Gesamtzahl von Werten
amount of data	essentialness	Größe des Ergebnisses
believability	error rate, credibility, trustworthiness	Ausmaß, in dem die Information als korrekt akzeptiert wird
completeness	coverage, scope, granularity, comprehensiveness, density, extent	Quotient aus der Anzahl Antwort-Items und der Anzahl der Items in der „realen Welt“
concise representation	attribute granularity, occurrence identifiability, structural consistency, appropriateness, format precision	Ausmaß, in dem die Struktur der Information an die Information selbst angepasst ist
consistent representation	integrity, homogeneity, semantic consistency, value consistency, portability, compatibility	Ausmaß, in dem die Struktur der Information der Struktur bei anderen Informationsquellen entspricht
customer support		Umfang und Brauchbarkeit der Online-Unterstützung durch Text, E-Mail, Telefon etc.
documentation	traceability	Umfang und Brauchbarkeit der Dokumente mit Meta-Informationen
interpretability	clarity of definition, simplicity	Ausmaß, in dem die Information den technischen Möglichkeiten des Nutzers entspricht
latency	response time	Zeitumfang bis die erste Information den Nutzer erreicht
objectivity		Ausmaß, in dem die Information frei von Bias und unvoreingenommen ist
price	query value-to-cost ratio, cost-effectivity	Monetäre Kosten pro Query
relevancy	domain precision, minimum redundancy, applicability, helpfulness	Ausmaß, in dem die Information die Bedürfnisse des Nutzers befriedigt
reliability		Ausmaß, in dem der Nutzer der Information trauen kann
reputation	credibility	Ausmaß, in dem die Information oder ihre Quelle ein großes Ansehen genießt
response time	performance, turnaround time	Zeitdauer bis die komplette Antwort den Nutzer erreicht
security	privacy, access security	Ausmaß, in dem Informationen vertraulich vom Nutzer zur Informationsquelle und zurück transportiert werden
timeliness	up-to-date, freshness, currentness	Alter der Information
understandability	ease of understanding	Ausmaß, in dem die Information vom Nutzer verstanden werden kann
value-added		Umfang des Gewinns den die Nutzung der Information liefert
verifiability	naturalness, traceability, provability	Ausmaß und Leichtigkeit, mit der die Information auf Korrektheit überprüft werden kann

Tabelle zu Naumann/Rolker 2000 (S. 147): Kriterien für Informationsqualität

- Die Qualitätsdimensionen lassen sich zu Kategorien zusammenfassen:
- Intrinsic: accuracy, objectivity, believability, reputation
- Contextual: value-added, relevancy, timeliness, completeness, appropriate amount of data
- Representational: interpretability, ease of understanding, representational consistency, concise representation
- Accessible: accessibility, access security

Attribute, Dimensionen und Kategorien lassen sich hierarchisch als Datenqualitäts-Baum anordnen. Für einen solchen hierarchischen Ansatz sind effektive Verfahren aus der Ursachenforschung bekannt, mit denen sich die letztendlichen Ursachen von Fehlern und Problemen, die auf höheren Hierarchieebenen auftreten, identifizieren lassen.

Brennan/Stead 2000

Die Autoren geben redaktionelle Kommentare für den Weg von Konkordanz über Korrektheit und Vollständigkeit zu valide verarbeitbaren Repräsentationen des realen Patienten-Zustands. Konkordanz kann sowohl innerhalb der Krankenakte als auch zwischen verschiedenen Datenhaltungssystemen definiert werden. In ersterem Fall bezieht sich die Konkordanz darauf, in wie weit die in einem Teil der Akte gefundene Evidenz konsistent ist mit der Evidenz in einem anderen Teil der Akte. Ein Spezialfall ist hier die Übereinstimmung der Einträge in Freitext-Feldern mit entsprechenden kodierten Feldern. Externe Konkordanz zwischen Datenhaltungssystemen kann z. B. zwischen einer papiergestützten Krankenakte und einem EDV-gestützten Dokumentationssystem geprüft werden. Konkordanz beschreibt die Übereinstimmung zwischen zwei in verschiedenen Dokumenten festgehaltenen Repräsentationen der Realität, ohne eine Aussage über den Wahrheitsgehalt und die Angemessenheit der Darstellungen zu machen. Zur Untersuchung der Frage nach der Wiederverwendbarkeit klinischer Daten sind zum einen prospektive Studien erforderlich, um die Genauigkeit bestehender Patientenakten zu untersuchen. Zum anderen muss in methodologischen Studien nach Strategien gesucht werden, um robustere Sprachkonstrukte zur Dar-

stellung klinischer Phänomene zu entwickeln. Die Genauigkeit von Daten kann durch eine Überprüfung ihrer Korrektheit und Vollständigkeit unter Hinzuziehung eines Goldstandards untersucht werden. Alle derartigen Untersuchungen werden begrenzt durch die zugrundeliegenden (natürlichsprachlichen) Sprachkonstrukte, deren Terme und Phrasen nach speziellen Worten und Ausdrücken durchsucht werden müssen, um Übereinstimmungen festzustellen. Hier wäre der Einsatz formalisierter Sprachen, die nicht nur syntaktische Bedeutungen sondern auch semantische Interpretationen darstellen können, eine Verbesserung.

Cassidy 1998

Eines der Ziele der vom Vorstand der AHIMA (American Health Information Management Association) initiierten Initiative Vision 2006 war die Schaffung neuer Karrieremöglichkeiten für im Gesundheitsinformationsmanagement Tätige. Als eines der zukünftigen Arbeitsgebiete wurde das Management von Datenqualität identifiziert. Daraufhin wurde eine Arbeitsgruppe eingerichtet, die u. a. die folgenden Anforderungen an die Datenintegrität formulierte:

- data accuracy
- data accessibility
- data comprehensiveness
- data consistency
- data currency
- data definition
- data granularity
- data precision
- data relevancy
- data timeliness

lezzoni 1997

Die aus der Verwaltung der Gesundheitsversorgung in den USA resultierenden administrativen Daten stammen hauptsächlich von Bundes- und Landesbehörden sowie Krankenkassen. Obwohl diese Daten außer demographischen Angaben, kodierten Diagnosen und Prozeduren keine weiteren medizinischen Daten enthalten, werden sie oft zur Evaluation der Qualität im Gesundheitswesen genutzt. Die Daten sind leicht verfügbar, einfach zu akquirieren, computerlesbar und umfassen

meist eine große Anzahl Patienten. Administrative Daten erlauben einen begrenzten Einblick in die Prozessqualität und Angemessenheit der Versorgung. Zusätzlich stellt sich die Frage nach der Vollständigkeit und Richtigkeit der administrativen Daten. Derzeit sind sie wahrscheinlich am nützlichsten als Screening-Instrumente zur Identifizierung von Bereichen, in denen die Qualität detaillierter untersucht werden sollte.

Lack/Schneider 2005

Die in der externen Qualitätssicherung eingesetzten univariaten Verfahren bergen die Gefahr mangelnder Spezifität und Effizienz. In einer Fallstudie wurden an Daten der bayerischen QS-Geburtshilfe die BQS Qualitätsindikatoren Geburtshilfe nach inhaltlichen Gesichtspunkten klassifiziert und mittels multivariater Statistik ausgewertet. Bei multipel auffälligen Kliniken wurden Beratungsgespräche geführt. Die multivariaten Analysen führten zu einer deutlich höheren Spezifität. Die Klassifikation der Indikatoren zeigt eine Unausgewogenheit. Beratungsgespräche erwiesen sich als sehr effektiv. Mängel in klinikinternen Organisationsprozessen wurden häufig erst durch gezielte Ursachenforschung vor Ort aufgedeckt.

Lin/Chen 2000

Es gibt eine starke Evidenz, dass die Daten in Informationssystemen des Gesundheitswesens eine signifikante Zahl von Fehlern enthalten. Da solche Systeme in einem breiten Anwendungsspektrum eingesetzt werden und direkten oder indirekten Einfluss auf die Leistungserbringung haben, ist ihre Qualität von großer Bedeutung. Einige im Gesundheitswesen weit verbreitete Methoden zum Management von Datenqualität sind:

- Bereinigung der Front-End-Daten
- Regelmäßige Audits einer Stichprobe der Front-End-Daten
- Bereinigung der Back-End-Daten in der Analyse-Datenbank
- Wiederverwendung von Datenbank-Komponenten

Die Autoren schlagen ein integriertes System für das Qualitätsmanagement vor, das aus vier Hauptbestandteilen besteht:

1. Förderung der Unterstützung und Kooperation (u. a. Werbung für die Vorteile einer guten Datenqualität)
2. Identifikation aller relevanten Aspekte der Organisation (Aufgabe, Management und Unternehmenskultur, Informationsfluss, Informationsressourcen)
3. Analyse und Evaluation der Informationsressourcen und Formulierung eines Maßnahmenplans zur Qualitätsverbesserung
4. Berechnung der Kosten für die Datenqualität und ihre Verbesserung

Lindquist 2004

Die Pharmakovigilanz (Arzneimittelsicherheit) befasst sich mit der Entdeckung, Bewertung, Untersuchung und Prävention von unerwünschten Arzneimittelwirkungen. Der Artikel diskutiert Qualitätsthemen in der Arzneimittelsicherheit, insbesondere bei der Sammlung und Eingabe, Speicherung und Pflege sowie der Auswahl und dem Retrieval von Daten.

Als Rollen lassen sich Datenproduzenten, Datensammler, Datenhüter, Output-Produzenten und Datennutzer identifizieren.

Fallberichte unerwünschter Arzneimittelwirkungen (ADR) sind oft papierbasiert. Häufig sind jedoch nicht alle benötigten Felder ausgefüllt. Die Daten werden in den meisten Ländern in ein EDV-System eingegeben. Das zugrundeliegende Datenmodell muss nicht nur relevant, umfassend und realistisch sein, sondern auch konsistent bzgl. seiner Inhalte und der Terminologie. Um bzgl. der Granularität von Informationen flexibel zu sein, können erfasste Daten mit der zugehörigen Klassifikation verlinkt sein. Die Qualität der Dateneingabe lässt sich bestimmen durch Vergleich des Outputs mit den Originaldaten. Das Ausmaß, in dem die Eingabekraft ihr eigenes Urteil mit einbringen muss (z. B. bei der Kodierung), beeinflusst die Datenqualität. Datenrichtigkeit (accuracy) kann sich auf Syntax und Semantik beziehen. Syntaktische Richtigkeit lässt sich durch kontrolliertes Vokabular erreichen. Semantische Richtigkeit ist schwerer zu prüfen – evtl. können verschiedene Datenquellen verglichen werden, hierbei ist jedoch zu bedenken, dass Übereinstimmung nicht unbedingt Richtigkeit bedeutet. Ein großes Problem in

der Arzneimittelsicherheit ist die Dunkelziffer, die z. T. auf 95 % geschätzt wird, und die Unvollständigkeit eingereicherter Berichte. Die Datenhaltung muss ein Sicherungskonzept sowohl für Daten wie für Programme und ausreichende Schutzmaßnahmen z. B. gegen Angriffe von außen vorsehen. In den letzten Jahren werden verstärkt Data Mining-Techniken eingesetzt, die jedoch zum „Fischen“ nach Zusammenhängen (Data Dredging) verführen können.

Die Qualität von Datenoutput lässt sich anhand seiner Relevanz, Richtigkeit, Genauigkeit, Vollständigkeit und Zeitnähe messen. Aufklärungs- und Schulungsmaßnahmen können zur Senkung der Dunkelziffer beitragen. Datenqualitätsmanagement darf sich nicht auf die Datenhaltung und -abfrage beschränken. Eine „Good Pharmacovigilance Practice“ kann die Risiken fehlerhafter und fehlinterpretierter Daten reduzieren. Die Autorin empfiehlt transparente, gut dokumentierte und funktionierende Qualitätssysteme.

Massey/Hoffman 1989

Das NCHS (National Center for Health Statistics) hat ein Data Quality Monitoring System (DQMS) entwickelt, um die Datenqualität über alle eingesetzten EDV-Systeme und Datenbanken hinweg zu vergleichen und zu prüfen. Es besteht aus mehreren Komponenten: Daten innerhalb der einzelnen Systeme prüfen mit einheitlichen Guidelines und Standards, Vergleich zwischen NCHS-Daten und externen Statistiken, spezielle Forschungsprojekte zum Messen und Bewerten der Qualität von NCHS-Statistiken, Programm zur Prüfung von Fehlern und Inkonsistenzen in den Statistiken. Das Paper beschreibt die Komponente der Prüfung von Daten zwischen den verschiedenen internen Systemen, die zunächst kurz beschrieben werden (National Health Interview Survey, National Hospital Discharge Survey etc.). Vergleiche erfolgen z. B. anhand von Benchmarks oder bereits validierten Zahlen oder die einzelnen Datenquellen müssen einzeln bewertet werden. Hierzu werden Zeitreihenanalysen durchgeführt, Datensysteme verglichen (z. B. bzgl. der Effekte unterschiedlicher Survey-Methoden, Effekte der verschiedenen Definitionen und Konzepte

(z. B. ab wann ist ein Bluthochdruck gegeben), Aufdeckung von Inkonsistenzen zwischen den Datenquellen). Da diese Vergleiche sehr aufwendig sind, werden die Surveys in Design und Durchführung stärker koordiniert werden müssen.

Nielsen et al. 2004

Vier prominente Vertreter des Qualitätswesens in der Medizin fassen die wichtigsten Aussagen zusammen und betrachten ihre Anwendbarkeit auf die medizinische Versorgung.

a) Phil Crosby ist am bekanntesten für zwei Konzepte: „Mach es sofort richtig!“ und „Null Fehler!“ Die zentrale Verantwortung liegt beim Management, die Mitarbeiter folgen den Anweisungen. Er weigerte sich, die gängige Haltung zu akzeptieren, dass Fehler unvermeidlich sind, und schlägt 14 Schritte zur Qualitätsverbesserung vor. Der erste davon ist die Verpflichtung des Managements auf Qualität. Eine weitere zentrale Aussage Crosbys – dass Prävention entscheidend für Qualität ist – wird in der zunehmenden Bedeutung der Patientensicherheit umgesetzt, und die Suche nach Fehlerquellen entspricht einem seiner Schritte (Prozesse prüfen, um aktuelle und potentielle Qualitätsprobleme zu identifizieren). Der Einbeziehung der unmittelbar am Prozess Beteiligten (Qualitätsbewusstsein aller Angestellten heben und die Hemmschwelle zu Meldungen an das Management senken) entspricht die Fehlermeldung durch Patienten und Personal. Die Gesamtqualität der medizinischen Versorgung hat sich allerdings noch nicht zufriedenstellend verbessert.

b) W. Edwards Deming hat 14 Punkte benannt und das Konzept des profound knowledge entwickelt. Die Kultur in der Medizin ist immer noch hierarchisch und handwerklich orientiert und nicht systemisch. Demings Ideen entstammen einer industriellen Gesellschaft. Nach Deming ist das Gesundheitssystem in den Punkten 3 (beendet die Abhängigkeit von Masseninspektionen), 6 (institutionalisiert Schulungen), 10 (eliminiert Slogans, Ermahnungen und Appelle an die Arbeiterschaft), 11 (eliminiert numerische Quoten) und 13 (fördert Fortbildung und Selbstverbesserung für alle) gut. Durch die traditionelle Orientierung

in der Medizin sind viele negative Folgen der industriellen Revolution (top down, Fließbandarbeit etc.) nicht wirksam geworden und das Konzept des lebenslangen Lernens ist fest verankert. Allerdings bezieht sich dies vor allem auf die wissenschaftlichen Aspekte.

Mittelmäßig steht das Gesundheitssystem bzgl. der Punkte 4 (beendet die Praxis, Geschäfte allein preisbezogen abzuschließen) und 12 (schafft Barrieren ab, die den Stolz auf Qualitätsarbeit rauben) da. In den letzten Jahren sind Aufträge für Gesundheitsleistungen verstärkt preisbezogen vergeben worden und die traditionell hohe Betonung von handwerklicher Qualitätsarbeit wurde vor allem durch die zunehmende Komplexität und Verwaltungsarbeit erodiert.

Einen weiten Weg vor sich hat das Gesundheitssystem bei Punkt 1 (schafft dauerhafte Zielsetzungen für die Verbesserung von Produkten und Dienstleistungen), Punkt 2 (eignet euch die neue Philosophie an), Punkt 5 (verbessert ständig und dauerhaft die Produktions- und Dienstleistungssysteme), Punkt 7 (setzt Leitung ein und institutionalisiert sie), Punkt 8 (vertreibt die Furcht) und Punkt 9 (reißt Abteilungsbarrieren ein).

c) Armand V. Feigenbaum führte 1961 den Begriff „Total Quality Control“ ein. Er sah Total Quality Control nicht als eine Ansammlung technischer Projekte und anregender Initiativen, sondern als einen Weg zum Management einer Organisation. Die Ergebnisse dieses Managements umfassen eine höhere Reliabilität und eine Reduzierung der betrieblichen Kosten. Der Fokus aller Bemühungen zur Qualitätsverbesserung liegt auf der Erfüllung der Kundenanforderungen. Im Gesundheitswesen übernehmen die Patienten und die Gesellschaft insgesamt die Rolle der Kunden. Es liegt im allgemeinen Interesse, dass die Kunden des Gesundheitswesens auf der Basis Evidenz basierter Medizin zufriedengestellt werden. Die Qualitätskosten werden von Feigenbaum folgendermaßen klassifiziert:

- Präventionskosten
- Bewertungskosten
- Interne Fehlerkosten
- Externe Fehlerkosten

Fallstudien in der produzierenden Industrie und im Dienstleistungssektor zeigen, dass In-

vestitionen in Präventionskosten durch Einsparungen bei den internen und externen Fehlerkosten aufgewogen werden. Im Gesundheitswesen tritt das Paradox auf, dass interne Fehler (z. B. Abrechnungsfehler) zu höheren Einnahmen führen können. Eine Beseitigung solcher Fehler hat für den Leistungserbringer negative finanzielle Auswirkungen und stellt damit einen negativen Anreiz zur Qualitätsverbesserung dar.

d) Heutzutage sind viele Krankenhäuser mit der von Joseph Juran eingeführten Trilogie vertraut: Qualitätsplanung, Qualitätslenkung, Qualitätsverbesserung. Der Fokus liegt dabei auf einer genauen Beobachtung aller Prozesse, um ihre Fehler und Schwächen zu finden und zu verbessern. Nach Jurons Ansicht besteht Qualität aus zwei Aspekten:

- Bereitstellung der Merkmale, die der Kunde wünscht
- Sicherstellung der Mängelfreiheit

Selbmann 2000

In einer empirischen Wissenschaft wie der Medizin gibt es keine perfekten Daten. Die Datenfehler sollten daraufhin untersucht werden, ob sie klein genug sind, damit die aus den Daten abgeleiteten Aussagen ihre Gültigkeit behalten. Ohne Information über die eigene Qualität lassen sich keine Defizite erkennen und existieren keine Anreize zur Qualitätsverbesserung. Zum Qualitätsmanagement gehören Qualitätsdefinitionen und praktikable Qualitätsmaße. Qualität kann definiert werden als Ausmaß, mit dem die erbrachte Qualität mit vorgegebenen Kriterien für eine gute Qualität übereinstimmt (Soll-Ist-Vergleich). Eine Definition aus dem Jahre 1997 definiert Qualität als das Ausmaß, in dem die Gesundheitsleistungen für Personen oder Bevölkerungsgruppen erstens konsistent mit den vorhandenen professionellen Erkenntnissen sind und zweitens die Präferenzen gut informierter Verbraucher berücksichtigen. Da ein umfassendes Qualitätsmaß fehlt, wird die Qualität in einzelne Teile zerlegt, die bekannteste Zerlegung ist die in Struktur-, Prozess- und Ergebnisqualität. Hinter der Diskussion um die Definition von Qualität verbergen sich zwei der wichtigsten Aspekte des modernen Qualitätsmanagements:

- das Unternehmerverständnis und
- die Kundenorientierung.

Zur Qualitätssicherung und -verbesserung wird der PDCA-Zyklus – entwickelt in den 50er Jahren des 20. Jahrhunderts für die japanische Automobilindustrie – fast überall eingesetzt. Zum Einstieg in den PDCA-Zyklus ist ein gewisser Leidensdruck, die Einsicht, dass Handlungsmöglichkeiten existieren, und die Überzeugung, selbst etwas tun zu können, notwendig.

Wang et al. 1993

Die Autoren beschreiben eine Methodologie, bei der – basierend auf dem Entity-Relationship-Ansatz – den einzelnen Datenitems Qualitätsmarken („tags“) zugeordnet werden, die zusammen mit den Datenitems in derselben Datenbank gespeichert und vom Nutzer abgefragt werden können. Datenqualität wird hier – wie häufig in der Literatur – als „Erfüllung von Anforderungen“ (der Nutzer) verstanden. Zur Operationalisierung werden verschiedene Begriffe definiert:

- (Daten-)Qualitätsparameter: qualitative oder subjektive Dimension, mit der der Nutzer Datenqualität evaluiert (z. B. Aktualität, Glaubwürdigkeit)
- (Daten-)Qualitätsindikator: Datendimension, die objektive (d. h. nach festen Regeln messbare) Information über die Daten vermittelt (z. B. Datenquelle)
- (Daten-)Qualitätsattribut: Zusammenfassung von Datenqualitätsparametern und -indikatoren
- Wert eines Qualitätsindikators: für konkrete Daten gemessener Wert eines Datenqualitätsindikators (z. B. „Wall Street Journal“ für den Indikator „Datenquelle“)
- Wert eines Qualitätsparameters: aus Werten von Datenqualitätsindikatoren direkt oder indirekt abgeleiteter Wert für einen Datenqualitätsparameter (z. B. Glaubwürdigkeit wird als „hoch“ eingestuft, da Datenquelle = „Wall Street Journal“)
- Datenqualitätsanforderungen: Spezifikation der notwendigen Qualitätsindikatoren, um beim Datenzugriff Daten einer spezifischen Qualität abrufen zu können
- Datenqualitätsadministrator: Verantwortliche Person oder System zur Sicherstel-

lung, dass die Datenqualitätsanforderungen von den Daten erfüllt werden

Die Modellierung der Datenqualität umfasst mehrere Schritte:

1. Festlegung der Anwendungssicht auf die Daten
2. Festlegung der (subjektiven) Qualitätsparameter für die Anwendung in Form einer Parametersicht
3. Festlegung (objektiver) Qualitätsindikatoren für die Anwendung in Form einer oder mehrerer Qualitätssichten
4. Integration der Qualitätssichten und Erstellung eines Qualitätsschemas

Das Qualitätsschema dokumentiert sowohl die Anforderungen an die Anwendungsdaten als auch den Sachverhalt der Datenqualität. Zur Unterstützung der Modellierung haben die Autoren im Anhang eine sehr umfangreiche Liste (179 Einträge) mit denkbaren Datenqualitätsattributen zusammengestellt.

2.4.7.2 Datenqualität in epidemiologischen Studien/Surveys

Collins et al. 1991

Das Papier betrachtet Qualität bei der Durchführung epidemiologischer Studien vor dem Hintergrund der Veränderung in Gesetzen, die die Einhaltung von GLP auch bei epidemiologischen Studien fordern. GLP sollen die Qualität und Integrität der Daten sichern, dafür sorgen, dass eine Studie rekonstruierbar ist und dass der Prozess zum Schutz von Umwelt und Gesundheit beiträgt.

Es wird ein Entwurf von GEP vorgestellt, die Ansichten zu Qualität von erfahrenen Studienmanagern dargestellt, die Bedeutung eines guten Prüfplans diskutiert und die Prinzipien der Qualitätssicherung auf die Durchführung von epidemiologischen Studien angewendet. Datenzugriff und Data sharing werden diskutiert und die Problematik von Leitlinien in epidemiologischen Studien wird dargestellt.

DAE 2004

Die „Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP)“ wurden gemeinsam von DAE, GMDS, DGSMP und der Deutschen Region der Internationalen Biometrischen Gesellschaft ver-

abschiedet. Neben einer Begriffsdefinition enthalten sie Empfehlungen zu folgenden Punkten:

- Ethik
- Forschungsfrage
- Studienplan
- Probenbanken
- Qualitätssicherung
- Datenhaltung und -dokumentation
- Auswertung
- Datenschutz
- Vertragliche Rahmenbedingungen
- Interpretation
- Kommunikation und Public Health

Die zeitlichen, organisatorischen und technischen Maßnahmen zur Qualitätssicherung und Qualitätskontrolle sollen im Studienplan und in einem Operationshandbuch beschrieben werden. Alle relevanten Instrumente und Verfahren sind einer begleitenden Qualitätssicherung zu unterziehen. Für die Erfassung, Haltung, Plausibilitätsprüfung und Auswertung der Daten ist vorab ein detailliertes Konzept zu erstellen. Eine Zweiterfassung wird insbesondere für solche Variablen empfohlen, die einer späteren Plausibilitätsprüfung nur eingeschränkt zugänglich sind (z. B. Datumsangaben). Die Plausibilitätsprüfungen können teilweise bereits automatisiert bei der Datenerfassung erfolgen. Es darf auch auf Originalerhebungsbögen oder andere Rohdatenquellen zurückgegriffen werden. Zwischenauswertungen im Studienverlauf können als Teil der internen Qualitätssicherung eingesetzt werden.

Lyberg et al. 1997

Dieses Buch erläutert die Grundlagen epidemiologischer Forschung. Es enthält u. a. eine Beschreibung statistischer Methoden zur Analyse der Datenqualität.

Morganstein/Marker 1997

Die Autoren diskutieren einen systematischen Ansatz, in dem eine Survey-Agentur bzw. statistische Behörde als ein System von Prozessen betrachtet wird, das ein Produkt (z. B. einen Survey oder ein Register) herstellt – stark prozessorientiert, basiert auf Methoden des Continuous Quality Improvement (CQI).

Um einen CQI-Plan aufzustellen, werden als Schritte empfohlen:

1. Identifikation kritischer Produktcharakteristika
2. Entwicklung eines Prozessflussdiagramms
3. Identifikation von Schlüsselvariablen
4. Bewertung der Messleistung (Measurement Capability = der mittlere quadratische Fehler des Messprozesses ist relativ klein verglichen mit den Gesamt-Fehleranforderungen)
5. Bewertung der Prozess-Stabilität (die Varianz des Prozesses hat nur zufällige Komponenten)
6. Bewertung der Prozessleistung (Process Capability)

Mullooly 1990

Die Arbeit beschreibt einen Entwurf zur kosteneffektiven Dateneingabe-Prüfung für epidemiologische Studien. Die Abschätzung der Größe des Effekts einer falschen Dateneingabe erlaubt die Quantifizierung des Nutzens, den eine Datenqualitätsprüfung hat.

Bei binären Variablen erfolgt die Abschätzung der Effektgröße durch den Bias, der bei Fall-Kontroll-Studien in die OR und bei Kohortenstudien in das RR eingeht. Das mathematische Verfahren ist ähnlich dem, das für Fehlerklassifikationen eingesetzt wird. Die Datenkorrektheit hängt von der Vollständigkeit der Überprüfung dieser Variable ab.

Es werden Gleichungssysteme für zwei zusammenhängende binäre Variablen aufgestellt, sowohl für Fall-Kontroll-Design wie auch für Kohortendesign. Bei stetigen Variablen erhöht sich durch Dateneingabe-Fehler die Varianz. Wenn die Fehler zufällig verteilt sind, ist der mittlere Fehler 0. Die Effekte der Dateneingabe-Fehler werden durch den Bias geschätzt, der in den Korrelationskoeffizienten eingeht. Der Autor stellt auch für stetige Variablen Gleichungssysteme für zwei zusammenhängende Variablen auf. Bei einer vermuteten Fehlerrate von weniger als 0,25 % pro Feld reicht eine Check-Rate von 5 % aus (Übersichtstabelle hinten in der Arbeit).

Norton et al. 1981

Die Autoren beschreiben die Fehlermuster bei der Eingabe von kirchlichen Registerdaten im Rahmen einer bevölkerungsbasierten epidemiologischen Studie in einem texanischen Ort. Da die Bevölkerung fast vollständig katholisch ist, sind die seit 1790 geführten kirchlichen Register die beste Quelle zu Geburts-, Tauf-, Heirats- und Todesdaten. Diese Daten wurden, ergänzt mit städtischen und Bezirksregisterdaten, zu Kernfamilien-Einheiten zusammengeführt. Hierbei ist der Zusammenführungsprozess (record linkage) von kritischer Bedeutung. Da in erster Linie Textfelder verarbeitet werden, sind automatische Datenprüfungen kaum anwendbar. Die Taufregister werden fotografiert und vom Dia über interaktive Masken direkt in den Computer eingegeben. Namen werden sofort bei Eingabe in einer Namensbibliothek gesucht. Wird der Name nicht gefunden, muss der Dateneingabe entschieden, ob er ihn in dieser Form eingibt. Wie sich gezeigt hat, können mit diesem System viele Tippfehler vermieden werden.

Um die Datenqualität zu prüfen, wurde eine Stichprobe von einem Dia (entspricht 10 Datensätzen) pro Woche und Dateneingabe ausgewählt und bei weniger als 10 Dateneingabern in der Woche mit weiteren Datensätzen auf 100 aufgefüllt. Die kleinste Stichprobe umfasste 69 Datensätze. Stichproben-Datensätze wurden von zwei Eingabekräften nochmals eingegeben, wobei sich mögliche Confounder dadurch ergeben, dass diese beiden Eingabekräfte wussten, dass sie eine Kontrolleingabe durchführen und dass auch von ihnen selbst zuvor eingegebene Datensätze in den Stichproben enthalten waren.

Die Auswertung erfolgte stratifiziert nach Eingabekraft, die drei eingegebenen Versionen wurden per Programm verglichen. Stimmten zwei Versionen überein, galt die dritte als fehlerhaft. Stimmten die drei Datensätze nicht überein, wurden sie alle als fehlerhaft gezählt. Jede Variation zwischen den Datensätzen wurde als Abweichung gewertet. Felder mit Informationen zu Firmung und Heirat wurden aus der Analyse ausgeschlossen.

Etwa 4 % der Felder weisen Fehler auf. Die Veränderungen über die Zeit (Wochen) sind zufällig und liegen unterhalb von zwei Stan-

dardabweichungen. Die Variable Zeit geht daher nicht in weitere Analysen ein. Eingabekraft und Feld gingen als Faktoren in eine zweifache Varianzanalyse ein. Sowohl Eingabekraft und Feld als auch die Wechselwirkungen waren hierbei signifikant. Die Wechselwirkung zwischen Eingabekraft und Feld macht einen großen Anteil der Gesamtvarianz aus. Die Eingabekräfte wurden nach Fehlerraten mit Hilfe des Student-Newman-Keuls-Tests in Gruppen eingeteilt, innerhalb der Gruppen ergeben sich dann keine signifikant unterschiedlichen Fehlerraten mehr. Als weitere Einflussvariable wurde die Geschwindigkeit der Dateneingabe untersucht. Dabei zeigte sich, dass die schnelleren Eingabekräfte weniger Fehler machen. Die Felder lassen sich bzgl. der Fehlerraten einteilen nach Namens- und Datumsfeldern, innerhalb der Namensfelder sind Personennamen mit mehr Fehlern behaftet als Ortsnamen und innerhalb der Personennamen sind die Namen der Taufpaten besonders problematisch.

Aufgetretene Fehler wurden nach Typen gruppiert. Fehlertypen: 1) Feld leer oder unvollständig, 2) existierender, aber falscher Name, 3) String, der nicht als Name erkannt wird, 4) Variante eines Standard-Namens, 5) drei abweichende Einträge, 6) richtiger Name im falschen Feld. Diese Fehlertypen unterscheiden sich in ihrer Bedeutung für die Zusammenführung: 1 und 6 können durch Suche in anderen Feldern ggf. abgefangen werden, 3 und 4 lassen sich automatisch durch den Abgleich identifizieren. 2 ist schwerwiegend, weil sich der Fehler schwer feststellen lässt. Insgesamt verteilen sich die aufgetretenen Fehler gleichmäßig auf alle Typen.

Die hier gefundene Bedeutung der Eingabekraft für die Fehlerrate impliziert die Bedeutung sorgfältiger Personalentscheidungen und -schulungen. Da auch während der Testphase die automatischen Datenprüfungen bei der Eingabe aktiviert waren, kann ihr Einfluss auf die Fehlerrate hier nicht überprüft werden. Die Autoren empfehlen, im Einsatz solcher Datenprüfungen zwischen Fehlervermeidung und Zeitverlust abzuwägen. Von besonderer Bedeutung sind Eingaberegeln, bei unklar definierten Feldern war die Fehlerrate am höchsten. Kodierung sollte nicht bei, sondern nach

der Dateneingabe stattfinden, um keine zusätzliche Fehlerquelle einzuführen.

Rossi 2005

Im Rahmen einer Diplomarbeit wurden mehrere SAS-Makros geschrieben, die Datenmanagern und Auswertern (z. B. Biometrikern) einen Überblick über vorhandene SAS-Dateien verschaffen, die z. B. den Datenbestand einer epidemiologischen Studie repräsentieren. Durch eine Analyse der Arbeitsprozesse von Datenmanagern und Auswertern wurden diejenigen Teilprozesse identifiziert, die programmtechnisch unterstützt werden können, und die Anforderungen an die Programmierung abgeleitet. Der Überblick über die SAS-Dateien umfasst eine Reihe von Plausibilitätsprüfungen auf den Daten mit automatischer Unterscheidung zwischen kategoriellen und stetigen Variablen (u. a. Anzahl fehlender Werte, Verteilung stetiger Variablen, Ausreißer). Die Ergebnisse der Plausibilitätsprüfungen werden in Tabellen, Listen und Graphiken dargestellt. Entdeckte Fehler können in nachvollziehbarer Weise korrigiert werden.

Whitney et al. 1998

Bei Kohortenstudien werden in weniger Zentren mehr Daten über einen längeren Zeitraum erhoben als in klinischen Studien. Qualitätssicherung findet vor, Qualitätslenkung während und nach der Datenerhebung statt. Zur Qualitätssicherung gehört die Entwicklung des Studienprotokolls, der Datenerhebungsinstrumente und der Eingabemasken. Selbstauffüller wie Interviewer-Fragebogen haben ihre jeweils eigene Fehlerproblematik. Ein Operationshandbuch ist zu erstellen und ggf. fortzuschreiben. Für gemessene Daten sollten standardisierte Messtechniken und -geräte verwendet werden. Nötige Veränderungen in den Studienprozessen müssen an alle Beteiligten weitergegeben werden. Das Personal ist adäquat zu schulen. Kommunikation ist ein Schlüsselement erfolgreicher Qualitätssicherung, und jede Studie sollte angemessene Kommunikationsstrukturen haben (z. B. für Qualitätsberichte oder Qualitätsbeauftragte in den Zentren vor Ort). Alle Studienprozeduren müssen getestet und validiert werden. Wo möglich

und sinnvoll, kann eine Pilotstudie durchgeführt werden.

Das Ziel der Qualitätslenkung ist es, während und nach der Datenerhebung Quellen von Bias oder Noise aufzudecken. Häufig erfolgt die Implementierung von Prozeduren zur Problemfindung eher als die der Problemlösung. Gute Qualitätsberichte können in den Zentren vor Ort als Anreiz dienen und sie daran erinnern, dass sie Teil einer größeren Studie sind und welche Bedeutung die von ihnen gelieferten Daten und deren Qualität haben (Feedback). Rekrutierung und Eligibilität ebenso wie Leistungsdaten (z. B. bei extremen Abweichungen vom Mittelwert aller Zentren) sollten regelmäßig berichtet werden. Interviewer können anhand von Aufnahmen beurteilt werden. Mehrfachmessungen können bei der Fehlerfindung helfen.

Bei Langzeitstudien spielt der Zeitfaktor (Personal- und Gerätewechsel, Inkonsistenzen in Teilnehmer-Antworten über die Zeit) eine besondere Rolle. Häufige Schulungen sind nötig, die Kosten lassen sich z. B. durch regionale Schulungstermine mit EDV-Unterstützung senken. Bei Gerätewechseln muss die Vergleichbarkeit der Messungen gewährleistet sein. Beim Einsatz von Referenzzentren und/oder Zentrallabors müssen gute Schätzwerte der biologischen Variabilität vorliegen, damit echte Veränderungen entdeckt werden können. Inter- und intrapersonelle Variabilität ist zu berücksichtigen, ein Trackingsystem sollte implementiert sein. Messfehler können sowohl zu Typ-I- als auch zu Typ-II-Fehlern führen. Auch bei automatisierten Prozeduren sollte auf Drift geprüft werden. Wenn bei der Analyse keine Adjustierung möglich ist, müssen die Daten u. U. neu aufgearbeitet werden. Das Datenzentrum bzw. koordinierende Zentrum muss Berichte und Qualitätsanalysen zeitnah erstellen und versenden. Das Personal sollte erfahren und qualifiziert sein. Die Größe eines Qualitätsmanagement-Systems für Langzeitstudien sollte nicht unterschätzt werden, Qualität ist jedoch von entscheidender Bedeutung für die Validität einer Studie.

2.4.7.3 Datenqualität in klinischen Studien¹

Akhtar-Danesh/Dehghan-Kooshkghazi 2003

Unter den verschiedenen Arten von Fehlverhalten in der medizinischen Forschung kann man die Erfindung von Daten als einen der schwersten Fälle ansehen. Es gibt Argumente dafür, dass die Korrelationskoeffizienten bei erfundenen Daten gewöhnlich größer sind als bei echten Daten. Die Autoren untersuchten an drei Beispielen die Unterschiede zwischen erfundenen und realen Daten an der Assoziation zweier Variablen.

Mit Hilfe der ersten beiden Beispiele sollte herausgefunden werden, wie gut Korrelationsstrukturen realer Daten mit erfundenen Daten nachgeahmt werden können. Die statistischen Kennzahlen zweier realer Datenbestände wurden den 34 Probanden (Mitglieder zweier Universitäten im Iran) mit der Bitte vorgelegt, ähnliche Daten für 40 Patienten zu erfinden. Im dritten Beispiel wurde der Mittelwert einer kontinuierlichen Variablen zwischen zwei Gruppen verglichen. Hier wurden von 17 Probanden die Daten für jeweils 30 Patienten pro Gruppe erfunden. Bei allen drei Beispielen wurden die von den Probanden erfundenen Daten mit 2.500 per Computer erzeugten Zufallsdatenbeständen verglichen, basierend auf den statistischen Kenngrößen der echten Daten und einer Normalverteilung bzw. als Zufallsstichprobe aus den echten Daten.

In den ersten beiden Beispielen wiesen die erfundenen Daten wesentlich höhere Korrelationskoeffizienten auf als die realen und simulierten Daten. Im dritten Beispiel zeigte sich, dass viele Menschen trotz Kenntnis der wahren Gruppenunterschiede dazu tendieren, Daten mit geringem oder keinem Unterschied zwischen den Gruppen zu erfinden.

Die Studie zeigte, dass hohe Korrelationskoeffizienten ein Hinweis auf Datenmanipulation sein können. Für die Unterschiede bei Mittelwerten lässt sich dies nicht sagen.

1 Der Bereich klinischer Studien wurde nur in dem Ausmaß betrachtet, als eine direkte Übernahme der Ergebnisse für Register und Kohortenstudien möglich erschien. Die Anwendung der in diesem Projekt zu entwickelnden Leitlinie war im Projektauftrag auf diese beiden Bereiche beschränkt worden.

Arndt et al. 1994

Große Forschungsprojekte werfen spezielle Probleme der Datenqualität auf. Die Fehler können eine größere absolute Zahl erreichen und können sich schwerwiegender auf die Analyse auswirken als bei kleinen und mittleren Projekten. Die Autoren hatten die Gelegenheit, ein aktives Datensystem für eine multizentrische Feldstudie zu entwickeln. Sie beobachteten Anzahl und Typen der auftretenden Fehler während der Datenverarbeitung, den potentiellen Einfluss der Fehler auf die Studienergebnisse und evaluierten Methoden zur Reduktion des Fehlereinflusses auf die Studienergebnisse.

Studienziel war die Evaluation von sechs verschiedenen diagnostischen Kriterien für Schizophrenie und verwandte psychotische Erkrankungen. Untersucht wurden die sieben in den USA teilnehmenden Zentren. Da die Schätzung der Reliabilität ein wichtiges Evaluationskriterium der sechs diagnostischen Kriterien darstellte, wurden mehr als 90 % aller teilnehmenden Patienten zweimal begutachtet. Die Daten wurden durch halb-standardisierte Interviews gewonnen und anschließend beurteilt. Die auf papiernen Erhebungsbögen gesammelten Daten wurden elektronisch mit einer SAS/FSP-Anwendung in einer Datenbank erfasst. „Schnappschüsse“ der Daten wurden an mehreren Zeitpunkten genommen und separat gespeichert, um sie mit den endgültigen Daten für die Analyse zu vergleichen:

- Zeitpunkt 1: Die Erfassungsbögen wurden photokopiert und in der Datenbank erfasst.
- Zeitpunkt 2: Die Daten wurden ein zweites Mal von einer zweiten Person in der Datenbank erfasst. Diskrepanzen zwischen Erst- und Zweiterfassung wurden aufgelöst.
- Zeitpunkt 3: Der zentrale Datenmanager überprüfte jeden Erfassungsbogen manuell auf offensichtliche Fehler (z. B. Monat und Tag vertauscht) und korrigierte die gefundenen Fehler. Danach wurden die Daten ein drittes Mal in der Datenbank erfasst. Dabei wurden automatisch Wertebereiche geprüft.
- Zeitpunkt 4: Die korrigierten Erfassungsbögen wurden erneut zweifach erfasst ein-

schließlich automatischer Überprüfung der Wertebereiche.

- Zeitpunkt 5: Ein Computerprogramm suchte in den Daten nach mehr als 100 logischen Fehlern. Gefundene Fehler wurden korrigiert und die korrigierten Daten gespeichert.
- Zeitpunkt 6: Die zehn zuletzt erfassten Fälle wurden mit allen anderen in der Datenbank gespeicherten Fällen verglichen, um Duplikate bzw. falsche Identifikationsnummern zu finden.
- Zeitpunkt 7: In den Daten wurde nach univariaten und bivariaten Ausreißern gesucht. Der nach der Fehlerkorrektur vorliegende Datenbestand wurde als „Goldstandard“ für die Qualitätsanalyse definiert.

In die Fehleranalyse einbezogen wurden 688 Erhebungsbögen mit 429 Variablen. Die meisten Fehler wurden bei der Ausfüllung der Erhebungsbögen gemacht. Zum Zeitpunkt 1 waren 2,4 % der Werte fehlerhaft. Nur 14 der 688 Erhebungsbögen waren zu diesem Zeitpunkt völlig fehlerfrei. Die Ergebnisse der Fehleranalyse wurden den beteiligten Zentren zurückgemeldet. Dies erwies sich als sehr effektiv. Die Fehlerrate einer einzelnen Variablen steht nicht generell im Zusammenhang mit dem Einfluss des Fehlers auf die Reliabilität oder seinem Einfluss auf die Studie. Die Fehler sind nicht immer unabhängig voneinander. Abhängigkeiten können jedoch auch vorteilhaft bei der Fehlersuche genutzt werden. Eine Abhängigkeit bedeutet, dass in der Umgebung eines Fehlers mit größerer Wahrscheinlichkeit weitere Fehler zu finden sind.

Bailey 1991

Das National Heart, Lung and Blood Institute (NHLBI) in den USA sponserte 1980 bis 1981 eine multizentrische Studie, in der Tiermodelle für den Schutz des ischämischen Myokardiums untersucht wurden. Mehrere Laboratorien sollten identische Studien an wachen bzw. narkotisierten Hunden durchführen und die Ergebnisse sollten beim National Institute of Health (NIH) zusammengeführt und ausgewertet werden.

Nach der Datensammlung ergaben sich Probleme mit den Daten eines bestimmten Labors,

die von den übrigen Laboren abweichende Ergebnisse zeigten. Nachdem ein Mitarbeiter des betroffenen Labors eingestanden hatte, in einer früheren Studie Daten gefälscht zu haben, wurden von der Koordinierungsstelle in Zusammenarbeit mit den anderen Forschern statistische Analysen durchgeführt, um die Datenfälschung nachzuweisen und festzustellen, ob ein Teil der Daten valide war. Wirkliche Beweise für die Datenfälschung wurden jedoch erst bei der Untersuchung von Gewebeproben der Versuchstiere gewonnen. Aus den Erfahrungen in dieser Studie lassen sich einige Schlüsse ziehen:

- Rein statistische Signifikanz mag notwendig sein, ist aber nicht ausreichend zum Nachweis einer Datenfälschung.
- Eine Grafik ist soviel wert wie 1.000 p-Werte.
- Die biologische Fachkenntnis der beteiligten Forscher war sehr wertvoll.
- Nur nicht-statistische, „harte“ Evidenz konnte den Verdacht tatsächlich bestätigen, obwohl die statistische Evidenz überzeugend war.

BMGS 2004

Die GCP-Verordnung (GCP = Good Clinical Practice) soll die Einhaltung der Guten Klinischen Praxis bei der Planung, Durchführung und Dokumentation klinischer Prüfungen am Menschen und der Berichterstattung darüber gewährleisten. Sie regelt Aufgaben, Verantwortungsbereiche und Verfahren hinsichtlich der Planung, Genehmigung, Durchführung, Überwachung, Dokumentation und Berichterstattung klinischer Prüfungen. Im Zusammenhang mit Inspektionen der Behörden werden Qualitätssicherungssysteme des Sponsors als zu überprüfende Ressourcen genannt. Die für die Inspektion zuständige Behörde muss über ein „umfassend geplantes und korrekt geführtes Qualitätssicherungssystem“ (§ 15) verfügen.

Califf et al. 1997

Die Methoden zum Nachweis der Richtigkeit der Daten, die klinischen Studien zugrunde liegen, sind bisher nicht systematisch evaluiert worden. Der Artikel stellt Ansätze zum kosteneffektiven Monitoring klinischer Studi-

en dar, die jedoch bisher nicht evaluiert werden konnten, da hierzu die Datenbasis fehlt. Einige bedeutende Einflussfaktoren auf die Durchführung klinischer Studien sind inzwischen deutlich: Es werden große Patientenzahlen benötigt, weil eine neue Therapie typischerweise eine relative Nutzenerhöhung von unter 25 % (selbst im Vergleich mit Placebo) bringt. Subgruppenanalysen bergen die Gefahr, dass zufällige Ergebnisse Signifikanzen zeigen. Trotzdem sollten klinische Baselinedaten erfasst werden. Es sollten nur solche Endpunkte verwendet werden, die für den Patienten von Bedeutung sind, da Surrogat-Endpunkte irreführend sein können.

Ein rationales, vertrauenswürdigen und gleichzeitig kostengünstiges Monitoring wird für klinische Studien immer wichtiger. Der Autor unterscheidet zwischen der Überwachung (engl. monitoring) einer Studie in allen relevanten Aspekten und dem Monitoring im engeren Sinne (engl. auditing), das in erster Linie den Originaldatenabgleich umfasst.

Das Monitoring ist von entscheidender Bedeutung z. B. im Aufdecken von Betrug, in der Leistungssteigerung für die Zentren, in der Verbesserung der Datenqualität. Es basiert auf der These, dass die Patientenakte die „Wahrheit“ über den Patienten wiedergibt. Dies hat dazu geführt, dass Akten so geführt werden, dass sie zum Studienprotokoll passen, z. B. indem zusätzliche Daten erhoben werden, damit sie für einen Originaldatenabgleich verfügbar sind (denn häufig werden für Studien benötigte Items im klinischen Alltag gar nicht dokumentiert). Bei der Beurteilung von Vergleichen zwischen Patientenakten und Studiendaten sollte zunächst die Reproduzierbarkeit der Beobachtung selbst berücksichtigt werden (Inter- und Intraobserver Variability). Darüber hinaus ist zu bedenken, dass die Dokumentation in Patientenakten häufig zufällig und unvollständig ist. Daher ist die Annahme, die Papierakte sei genauer als die Studiendaten und stelle den Goldstandard dar, fehlerbehaftet und sollte überdacht werden.

Maßnahmen, die Notwendigkeit für und Kosten des Monitoring reduzieren, sind von der Planungsphase an zu implementieren.

- Das Studienprotokoll sollte möglichst einfach und praxisnah sein. Endpunkte soll-

ten klar und einfach zu messen sein. Die zu erhebenden Daten sind auf das Notwendige zu beschränken.

- Es sollte eine Leitkommission mit Erfahrungen in dem jeweiligen medizinischen Feld und in der Durchführung von Studien einbezogen werden.
- Von großer Bedeutung ist die Schulung von Prüfärzten und Koordinatoren, insbesondere auch die Durchführung von Prüfärzttreffen und zentralen Schulungen als standardisierter Maßnahme zur Sicherstellung der Protokollcompliance. Erfahrene Koordinatoren sollten die CRFs einem Review unterziehen.
- Regelmäßige Kommunikation per Telefon, Newsletter und E-Mail, ggf. Einrichtung einer Hotline senkt den Reisebedarf.
- Plausibilitätsprüfungen während der Dateneingabe erleichtern die frühe Korrektur von Datenfehlern. Die CRF sollten früh einer zentralen Sichtkontrolle unterzogen werden, damit systematische Fehler rasch korrigiert werden können.
- Wenn ein wichtiger Endpunkt subjektiver Beurteilung unterliegt, muss eine zentrale Kommission die möglichen Ereignisse klassifizieren und über ihre Beurteilung entscheiden.

Zur Beurteilung von Monitor-Ergebnissen ist eine Methode zur Klassifikation von Fehlern notwendig. Hierzu sollte eine hierarchische Liste essentieller (z. B. Primärziel), kritischer (z. B. Ein-/Ausschlusskriterien) und weniger wichtiger Variablen erstellt und den Monitorprioritäten zu Grunde gelegt werden. Darüber hinaus sind Fehler nach den zu ihrer Aufdeckung nötigen Maßnahmen einzuteilen (einfache Aufdeckung durch automatische Plausibilitätsprüfungen usw.). Am meisten Probleme bereiten Fehler, die nicht ohne einen Besuch im Zentrum aufgedeckt werden können. Häufig wird ein unabhängiges Datenüberwachungs- und Sicherheitskomitee eingesetzt, das Empfehlungen an die Studienleitung ausspricht. Dies ist eine wichtige Ressource auch über Zwischenanalysen hinaus. Zu Monitoring-Methoden gibt es überraschend wenig quantitative Daten. In Zulassungsstudien für die FDA werden häufig alle Variablen einem Onsite-Monitoring unterzo-

gen – dies ist sehr kostspielig, sichert aber die Studie gegen Rückfragen der Behörden ab. Einen ganz anderen Ansatz verfolgt die ISIS-Gruppe, die ihr CRF auf eine Seite gekürzt hat, völlig auf Monitorbesuche verzichtet und statt dessen zentrale Prüfarzt-Treffen mit Prüfung der Daten durchführt. Dieser Ansatz setzt eine ausreichende Patientenzahl und eindeutige Endpunkte voraus.

Der von den Autoren verfolgte Ansatz basiert auf der systematischen Evaluation einer Stichprobe von Datensätzen im Zentrum. Nach zentralen Schulungen wird nur für Zentren, die bisher keine exzellente Dokumentationsqualität hatten, eine Studieninitiierung vor Ort durchgeführt. Während der Studienlaufzeit wird jedes Zentrum mindestens einmal besucht, im Allgemeinen werden mindestens 10 % der Daten vor Ort geprüft. Für jede kritische Variable werden die Fehlerraten mitgeführt, die unterhalb einer vorher festgelegten Grenze bleiben sollen. Diese Grenze variiert je nach Bedeutung der Variable und ihrer inhärenten Variabilität.

Zu beim Onsite-Monitoring gefundenen Fehlerraten existiert fast keine Literatur. Das NCI hat kürzlich ein auf kritische Variablen beschränktes Audit des NSABBP-Protokolls B06 durchgeführt, nachdem in einem Zentrum Fälschungen aufgetreten waren. Über 95 % der Variablen konnten mit Originaldaten abgeglichen werden und die Rate der Abweichungen lag unter 1 %. Die Methodik wurde nicht berichtet. Ein Bericht der EORTC von 1989 über die Evaluation von teilnehmenden Zentren zeigte mediane Fehlerraten von 2,8 %, die häufigste Fehlerursache waren Übertragungsfehler. Die Kosten steigen dramatisch mit der Zahl der Zentren und eingebrachten Patienten. Eine Reduktion der Besuche auf einen pro Jahr und des Anteils überprüfter Daten auf 10 – 20 % senkt die Kosten erheblich.

In Zukunft könnten die Kosten durch Einsatz von klinischen Datenbanken mit sekundärer Datenanalyse weiter sinken.

Castellion 1983

Der Artikel stammt aus der Zeit vor der Finalisierung der GCP-Regularien für Sponsor und Monitor. Er stellt die Vorgehensweise der

Firma Norwich Pharmaceuticals, Inc im Rahmen eines Qualitätssicherungsprogramms dar. Das Qualitätssicherungs-System von Norwich Pharmaceuticals basiert auf der Idee, dass Qualität mit den Mitarbeitern beginnt und jedes System so gut ist wie seine schlechteste Komponente, die identifiziert und durch Management-Entscheidungen verbessert werden muss. Zentrales Instrument sind die SOP-Handbücher, die nutzerorientiert auf die jeweiligen Aufgabenbereiche zugeschnitten sind.

Um Diskrepanzen aufzulösen, müssen Daten so dokumentiert und Dokumente so archiviert werden, dass ein Audit Trail möglich ist. Der Artikel führt Dokumente an, die einem GCP-Audit unterzogen werden können und benennt die relevanten SOPs. Die Firma archiviert, wo möglich, Originale – die Kopien erhält der Prüfarzt. Die Originale werden auf Mikrofilm archiviert.

Ein eigenes Programm sichert die Compliance und Konformität mit den Unternehmens-Qualitätszielen in der Forschung und Entwicklung. Es gilt auch für externe Kooperationspartner in klinischen Studien. Die entsprechende Projektgruppe ist in die F&E-Abteilung eingebunden (als „Dienstleister“, nicht als „Polizei“). Ihre Aktivitäten sind präventiv, prüfend und korrigierend mit Betonung auf Prävention und Prüfung (z. B. Review aller relevanten Dokumente vor Beginn einer Studie, Pflege der SOPs, Durchführung interner Audits). Ein computergenerierter Zeitplan, der alle Studien umfasst, dient der Terminverwaltung. On-Site-Monitoring wird durch Firmenpersonal durchgeführt.

Chilvers et al. 1988

Wichtige Anforderungen an klinische Studien sind: Bias bei Patientenzuweisung vermeiden (Random), genügend große Fallzahl (häufig nicht erfüllt, daher zunehmend multizentrische Studien), Daten müssen zuverlässig (reliable) und vollständig sein. Das Datenmanagement wird durch DBMS und automatisierte Datenchecks unterstützt. Hierzu wurde COMPACT (Computer Package for Cancer Trials) in FORTRAN entwickelt. In COMPACT werden die CRF-Typen und Variablen pro CRF definiert. Dabei können Plausibilitätsschecks

programmiert werden. Dann werden die Daten interaktiv in COMPACT eingegeben. Bei Fehlern wird ein Flag gesetzt und ein Hinweis generiert, und schwere Fehler werden in einer Problem-Datei abgelegt. Diese kann am Ende der Datenerfassung durchgesehen werden. Wird die Korrektur später eingegeben, wird der Eintrag aus der Problemdatei gelöscht. Aus ihr können Queries generiert werden. COMPACT enthält auch eine Terminverwaltung bzgl. Patientenvisiten und Follow-Up und kann flat files für den Export zum Statistikprogramm und Variablenbeschreibungen für SAS und SPSS erzeugen. Für gute Datenqualität sind darüber hinaus die Qualität der CRFs, die Motivation der Beteiligten und die Art der abgefragten Information von Bedeutung.

Collins 1998

Der Artikel bezieht sich in erster Linie auf Studiendesign und -planung und betont die Notwendigkeit von Schulungen. Darüber hinaus wird das Datenmanagement in der pharmazeutischen Industrie in den Grundzügen vorgestellt.

Nach Erstellung, Überarbeitung und Finalisierung eines Studienprotokolls können die zu erhebenden Daten festgelegt werden. Das CRF-Design hat einen unmittelbaren Einfluss auf die Datenqualität. Erfahrungen in der pharmazeutischen Industrie haben gezeigt, dass sich bei abteilungsübergreifendem Review der CRFs eine höhere Datenqualität ergibt. Es sollten nur die Daten erhoben werden, die für die Erreichung der Studienziele notwendig sind. Vor Beginn einer Studie wird häufig ein Prüfarzttreffen durchgeführt. Auch dabei können die CRF nochmals einem Review-Prozess unterzogen werden. Häufig werden die Prüfarzte bei dieser Gelegenheit auch im Umgang mit den CRF geschult. Fragen und Rückmeldungen werden von den beteiligten Monitoren an den Sponsor zurückgemeldet. Auch der Pretest von Formularen wird empfohlen. Nach Einbringung der ersten Patienten sollte ein Monitorbesuch im Zentrum stattfinden, um die Genauigkeit der Dokumentation sicherzustellen.

Datenmanagement ist der Prozess, die Daten vom CRF in ein per EDV verarbeitbares Format zu bringen und ihre vollständige und kor-

rekte Eingabe sicherzustellen. Er beginnt mit dem Originaldatenabgleich durch den Monitor. Danach werden die Daten (häufig doppelt) eingegeben und geprüft (Sichtkontrolle, programmgesteuerte Plausibilitätsprüfungen). Wenn Fehler nicht durch Abgleich mit dem CRF auflösbar sind, wird der Prüfarzt kontaktiert. Wenn die Datenbank „sauber“ und geschlossen ist, können die Daten ausgewertet werden.

Crerand et al. 2002

Der Artikel gibt einen Überblick über die Ansätze der pharmazeutischen Industrie, die darauf abzielen, von vornherein Qualität in den Datenmanagement-Prozess einzubauen. Dies beginnt mit dem Entwurf guter CRFs. Eine Bibliothek von Standard-CRF-Modulen unterstützt diese Entwicklung. Die CRF sind einem ausführlichen Review unterworfen. Nach Fertigstellung der CRF wird die Datenbank entwickelt und mit Hilfe von Testskripten validiert. Sie enthält automatische Datenprüfungen, die während der Eingabe ausgeführt und in einem Datenvalidierungs-Dokument schriftlich festgehalten werden. Darüber hinaus werden Leitlinien entwickelt zum Ausfüllen der CRF, zum Monitoring, zur Dateneingabe, zu erlaubten Korrekturen in der Datenbank und zum Datenfluss. Ursprünglich wurden CRFs im Zentrum erfasst und erst bei der Dateneingabe wurden Fehler identifiziert. Inzwischen geht die Tendenz dahin, die Datenbereinigung quellennäher durchzuführen. Hierzu werden entweder

- die ersten eingehenden CRFs sofort nach Eingang geprüft und dem Zentrum zeitnah ein Feedback gegeben – zusätzlich führt das Datenmanagement Trendanalysen bzgl. der Fehler durch – oder
- die Daten werden zeitnah eingegeben und überprüft, so dass Rückfragen rasch generiert werden können.

Diese Strategien führen zur Verbesserung der Datenqualität, u. a., da Änderungen an den CRFs vorgenommen werden können, bevor der Großteil der Daten erhoben wird. Häufig werden CRFs eingescannt und zusammen mit den dazugehörigen Queries indexiert. Aus dem Scanning-System heraus werden bereits erste Berichte bzgl. fehlender Items erzeugt.

Diese können bereits vor der Dateneingabe bearbeitet werden. Die Daten werden doppelt eingegeben, Abweichungen zwischen Erst- und Zweiteingabe von der erfahreneren Eingabekraft oder einem dritten Mitarbeiter geklärt. Der Datenmanager ist für das Klären von Datenfehlern/Diskrepanzen verantwortlich – was nicht anhand der CRF aufgeklärt werden kann, geht als Query an das Zentrum. Kritische Felder werden anhand von Datenlistings überprüft (patientenübergreifend). Vor der Übergabe der Daten an die Biometrie sollte sichergestellt sein, dass ein gewisser Datenqualitätsstandard eingehalten wird. Die Freigabe der Datenbank wird von einer Checkliste begleitet (alle CRF wurden eingegeben, alle Plausibilitätsprüfungen wurden durchgeführt und Problemfälle gelöst, die Datenbank ist konsistent mit anderen Datenbanken (z. B. der Dokumentation unerwünschter Ereignisse), bei der Kodierung wurden Standardnachschaugerwerke eingesetzt). Eine unabhängige Prüfung der Datenqualität wird durchgeführt und dokumentiert. Um studienübergreifende Analysen zu ermöglichen, werden Messgrößen wie

- Anzahl der Queries (pro Patient),
 - Anzahl der Queries mit nachfolgender Änderung in der Datenbank (pro Patient),
 - Anzahl der Datenbank-Versionen
 - Queries nach Typ (prozentual)
 - Zeit von last patient out bis Datenbank-Freigabe
- festgehalten.

Das Management von Datenqualität in klinischen Studien spiegelt das generelle Datenqualitätsmanagement klinischer Daten wider.

Drabig et al. 2005

Zur standardisierten Bezeichnung von Variablen in klinischen Studien wurde eine dreigliedrige Systematik mit einer hierarchischen Gruppierungsoption und einer zusätzlichen Identifikation des Skalenniveaus entwickelt und erfolgreich im mehreren umfangreichen klinischen Studien eingesetzt. Die Systematik war einfach handhabbar und erlaubte eine flexible Anpassung an die Erfordernisse der jeweiligen Indikation. Der Einsatz einer Systematik ist einer unsystematischen Variablenbenennung vorzuziehen.

EMEA 2004

In klinischen Studien, die über einen längeren Zeitraum durchgeführt werden, ist es erforderlich, die Studienteilnehmer vor einer unvermeidbaren Risikosteigerung zu schützen. Andererseits sollte eine Studie nicht zu früh beendet werden, um die Beantwortung der wissenschaftlichen Fragestellungen nicht zu beeinträchtigen. Beide Aufgaben werden von einem Gremium Studien-unabhängiger Experten (Data Monitoring Committee DMC) übernommen. Das DMC überprüft regelmäßig anhand akkumulierter Studiendaten die wissenschaftliche und ethische Integrität und spricht Empfehlungen für den Fortgang der Studie aus.

Gassman et al. 1995

Qualitätssicherung sollte ein Teil des Datenmanagement-Systems einer Studie sein und parallel zu anderen Datenmanagement-Aktivitäten stattfinden. Da die Verantwortung für Datenqualität bei zentralen und lokalen Organisationen und Personen liegt, müssen Qualitätssicherungs-Maßnahmen funktions- und institutionsübergreifend implementiert werden. Ein typisches Qualitätssicherungs-Programm für eine Studie umfasst Schulung des Studienpersonals, umfassende Dokumentation, gute CRF, Datenmanagement-Systeme mit Plausibilitätschecks und Editiermöglichkeiten, Terminverwaltung, Qualitätsüberwachung, Monitoring, SDV und Abgleich zwischen Datenbank und CRF, regelmäßige Fortschrittsberichte, Berichte zur Datenqualität, Fehlerkorrekturen, wenn nötig, Revision von Studienprozeduren.

In multizentrischen Studien sind Qualitätssicherungs-Maßnahmen oft als Regelkreisläufe (feedback loop) angelegt, z. B. in der Query-Bearbeitung. Mögliche Fehlerquellen sind Protokollverletzungen, unzureichende oder ungeeichte Messgeräte, ungenaue/unlesbare/unvollständige Aufzeichnungen, ungenaue/unvollständige Dateneingabe, Fehler beim Daten- oder Materialtransport, fehlende Quelldaten, Übermaß an zu dokumentierenden Daten, unzureichende Schulungen, Datenfälschung, undokumentierte Änderungen an Datenbanken, Programmierfehler, Fehler beim Zusammenfassen und Berichten von Daten,

Missbrauch von Statistiksoftware oder Einsatz ungeeigneter statistischer Methodik. Regelmäßige Berichte und Informationen über die Qualität der Daten in der Studiendatenbank tragen zur Transparenz bei.

Automatisierte Plausibilitätschecks und Datenedits: die zu prüfenden Variablen sollten nach Relevanz ausgewählt werden. Das Editingsystem sollte vor Beginn der Datenerhebung geplant und implementiert werden. Das Personal in den Prüfzentren kann viele Prüfungen selbst durchführen, muss jedoch hierzu entsprechend angeleitet werden. Möglichkeiten zur Datenkorrektur sollten in allen Studienstufen gegeben sein. Automatisiert oder teilautomatisiert sind: Prüfungen auf Variablentyp, Range (Labor-Referenzbereichen für kranke Patienten oft schlecht anwendbar, ggf. eigene Referenzbereiche nach den ersten Patienten entwickeln, Unterschiede in Referenzbereichen verschiedener Labore beachten), Konsistenzprüfungen innerhalb eines CRF (Zeit- und Datumsfelder immer validieren, geschätzte Daten von genauen Daten unterscheiden) und CRF-übergreifend (alle CRF identifizierbar, alle benötigten CRF vorhanden, Zeitfenster eingehalten) sowie Konsistenzprüfungen über die Zeit (z. B. Visitenabfolge, Konsistenz bzgl. Variablen, die sich über die Zeit hinweg nicht oder kaum ändern). Alle Daten müssen eindeutig einem Patienten zuzuordnen sein. Für ID-Variablen ist ggf. eine Prüfsumme oder sprechende Patienten-ID sinnvoll. Ein Audit Trail soll ins Dateneingabesystem eingebaut sein. Wenn Datenerfassung ohne ein Datenbankmanagementsystem erfolgt, können keine übergreifenden Plausibilitätsprüfungen durchgeführt werden. Bei zentral und lokal verteilten DB müssen Synchronisationsprobleme ausgeschlossen werden. Werden Daten lokal eingegeben und dann in die zentrale Datenbank übertragen, werden in der Regel bei Übertragung weitere Checks durchgeführt. Weitere Plausibilitätsprüfungen werden in regelmäßigen Zeitabständen durchgeführt (z. B. Überprüfung von Mustern, Varianzen, zentrumsbezogene Vergleiche etc.) Editierprogramme sollten sorgfältig getestet und validiert werden. Neu eingeführte Plausibilitätsprüfungen sollten retrospektiv auf vorhandene Daten an-

gewendet werden, damit der Gesamtdatenbestand konsistent ist. Als letzte Stufe werden Daten vor der Auswertung nochmals editiert, wenn nötig.

Korrekturen müssen zeitnah angefordert und durchgeführt werden, ggf. kann der Status datensatz- oder merkmalsbezogen mit abgespeichert werden.

Queries werden an den Prüfarzt gesandt, von diesem beantwortet und abgezeichnet zurückgeschickt und dann erneut geprüft. Sie können manuell erzeugt oder automatisch generiert werden. Sie sollten ausführlich dokumentiert werden. Korrekturen können auch auf dem Original-CRF farbig markiert werden. In der Regel werden Queries und Antworten in einer Datenbank abgespeichert und die Zeit bis zur Antwort überwacht. Sobald die Query beantwortet wurde, werden die Daten korrigiert, entweder manuell oder automatisch in der Programmiersprache, die auch im DBMS eingesetzt wird. Dann werden sie den Plausibilitätsprüfungen unterzogen. Ungewöhnliche Daten werden, wenn als korrekt bestätigt, abgespeichert; dies muss dokumentiert werden. Das DBMS sollte einen vollständigen Audit Trail enthalten.

Die Vollständigkeit und Aktualität der Datenbank muss ständig überwacht werden, die Zentren erhalten regelmäßige Statusberichte und ggf. Aufforderungen. Die Zeitfenster für Erhebung und Transfer der Daten müssen eingehalten werden. Werden Daten nicht geliefert, sollte der oberste Prüfarzt im Zentrum mit angeschrieben werden. Aus Qualitätssicherungs-Gründen wird hier die Verwendung papierbasierter CRF empfohlen, bei computerbasierten CRF wird ein Ausdruck für SDV verwendet. Fehlerraten von 10 pro 10.000 Felder sind erreichbar. Werden Daten elektronisch übermittelt, muss die Zuverlässigkeit und Fehlerfreiheit der Übermittlung geprüft werden. Das wichtigste Maß für die Zuverlässigkeit eines Datenmanagement Systems ist die Restfehlerrate, d. h., die Anzahl der Fehler in der Datenbank, nachdem alle Qualitätssicherungs-Maßnahmen durchgeführt wurden. Sie muss in der Regel manuell überprüft werden. Ist die Fehlerrate hoch, sollte darüber berichtet und die Fehlerursachen identifiziert werden.

Datenqualitätsberichte enthalten z. B. zentrumspezifische Informationen zu Rekrutierung, Dokumentation, Fehlerraten etc. Referenzzentren und Zentrallabore sollen ähnlich überwacht werden wie die Prüfzentren. Die Studienleitung entscheidet über Ausmaß und Umfang dieser Überwachung. Auch hier werden ausgewählte Daten gemonitort. Bei Laborproben sollte eine Stichprobe doppelt eingesandt werden. Wo möglich, sollten Referenzbegutachtungen verblindet stattfinden. Die Anzahl nicht verwendbarer Proben soll berichtet werden. Das Datenzentrum sollte seine eigenen Qualitätssicherungs-Maßnahmen haben, u. a. Dokumentation aller durchgeführten Prozeduren, ungewöhnlichen Ereignisse und Pläne für den Fall, dass Daten verloren gehen oder Personal fehlt, Fehlerraten, Randomisationsüberwachung etc.

Um die zeitnahe und genaue Datenerhebung zu fördern, sollte alles Personal in den studienspezifischen und allgemeinen Prozeduren sowie medizinischen Aspekten und dem Qualitätssicherungs-System zentral geschult und ggf. zertifiziert werden. Pro Prüfzentrum sollten mindestens zwei Personen pro Aufgabe geschult werden, um eine Vertretung zu gewährleisten. Der Studienablauf wird durch Hilfen für die Prüfzentren (z. B. Terminpläne für einzelne Patienten oder pro Monat, Patientenlisten, Listen fälliger CRFs, Etiketten etc.) unterstützt.

Monitorbesuche können zu Beginn die Rekrutierung in den Vordergrund stellen, später wird meist die Protokollcompliance geprüft und SDV durchgeführt. Das Prüfpersonal kann Fragen stellen und die Datenhaltung überprüft werden. Zum SDV sollten die Papier-CRF und die Daten in der Datenbank herangezogen werden. Art und Umfang müssen vorher festgelegt sein. Jeder Monitorbesuch wird berichtet.

Zusammenfassung: Qualitätssicherung und Monitoring sind integraler und ständiger Bestandteil einer Studie, es müssen proaktive Maßnahmen von vornherein eingebaut werden. Die Datenqualität sollte ständig überwacht und dokumentiert werden. Zum Studienende sollten die Resultate der Qualitätssicherung in einen Bericht einfließen. Feedback

und corrective action sollten im Problemfall prompt erfolgen.

Gaus 2003

Die Qualitätssicherung in klinischen Studien sollte durch Personen erfolgen, die unabhängig von den Personen sind, die die Studie tatsächlich durchführen, und routinemäßig in den Studienablauf eingebunden sein. Die Qualitätssicherung kann sowohl intern als auch durch eine externe Institution oder Firma vorgenommen werden. Ein externes Auditing orientiert sich an der Struktur-, Prozess- und Ergebnisqualität und folgt den einschlägigen Normen DIN/ISO 9000 bis 9004.

Zur Qualitätssicherung gehören:

- eine vollständige Dokumentation und Nachprüfbarkeit aller Vorgänge innerhalb einer klinischen Studie
- Standard-Arbeitsanweisungen (SOPs) als zentraler Bestandteil
- Protokollierung der Tätigkeit der Studienorgane
- von Prüfärzten unabhängiges Monitoring zur Sicherung der Datenqualität
- Überprüfung der Existenz und Anwendung von Verfahrensvorschriften (internes oder externes Auditing)
- Zertifizierung von Studieneinrichtungen oder Studien
- Ablage aller papierbasierten Studiendokumente in einem Trial Master File

Durch das Monitoring soll die Studientätigkeit des Prüfarztes kontrolliert werden. Nach einem Initiierungsbesuch besucht der Monitor jeden Prüfarzt regelmäßig. Die Besuchsfrequenz kann zeitgesteuert oder fallzahlgesteuert sein. Bei jedem Besuch prüft der Monitor, ob der Prüfplan eingehalten wird, dass alle in die Studie aufgenommenen Patienten vollständig dokumentiert sind und die Randomisierung ordnungsgemäß erfolgte. Zur Überprüfung der Existenz der Studienpatienten überprüft der Monitor die Originalkrankenakten dieser Patienten. Stichprobenweise oder für alle Patienten überprüft der Monitor, ob die Studiendaten mit der Originalkrankenakte übereinstimmen (Source Data Verification, SDV). Besonders wichtig ist die SDV bei den Haupt- und Nebenzielgrößen der Studie. Der Umfang der SDV ist in den SOPs festzulegen,

jedoch kann der Monitor bei schlechter Datenqualität den Umfang von sich aus erweitern. Die ausgefüllten Erhebungsbögen werden auf Vollzähligkeit, Vollständigkeit, Übereinstimmung mit dem Studienplan sowie Plausibilität und Widerspruchsfreiheit geprüft. Zu den weiteren Aufgaben des Monitors (oder Studiensekretariats) gehört die Überwachung des Rekrutierungstempos. Jeder Prüfarzt erhöht die Inhomogenität der Daten und diese wird nur durch eine entsprechende Anzahl an Patienten kompensiert. Deshalb werden u. U. Zentren mit nur wenigen Patienten von der According-to-Protocol-Auswertung ausgeschlossen. Monitore müssen regelmäßig geschult werden. „Diese Schulung ist besonders wichtig, wenn an einer Studie mehrere Personen das Monitoring durchführen, damit diese gleichartig arbeiten und somit für die Studie ein einheitlicher Datenkörper entsteht.“ Ein gutes Monitoring gewährleistet eine hohe Datenqualität, verursacht aber auch Kosten. Deshalb muss zwischen die Balance zwischen einer angemessenen Datenqualität und vertretbaren Kosten gefunden werden.

Grimes et al. 2005

Die Autoren stellen ihre kritische Haltung zur Guideline Good Clinical Practice der ICH (International Conference on Harmonisation) dar. Prinzipiell wird bemängelt, dass die Guideline nicht Evidenz basiert entwickelt wurde, sondern durch einen informellen Konsens zustande gekommen ist.

Das in der Guideline geforderte Monitoring bestätigt lediglich die Übereinstimmung zwischen den CRFs und der Originaldokumentation. Wenn die Originaldokumentation fehlerhaft ist, liefert das Monitoring keinen Vorteil. Entgegen einem weit verbreiteten Missverständnis verlangt die Guideline kein Monitoring aller Daten; das Monitoring einer Zufallsstichprobe kann ausreichend sein.

Hasford/Staib 1994

Betrug in klinischen Studien untergräbt die ethische Legitimation der Therapieforschung und ihre Akzeptanz in der Öffentlichkeit und gefährdet unter Umständen das Wohl der Patienten. Deshalb sollten bei klinischen Studien immer Maßnahmen zur Prävention und Er-

kennung von Fälschungen vorgesehen werden. Als Fälschung (oder Betrug) wird im Folgenden die absichtliche Täuschung durch das Erfinden, Verändern oder Verschweigen von Daten bezeichnet.

Zur Prävention sind verschiedene Maßnahmen sinnvoll:

- Aus- und Fortbildung in den Regeln der Forschung und Forschungsethik
- Offenlegung der Daten und ihre externe Begutachtung
- Vermeidung von Interessenskonflikten
- Leistungsorientierte Bezahlung der Prüfarzte
- Klare Regelungen der Autorenschaft bei Publikationen
- SOPs für die Berichterstattung von Betrugsfällen
- Vertragliche Regelungen mit externen Leistungserbringern (z. B. Labor)
- Öffentlichmachung von Betrugsfällen
- Maßnahmen den Prüfplan betreffend (siehe Seite 188 – 190)

Die folgenden Indizien können auf Fälschungen hinweisen:

- Schwierigkeiten, mit einem Prüfarzt einen Termin für das Monitoring zu vereinbaren
- Originaldokumente fehlen oder können durch den Monitor nicht eingesehen werden
- Patienteneinverständniserklärungen fehlen
- Alle CRFs sind mit dem gleichen Stift ausgefüllt
- Sehr akkurat und komplett ausgefüllte CRFs ohne irgendeine Durchstreichung/Korrektur
- Inkompatible Datumsangaben (z. B. ärztliche Untersuchungen am Wochenende)

Auch statistische Auswertungen können entsprechende Hinweise liefern:

- Deskription der Rekrutierungsmuster
- Abgleich der gescreenten mit den rekrutierten Patienten
- Abgleich der Randomisierungs- und Patientensequenz
- Plausibilitätskontrolle aller Datumsangaben
- Anzahl der Protokollverstöße pro Zentrum

- Anzahl der Rückfragen pro Zentrum/pro Patient
- Fehlende Angaben zu unerwünschten Ereignissen, wenn bekannt ist, dass solche auftreten müssten
- Auftreten unsinniger klinischer Muster
- Auftreten von Zahlenpräferenzen
- Übereinstimmung der Streuung wichtiger Variablen zwischen den Zentren bzw. mit Angaben aus der Literatur
- Nachweis bekannter Korrelationen
- Fehlen von Ausreißern

Khosla et al. 2000

Die Autoren geben eine Einführung in die Source Data Verification (SDV) und machen Vorschläge für ihre effiziente Durchführung. Die international beachteten Leitlinien zur Good Clinical Practice (GCP) verlangen SDV für klinische Studien der Phasen I bis IV. Als Originaldaten („Source Data“) werden alle Informationen aus Originalakten und zertifizierten Kopien von Originalakten betrachtet, die für eine klinische Studie benutzt werden. Die Originaldokumente („Source Documents“) umfassen alle Dokumente, die Originaldaten für eine klinische Studie enthalten, und erlauben die Evaluation der Studie und der Datenqualität der Studiendaten. Die Source Data Verification ist eine Evaluation der Konformität der Daten, die in den CRFs einer Studie eingetragen sind, mit den Originaldaten. Jeder Wert, der auf einer CRF eingetragen ist, muss sich irgendwo in der Originaldokumentation wiederfinden lassen. Die Sicherstellung der Vertraulichkeit der Daten der Studienteilnehmer durch alle Personen, die mit diesen Daten in Berührung kommen, ist bei der SDV von entscheidender Bedeutung. Die Studienzentren sollten alle Originaldokumente sicher aufbewahren und Maßnahmen treffen, um eine versehentliche oder vorzeitige Vernichtung von Unterlagen zu verhindern.

Bei einer SDV gehören die folgenden Daten zu den Schlüsseldaten, bei denen größere Fehler der wissenschaftlichen und ethischen Qualität einer Studie abträglich sein können:

- Daten im Zusammenhang mit dem Wirksamkeitsnachweis (Hauptzielparameter)
- Einschluss- und Ausschlusskriterien
- Anamnese und Vorbehandlung

- Ärztliche Untersuchung und Vitalfunktionen
- Datumsangaben der Visiten
- Unerwünschte Ereignisse
- Begleitmedikation
- Angabe, dass der Patient an der klinischen Studie teilnimmt, und das Datum der Einverständniserklärung

Für die Durchführung der SDV gibt es zwei Methoden:

- Zug-um-Zug Methode: Der Prüfarzt besitzt die Originaldokumente und beantwortet spezifische Fragen des Monitors zu den Daten ohne dem Monitor Einblick in die Originaldokumente zu gewähren.
- Direkte Methode: Der Monitor hat direkten Zugang zu den Originaldokumenten.

Die Vorgehensweise bei der SDV sollte in Standard Operating Procedures (SOPs) festgelegt werden. Die SDV sollte abhängig vom Studienfortgang und so oft wie möglich durchgeführt werden. Am besten werden einige SDVs in den frühen Stadien einer Studie durchgeführt, damit aufgedeckte Probleme gelöst werden können, bevor sie einen signifikanten Einfluss auf die Datenintegrität erlangen. Die GCP-Leitlinien machen keine Aussage über den Umfang der SDV. Jedes Unternehmen muss selbst entscheiden, ob es 100 % SDV für alle Patienten oder 100 % SDV für die Schlüsseldaten und eine Stichprobe der übrigen Daten durchführt. 100 % SDV ist z. B. bei einer kleinen Patientenzahl zweckmäßig. Der Umfang der SDV hängt von verschiedenen Faktoren ab (z. B. Datenvolumen, personelle Ausstattung). Ein gebräuchlicher Ansatz ist die Kategorisierung der Daten für die SDV, z. B. in „kritische“ und „unkritische“ Daten. Kritische Daten (z. B. Einverständniserklärung) sollten für jeden Patienten zu 100 % durch die SDV überprüft werden. Unkritische Daten werden nur für einen Teil der CRFs überprüft, üblicherweise eine zufällige Stichprobe von 15 % bis 25 %. Der Prozess der SDV sollte sorgfältig dokumentiert werden (u. a. Häufigkeit und Natur der festgestellten Fehler).

Die Ergebnisse der SDV sollten kritisch evaluiert werden. Ein akzeptabler Standard sollte in der SOP zur SDV festgelegt werden, z. B. 0 % Fehler bei kritischen Daten und < 0,5 % Fehler bei unkritischen Daten.

Knatterud 1981

Der Erfolg einer kooperativen klinischen Studie hängt entscheidend davon ab, dass die gesammelten Daten von guter Qualität sind und richtig behandelt und verarbeitet werden. Qualitätslenkende Maßnahmen sollten bereits in der Planungsphase einer Studie beginnen und bis zum Studienende (Berichterstellung) beibehalten werden. Das Ziel einer völlig fehlerfreien Studie ist mit vertretbarem Aufwand nicht erreichbar. Stattdessen sollte angestrebt werden, dass die Zahl der Fehler klein ist und die Fehler zufällig zwischen den Behandlungsgruppen verteilt sind.

Eine der ersten Maßnahmen zur Qualitätslenkung sollte die Einrichtung eines Komitees oder einer Gruppe sein, die für die Sicherstellung der Datenqualität verantwortlich ist und die autorisiert ist, Maßnahmen der Qualitätslenkung zu implementieren. Neben Prüfärzten sollten auch externe Experten in dieses Gremium aufgenommen werden.

Verschiedene Maßnahmen sollten zur Sicherstellung einer qualitativ hochwertigen Datensammlung in den teilnehmenden Studienzentren implementiert werden:

- Definition von Standardabläufen und Erstellung eines Operationshandbuchs
- Training und Zertifizierung des Personals für Datensammlung und Kodierung
- Besuche in den Zentren zur Kontrolle der Einhaltung des Studienprotokolls
- Durchführung von Datenanalysen zur Aufdeckung von Problemen bei der Datensammlung und -aufzeichnung

Die Überwachung der Arbeit des zentralen Koordinierungszentrums erfährt meist wenig Aufmerksamkeit. Der Autor hält ein internes und externes Audit des Koordinierungszentrums für sinnvoll. Das Audit sollte die Dokumentation, Randomisierung, Datenverwaltung und Datenanalyse umfassen.

Knatterud et al. 1998

Die Qualitätssicherung zieht sich als durchgängiger Prozess durch alle Studienstufen – Prävention, Auffindung und Korrektur von Fehlern bzw. Problemlösung. Die Auswahl ehrlicher Prüfärzte ist von großer Bedeutung, das Monitoring soll zufällige wie systematische Fehler aufdecken und Probleme sollen

zeitnah und effektiv gelöst werden. Jede Studie sollte Leitlinien für den Umgang mit verdächtigen Daten oder Fehlverhalten haben. Datenfehler dürfen nicht die Primärvariable beeinflussen.

Gute Prüfpläne und CRF können Datenfehler vorbeugen. Wenn möglich, sollte ein Pretest durchgeführt werden. Der Qualitätssicherungs-Plan sollte in ein Qualitätssicherungs-Handbuch eingehen und für kritische Daten die Zeitfenster, Methodik und Verantwortlichen für Datenprüfung benennen. Vor Beginn sollte der site and data monitoring plan feststehen. Datenqualitäts-Ziele sollten definiert werden. Auch wenn der Hauptzweck des Monitoring nicht die Aufdeckung von Fälschungen ist, sollte es doch so geplant sein, dass man damit rechnet.

Prävention: Prüfärzte und ihr Personal sollten geschult werden, Review Committees sollten verblindet sein. Bei speziellen Prozeduren muss sichergestellt sein, dass Geräte und Know-How vorhanden sind. Zentrallabore und Referenzzentren bedürfen interner wie externer Qualitätssicherungs-Maßnahmen. Ein Monitorplan mit Benennung der Verantwortlichen, Checkliste, Ausmaß und Umfang des SDV sollte vorliegen. Wenn möglich: mindestens ein Besuch nach Rekrutierungsbeginn und zum Center Closing, während der Studie bei Laufzeiten ≥ 3 Jahre. Klare Prozeduren für die Aufbewahrung von Quelldokumenten, Audit Trail und Archivierung müssen definiert sein.

Problemfindung: Prüfärzte sollen die CRF durchgehen, Plausibilitätsprüfungen programmgestützt bei Eingabe und Edit-Prozeduren, Überprüfung auffälliger Laborwerte, Abgleich der endgültigen Daten stichprobenartig mit den Originaldaten, Diskrepanzen müssen von den Prüfärzten gelöst werden. Beim Monitoring soll alles Personal anwesend sein, Prozeduren und Datenhaltung werden geprüft und ein SDV sowie die Überprüfung der Einwilligungserklärungen sowie der Randomisierung durchgeführt. Es wird ein Monitorbericht erstellt. Zwei- bis dreimal jährlich sollen Berichte über den Fortschritt in den Zentren und der Gesamtstudie erstellt werden, auch bezüglich Fehler. Mit Hilfe statistischer Methoden können Fehler aufgedeckt werden. Der Daten-

erheber soll persönlich identifizierbar sein. Zentrallabore und Referenzzentren werden regelmäßig überwacht. Zur Auswertung sollten validierte Programme eingesetzt werden, diese werden archiviert. Die Prüfarzte sollten einem standard of performance folgen, Doppeleingabe ist zwar kostenintensiv, reduziert aber bestimmte Fehlertypen.

Problemlösung/Korrekturen: Außer bei vermuteter Datenfälschung geht es primär um Fehlerkorrektur. Diese sollte möglichst zeitnah zur Aufdeckung erfolgen (Query-Bearbeitung). Systematische Fehler können z. T. durch erneutes Review der Quelldaten korrigiert werden. Verblindung dient der Ausschaltung von Bias. Protokollverletzungen müssen berichtet werden. Alle eingebrachten Patienten sollten nach dem ITT (intent-to-treat) Prinzip ausgewertet werden. Bei Problemen mit Einzelpersonen kann Nachschulung oder Einzelaudit, ggf. auch Ausschluss eines Zentrums notwendig sein. Täuschung und Fälschung dürfen nicht geduldet werden, weil sie die Glaubwürdigkeit von Studien (oft überproportional) gefährden. Jeder Fall von Fehlverhalten oder Täuschung muss untersucht und an die entsprechenden Behörden oder Sponsoren und Studienleiter gemeldet werden. Für das Monitoring ist eine Kosten-Nutzen-Analyse sinnvoll. „Reducing the number of visits and reviewing only a sample of the records would allow a much less expensive quality assurance plan but may increase the number of undetected problems.“

Empfehlungen: am besten und kosteneffizientesten ist Prävention. Es sollte Monitoring bzw. Auditing durchgeführt werden, besonders zu Studienbeginn. Probleme sollten möglichst früh identifiziert und möglichst zeitnah gelöst werden. Bei Publikationen sollten die Qualitätssicherungs-Maßnahmen beschrieben werden.

Lachin 2004

In diesem Artikel steht Data Monitoring für (Zwischen-)analysen im Mittelpunkt. Der Autor weist darauf hin, dass in der FDA Guidance zu möglichen Interessenkonflikten (die entstehen, wenn ein Wissenschaftler beim Design und bei der Auswertung einer Studie beteiligt ist) nicht zwischen industriefinanzier-

ten und regierungsfinanzierten Studien unterschieden wird. Der Autor ist der Ansicht, dass die bisher z. B. von den NIH eingesetzte Vorgehensweise für Zwischenanalysen nicht geändert werden soll, obwohl sie nicht den FDA-Empfehlungen entspricht. Die NIH-Vorgehensweise wird kurz beschrieben. Als Unterschied zwischen industrie- und regierungsfinanzierten Studien wird angeführt, dass die Sponsoren und Prüfer einer NIH-finanzierten Studie in der Regel aus dem universitären Bereich stammen und üblicherweise keine versteckten finanziellen oder sonstigen Interessen bzgl. der Studienergebnisse haben und dass die Einrichtungen Regelwerke haben, in denen der Umgang mit Interessenkonflikten vorgeschrieben ist. Das Hauptproblem, das in der FDA Guidance angesprochen wird, ist der mögliche Einfluss von Zwischenanalysen auf weitere Entscheidungen in der Studie. Dies sollte auch in NIH-finanzierten Studien stärker beachtet werden. Der Autor sieht jedoch keine Notwendigkeit, die FDA Guidance auf öffentlich finanzierte Studien auszuweiten.

Marks 2004

Datensammlung in klinischen Studien erfolgte und erfolgt auf Papier mit Übertragung in eine Datenbank. Obwohl inzwischen viele EDC-Systeme (electronic data capture) existieren, ist die Effizienz der Datensammlung nicht gewachsen. Allmählich werden papierbasierte Quelldaten durch elektronische Quelldaten ersetzt – hierbei ergibt sich aber das Problem, wie diese Daten validiert werden können. Außerdem wird immer noch (zu) viel Papier eingesetzt.

Die Entwicklung ging von der zentralisierten Datensammlung (CRF werden eingesammelt und zentral in die Datenbank eingegeben, Prüfarzt muss Papier-CRF ausfüllen und Queries beantworten – lange Zeitspanne, viele Fehlerquellen) über Datensammlung am PC vor Ort (Remote Data Entry, Prüfarzt füllt Papier-CRF aus und diese werden vor Ort eingegeben mit Edit checks – dann per FTP an die zentrale Datenbank übermittelt) zur webbasierten Dateneingabe (EDC, erfolgte häufig noch immer vom Papier-CRF, Datentransfer schneller über WWW, mehr Edit Checks eingebaut). EDC hat sich als weniger effizient er-

wiesen als erwartet, vor allem wegen mangelnder Unterstützung durch die Anwender. In der Realität hat sich der Datenerhebungsprozess nicht wirklich geändert.

Verschiedene Probleme treten im Zusammenhang mit EDC auf: es wird immer noch Papier eingesetzt, die FDA verlangt eine Validierung (CFR Part 11 wird oft missverstanden), es gibt zu wenig mobile Hardware und die Einstellung vor allem der Monitore muss sich ändern (Datenvalidierungsprozesse sind auf Papier ausgerichtet).

Wie validiert man elektronische Quelldaten? Quelldaten kommen aus Akten, direkt vom Patienten und aus Laborsystemen. Letztere sind Präzedenzfälle für elektronische Quelldaten. Wenn man das Papier abschaffen will, was ist dann das „offizielle“ Dokument? Beim Papier ist die Patientenakte das Quelldokument. Wenn die Daten mit dem gleichen System sowohl für die elektronische Akte als auch für die Studie erhoben werden, können diese direkt bei der Eingabe validiert werden und sind dann auch für die Studie validiert. Dadurch ändert sich der Workflow für den Prüfarzt, er braucht mehr Zeit zur ersten Erhebung, weil er dabei direkt die Daten verifiziert. Danach wird aber die Zeit für das Monitoring eingespart. Auch andere Prozesse in klinischen Studien sollten auf elektronische Daten umgestellt werden, z. B. Drug Management.

Neaton et al. 1990

Der Artikel diskutiert zwei Methoden zur Sicherstellung hoher Datenqualität in klinischen Studien (Schulung von Prüfpersonal und doppelte Dateneingabe) anhand zweier Studien. Die Schulung des Prüfpersonals sollte zu Beginn und bei Langzeitstudien auch im Verlauf erfolgen und nicht nur die Studienunterlagen umfassen, sondern auch auf die Bedeutung schlechter Datenqualität für die Studienergebnisse eingehen. Bei Einsatz spezieller Messmethoden muss auch hierin geschult werden. An den Kosten für Schulungen sollte nicht gespart werden, da Fehlervermeidung effizienter ist als Fehlerfindung und -korrektur. Insgesamt können Schulungen daher die Kosten senken.

In der Treatment of Mild Hypertension Study wurden Schulungen vor Studienbeginn und

danach jährlich durchgeführt, wo nötig (z. B. bei neuem Personal) auch öfter. Die Kosten umfassten ca. 1–6 % des Gesamtbudgets der Studie für die 1. Phase. In der 2. (geplanten) Studienphase soll es regionale Schulungszentren und computerunterstützte Schulungen geben.

Weiterhin wurden in zwei Studien vier Dateneingabe-Methoden bzgl. der Fehlerrate auf 10.000 Felder verglichen: zentrale Doppeleingabe vom Band mit ausführlicher Prüfung nach der Eingabe und Beurteilung der Abweichungen durch eine dritte Kraft (3,5 Fehler), zentrale Einfacheingabe mit Prüfungen während der Eingabe (9,5 Fehler – bei Abzug eines durch einen Programmierfehler entstandenen Fehleranteils nur noch 1,1 Fehler), Eingabe im Prüfzentrum mit wenigen Prüfungen während und ausführlichen Prüfungen nach der Eingabe (26,1 Fehler) und Eingabe im Prüfzentrum mit ausführlichen Prüfungen während der Eingabe (24,6 Fehler). In anderen Studien lassen sich ähnliche Zahlen zeigen. Die Autoren schließen, dass eine Einfacheingabe zu vier bis fünf mal mehr Fehlern führt, dass die Fehlerraten für Textfelder höher sind als für numerische Felder, dass bei ausgedehnten Plausibilitätsprüfungen inklusive Cross Checking dafür geeigneter Daten Fehlerraten erreicht werden können, die mit Doppeleingabe vergleichbar sind, und dass in jedem Fall ein Qualitätssicherungsprogramm implementiert sein sollte.

O'Kelly 2004

Um die Effektivität statistischer Methoden zur Aufdeckung von Betrug bei klinischen Studien zu prüfen, wurde ein Experiment durchgeführt. Drei mit dem Montgomery-Asberg Depression Rating Scale (MADRS) vertraute Mediziner erfanden für eine Anzahl nicht existierender Patienten in drei nicht existierenden Zentren Werte für den MADRS. Die gefälschten Daten wurden mit echten Daten einer klinischen Studie aus 18 realen Zentren zusammengeführt und von einem verblindeten Statistiker analysiert. Der Median der Zentrumsgröße betrug zehn Patienten, der Median der Visits lag bei vier Visits pro Patient. Der Statistiker wusste nur, dass drei erfahrene Mediziner eine unbekannte Zahl von MADRS-Wer-

ten bei einer unbekanntem Zahl von Patienten und Zentren erfunden hatten, und er versuchte, über die Suche nach ungewöhnlichen Mittelwerten und Korrelationen die gefälschten Daten ausfindig zu machen.

Die statistischen Analysen zeigten nur eine geringe Spezifität zur Aufdeckung der gefälschten Daten. Eines der drei erfundenen Zentren wurde korrekt identifiziert, die anderen beiden erschienen unauffällig. Stattdessen wurde ein echtes Zentrum fälschlicherweise als möglicherweise erfunden eingeschätzt.

Um eine größere Spezifität zu erreichen, sollten mehrere statistische Tests miteinander und mit anderen in der Literatur beschriebenen Prüfungen kombiniert werden, z. B.:

- Daten, die sich von einem Visit zum nächsten nicht ändern, obwohl eine Änderung zu erwarten wäre
- Bevorzugung bestimmter Endziffern bei numerischen Werten
- eine zu geringe Zahl unerwünschter Ereignisse
- eine zu geringe Zahl von Begleitmedikationen

Das Experiment hat gezeigt, dass ein erfahrener Mediziner Daten in einer Weise fälschen kann, dass sich ihre Verteilung kaum von der realer Daten unterscheiden lässt.

O'Sullivan et al. 1996

In Großbritannien wurden in einer klinischen Studie der Phase IV 10.361 Patienten rekrutiert, um das Sicherheitsprofil eines zugelassenen Medikamentes zu untersuchen. Eine Source Data Verification wurde bei zufällig ausgewählten 13 % der Patienten durchgeführt.

Ose et al. 2004

Die Planung, Durchführung und Auswertung klinischer Studien der Phase III ist sehr teuer. Auf das Monitoring entfallen dabei ungefähr 25 % der Kosten. In der ICH-GCP wird unter Punkt 5.18.3 zum Monitoring ausgeführt: „Der Sponsor sollte sicherstellen, dass die klinischen Prüfungen angemessen überwacht werden. Der Sponsor sollte Umfang und Art des Monitorings festlegen, basierend auf den Überlegungen zu z. B. Zielsetzung, Zweck, Design, Komplexität, Verblindung, Umfang

und Zielparame-ter der klinischen Prüfung. Im Allgemeinen muss ein Monitoring vor Ort vor, während und nach der klinischen Prüfung stattfinden. In Ausnahmefällen kann der Sponsor jedoch entscheiden, dass ein zentrales Monitoring zusammen mit Maßnahmen wie Schulungen und Prüfertreffen und ausführlichen schriftlichen Anleitungen eine ordnungsgemäße Durchführung der klinischen Prüfung in Übereinstimmung mit GCP gewährleisten kann. Eine statistisch kontrollierte Stichprobennahme kann ein annehmbares Verfahren zur Auswahl nachzuprüfender Daten darstellen.“ Die hier aufgezeigten Möglichkeiten zur Aufwandsreduzierung beim Monitoring sollten insbesondere von universitären Einrichtungen mit ihren eingeschränkten finanziellen Möglichkeiten genutzt werden. Wichtig ist in jedem Fall die Vermeidung einer relevanten Qualitätsminderung.

Zur Aufwandsreduzierung gibt es verschiedene Möglichkeiten, die einzeln oder kombiniert genutzt werden können:

- Sequentialplan: Alle Zentren werden nach einer geringen Anzahl von rekrutierten Patienten vom Monitor besucht. Die Zeitspanne bis zum nächsten Monitorbesuch wird individuell für jedes Zentrum aus der vorgefundenen Qualität abgeleitet: Bei hoher Qualität wird die Zeitspanne verlängert, bei niedriger Qualität verkürzt.
- Zufällige Auswahl von Zentren/Patienten/Daten: Eine zufällige Auswahl von Zentren eignet sich besonders bei Studien mit hoher Zentrenzahl (ca. > 50) und geringer Patientenzahl pro Zentrum. Bei der zufälligen Auswahl sollten Zentren, die noch nie besucht wurden, mit einer höheren Wahrscheinlichkeit berücksichtigt werden als die anderen Zentren. Ein zufällige Auswahl von Patienten eignet sich besonders bei einer geringen Zahl von Zentren mit jeweils hoher Patientenzahl (ca. > 25).
- Vermehrter Einsatz von Prüfertreffen/Telefonkonferenzen: Bei Prüfertreffen kann unter Umständen ein Originaldatenabgleich stattfinden. In vielen Fällen kann auch der zeitaufwändige Studieninitiationsbesuch des Monitors durch ein Prüfertreffen ersetzt werden. Telefonkonferenzen können dazu dienen, medizini-

sche, technische und dokumentarische Probleme zu erörtern. Sie sind allerdings nur bei einer kleinen Zentrenzahl (max. 10) geeignet.

- Einsatz von Remote-Data-Entry: Die Datenerfassung erfolgt lokal in den Zentren und die Daten werden in die zentrale Datenbank der Studienzentrale übertragen. Damit stehen dem Monitor die Studierendaten zeitnah zur Verfügung und er kann von der Studienzentrale aus intervenieren.

Die folgenden Möglichkeiten können in Kombination mit den o. g. eingesetzt werden:

- One Person = One Study: Eine Person ist hauptverantwortlich für eine Studie von der Prüfplanerstellung bis zur Auswertung. Diese Methode eignet sich gut für kleinere Studien (ca. 10 Zentren und maximal 300 Patienten). Es ist keine Übergabe und kein detaillierter Informationsfluss zwischen Studienleiter, Projektmanager, Biometriker, Datenmanager und Monitor notwendig.
- Vermehrter Einsatz von preiswerten Allroundkräften, z. B. medizinischen Dokumentaren
- Telefonische Information statt Initiierungsbesuch
- Einsatz von dezentralen Monitoren: Dieses Verfahren eignet sich hauptsächlich für große Langzeitstudien mit einer zentralen Kontrolle durch das Datenmanagement.
- Verstärkte Auswahl von Zentren in räumlicher Nähe zur Studienzentrale
- Frühzeitiges Schließen von Zentren mit mangelhafter Qualität und schlechter Rekrutierungsquote
- Zeitnaher Versand der Prüfbögen an die Studienzentrale mit dortiger Kontrolle

Bei der Auswahl einer Methode zur Kostenreduzierung sollten folgende Punkte bedacht werden:

- Anzahl der Zentren und Patienten
- Dauer der klinischen Prüfung
- Indikation/Schwere der Erkrankung
- Rekrutierungsgeschwindigkeit
- Komplexität der klinischen Prüfung
- Phase der klinischen Prüfung

Pogash et al. 2001

Die Autoren beschreiben das Datenmanagement-System im Asthma Clinical Research Network.

In der Entwicklungsphase werden die Prüfpläne vom Steering Committee entwickelt und nach Review genehmigt. Zu den CRF gehören auch Verwaltungs- und Referenz-Unterlagen, sie werden von einem interdisziplinären Team unter Verwendung einheitlicher Formatvorlagen entwickelt und mit Ausfüll-Anweisungen und Procedure Checklists versehen. In den ersten fünf Studien wurden standardisierte Module entwickelt, die wieder verwendet werden können. Für große Studien wird Durchschreibpapier verwendet, die CRF sind wo möglich vorkodiert. Zu jeder Studie gibt es ein Operationshandbuch, die Patienten erhalten eine Patientenkarte. Die Datenbanken sind modular aufgebaut und werden zentral gehalten. Die Zentren übernehmen Datenmanagement-Aufgaben wie z. B. die Meldung und Randomisation.

Vor der 1. Eingabe von Daten muss ein Patient registriert werden, dabei wird eine ID erzeugt und auf Duplikate geprüft. Vor Datenerfassung wird eine Sichtkontrolle durchgeführt, die Masken sehen wie die CRF aus. Programmierte Plausibilitätsprüfungen sind eingebaut, bei Fehlern in Pflichtfeldern muss ggf. die Datenerfassung abgebrochen werden. Nach Ersteingabe werden die Durchschreibsätze getrennt und ans Datenzentrum verschickt, wo erneute Sichtkontrollen und die Zweiteingabe durchgeführt werden. Das DBMS enthält ein Validierungsmodul mit Plausibilitätsprüfungen, die nachts über alle Daten durchgeführt werden. Die Validierung kann auch interaktiv erfolgen. Pro Problem wird eine Query verschickt per e-Mail mit drei Standardvorlagen (fehlende Daten, unklare Daten, fehlende CRF). Ein Query Tracking Modul unterstützt die Query-Verwaltung.

Zur Überwachung der Datenqualität werden Protokollverletzungen überwacht und berichtet ebenso wie die Rekrutierung und das Follow-Up pro Patient.

Als Parameter für die Datenqualität dienen u. a.:

- die Anzahl fehlender oder nicht zeitgerechter Visits,

- die Zeitfenster zwischen den Datenverarbeitungs-Schritten (Eingang, Queries etc.),

- Anzahl und Bearbeitung der Queries.

Wichtige Daten müssen auf speziellen Formularen authentifiziert werden, z. B. für kritische Sicherheitsparameter. Alle 18 Monate werden die Zentren besucht und auditiert, bei schlechter Datenqualität auch öfter. Um den Datentransfer zu prüfen, werden 5 % der wichtigen und 1 % der anderen CRF zufällig ausgewählt und gegen die Datenbank verglichen. Werden mehr als zehn Abweichungen pro 10.000 Felder gefunden, müssen weitere 5 % der CRF abgeglichen werden. Die Resultate werden in einem Auditmodul festgehalten und dem Studienstatistiker zugeleitet.

Inzwischen ist das DBMS um ein Validierungs- und Editmodul für die Zentren vor Ort erweitert worden, so dass mehr Korrekturen schon bei der Ersteingabe durchgeführt werden können. Ein weiteres Modul erlaubt elektronisches Tracking von Protokollverletzungen, die Verifikation des Zertifizierungs-Status für das Personal vor Ort und die Eingabe von Quelldokumenten wurden ebenfalls automatisiert.

Pollock 1994

In der pädiatrischen Onkologie sind multizentrische Studiengruppen von großer Bedeutung, da die Inzidenz gering ist und fast alle Kinder nach Studienprotokollen behandelt werden. Diese Gruppen erfordern nicht nur die üblichen Qualitätssicherungsmaßnahmen, sondern auch Kommunikation, zentrales Datenmanagement, Monitoring und hohe Compliance bzgl. der Behandlungsstandards. Die Pediatric Oncology Group umfasst mehr als 1.000 Wissenschaftler in 106 Institutionen.

Die Prüfpläne werden im Statistischen Zentrum entworfen, parallel zur Planung erfolgt der Entwurf von CRF, Datenbank inklusive Plausibilitätsprüfungen und Regeln für das Monitoring. Die Pläne werden internen und externen Reviews unterzogen. Die Meldung und Randomisierung erfolgt zentral über eine automatisierte Telefon-Hotline. Wo möglich, werden einheitliche CRF-Module verwendet. Die Papier-CRF sind vorkodiert und werden

zeitnah doppelt eingegeben. Alle zwei Wochen erfolgen automatisierte Plausibilitätsprüfungen im Batch-Modus, der Datenmanager prüft die Fehler und schickt eine Query an das Zentrum, wenn sich der Fehler nicht anhand der bereits vorhandenen Daten auflösen lässt. Für die halbjährlich stattfindenden Prüfarzttreffen werden Evaluationsbögen für jeden Studienkoordinator zusammengestellt, auf dem die Tumorresponse, Toxizitäten usw. dokumentiert werden.

Die Qualitätssicherung erfolgt durch Zuweisung von Scores für Compliance bei der Patienteneinbringung, dem Behandlungsschema und der Auswertbarkeit. Schon bei der Meldung finden Referenzbegutachtungen und -klassifikationen statt, damit nur eligible Patienten eingebracht werden. Die Auswertbarkeit bezieht sich auf Daten, Proben und Befunde, auf die Einhaltung der Behandlungsprotokolle und die Patientencompliance. Je nach Studienphase (I bis III) sind die Auswertbarkeitskriterien in Teilen unterschiedlich formuliert. Die Scores werden folgendermaßen gebildet:

- Score 1 summiert die gemeldeten Patienten nach früh (Meldung vor kurzem erfolgt, Daten noch nicht vorhanden), vollständig auswertbar, teilweise auswertbar, gemeldet für Gruppenprotokolle und gemeldet für Phase-I-Studien bei kooperierenden Forschungsinstituten.
- Score 2 (Kurzzeit-Auswertbarkeit) ist der Anteil der teilweise und vollständig auswertbaren Patienten an der Gesamtzahl der vollständig und teilweise auswertbaren, nicht eligiblen und nicht auswertbaren Patienten, die in den letzten 12 Monaten gemeldet wurden.
- Score 3 (Langzeit-Auswertbarkeit) entspricht Score 2, umfasst aber die gemeldeten und ausgewerteten Patienten der letzten 12 Monate.

Als Feedback-Mechanismus werden vor den halbjährlichen Treffen Berichte generiert. Daraus resultieren Berichte an die Leitkommission bzgl. Ausschluss aus der Gruppe. Von Zentren wird erwartet, dass der durchschnittliche Score bei 2 liegt, andernfalls gibt es vier 6-monatige Zyklen Zeit zur Korrektur. Werden in dieser Zeit keine Verbesserungen erreicht,

wird das Zentrum ausgeschlossen. Für alle laufenden Studien werden, wo sinnvoll, Follow-Up-Scores gebildet und die Rekrutierungsrate wird ebenso wie das Auftreten unerwünschter Ereignisse engmaschig überwacht. Die Datenüberwachung (Data Monitoring) ist der wichtigste Aspekt der Qualitätslenkung in der Pediatric Oncology Group (POG). Zwei Kommissionen, die Leitkommission und das Institutional Performance Review Committee, überwachen Qualitätslenkungsaufgaben. 35 zufällig ausgewählte Zentren werden pro Jahr einem Onsite-Monitoring unterzogen, pro Zentrum findet mindestens alle drei Jahre ein Monitorbesuch durch ein Team aus einem erfahrenen Prüfarzt und einem erfahrenen Datenmanager statt. Pro Zentrum werden 16 Patienten zufällig ausgewählt, stratifiziert nach Studientyp. Neben dem Originaldatenabgleich findet eine Prüfung der Protokollcompliance und Medikamentenverwaltung (Drug Accountability) statt. Ernste Probleme führen zu einem zweiten Besuch oder der Suspension der Patientenmeldung. Auf den halbjährlichen Treffen finden regelmäßige Schulungen statt (eintägige Einführung für neue Datenmanager sowie je ein Tag Fortbildungsprogramm). Aus den Ergebnissen einer Leistungsprüfung der Zentren für 11.719 zwischen 1981 und 1991 eingebrachte Patienten ergibt sich, dass Zentren mit höheren Rekrutierungsraten bessere Leistungen zeigen. Hauptzentren schneiden besser ab als assoziierte Zentren, und die Werte bessern sich insgesamt deutlich über die Zeit. Dies ist u. a. auf rasches Feedback zurückzuführen.

Poy 1993

Obwohl die FDA die erste Instanz war, die GCP etabliert hat, haben die europäischen GCP-Leitlinien als erste die Notwendigkeit eines Qualitätssicherungssystems betont. Der Sponsor ist für die Einführung und Pflege eines Qualitätssicherungssystems verantwortlich, das Qualitätslenkungsmaßnahmen umfasst.

Die Qualität von Daten in klinischen Studien ist zunächst von der Qualität der Quelldaten abhängig. Nach ihrer Erhebung und Erfassung durch den Prüfarzt werden sie mehrfach bearbeitet und transformiert, bevor ein Ab-

schlussbericht generiert wird. Daher müssen Maßnahmen zur Qualitätslenkung in allen Schritten des Datenverarbeitungsprozesses integriert sein. Sie werden in SOPs festgeschrieben. Zusätzlich werden Audits durchgeführt durch Personen, die von den entsprechenden Abteilungen unabhängig sind.

Da klinische Studien in einem multidisziplinären Umfeld durchgeführt werden, ist jeder Mitarbeiter, der mit den Daten zu tun hat, Teil einer Qualitätskette und jedes Kettenglied ist von Bedeutung. Da die Stärke einer Kette von ihrem schwächsten Glied abhängt, sollten klinische Studien mit einem Quality In-Quality Out Ansatz durchgeführt werden, bei dem für jedes Kettenglied entsprechende Qualitätslenkungsprozeduren und gute Kommunikationsstrategien vorgesehen sind. Schlüsselfiguren sind der Sponsor, das Monitorteam und die Prüfarzte. Zu Beginn sollten klare Definitionen und Informationen dafür sorgen, dass in einem Klima der Transparenz und des Vertrauens kommuniziert werden kann.

Der Sponsor übernimmt die Verantwortung für die Durchführung, wählt qualifizierte Kooperationspartner aus und benennt die Monitore. Die Rolle der Monitore und Clinical Research Assistants ist essentiell für die Sicherstellung von Qualität, da sie das Bindeglied zwischen Prüfarzt und Sponsor darstellen und die Prüfung der Datenqualität eine der Haupttätigkeiten beim Monitoring im Zentrum ist. Das Monitorteam sollte Probleme zeitnah aufdecken und berichten, damit Korrekturmaßnahmen eingeleitet werden können. Monitore müssen neben Fachkenntnissen über gute Kommunikationsfähigkeiten mit einer Mischung aus Überzeugungskraft und Diplomatie verfügen. Auch die Auswahl von Prüfarzten hat einen maßgeblichen Einfluss auf die Datenqualität. Sie sollten über die Studienziele ebenso informiert sein wie über ihre Verpflichtungen im Rahmen der GCP. Die Bedeutung von Originaldaten und Originaldatenabgleich muss zu Beginn der Studie klar formuliert werden.

Zu guter Datenqualität trägt neben entsprechend qualifiziertem Personal auch die Integration von Prüfungs- und Kontrollsystemen bei, die Fehlerraten auf ein Minimum reduzieren. Die geplante Rekrutierungsrate sollte rea-

listisch sein. Bei Monitorbesuchen wird neben der Durchführung des Originaldatenabgleichs auch die Verfügbarkeit aktueller Studienunterlagen (Studienordner, Prüfärztbroschüre etc.), die Einhaltung der Ein-/Ausschlusskriterien und der Umgang mit der Prüfmedikation (Drug Accountability) und die fortlaufende Eignung des Zentrums für die Studie geprüft. Auf einer höheren Ebene überprüft ein Clinical Quality Assurance Auditor die Studienaktivitäten und berichtet direkt dem höheren Management. Ggf. wird eine entsprechende Arbeitsgruppe eingerichtet, die von den Forschungs- und Entwicklungsabteilungen unabhängig ist. Sie führt stichprobenartige Audits in allen relevanten Bereichen durch. Dabei werden zum einen Studiendurchführung und -dokumentation überprüft, zum zweiten die externen Partner (Auftragsforschungsinstitute) und zum dritten die Effizienz der Qualitätslenkungsmaßnahmen selbst. Darüber hinaus unterstützt ein solches Team die anderen Abteilungen durch Schulungen und Beratungen und bereitet sie auf externe Audits z. B. durch Behörden angemessen vor.

Renard/van Glabbe 1989

Der Artikel stellt das Datenmanagementsystem des EORTC-Datenzentrums vor. In klinischen Studien werden Therapieeffekte verglichen. Für die Auswertung sind Endpunkte, Baseline-Werte und Follow-Up-Daten von besonderer Bedeutung. Qualitätslenkungsmaßnahmen sollten auf die Endpunkte und prognostischen Faktoren fokussieren. Follow-Up-Daten spielen eine wichtige Rolle bei Konsistenzprüfungen, da die Endpunkte häufig erst am Studienende berichtet werden, und können für Zwischenanalysen herangezogen werden. Die Daten im Master Data File sollten richtig, über die Zentren hinweg vergleichbar kodiert und vollständig sein. Die Vergleichbarkeit der Bewertung zwischen den Zentren spielt ebenfalls eine große Rolle. Werden zu viele Patienten wegen unvollständiger Daten ausgeschlossen, so entsteht ein „statistischer Bias“.

Die EORTC entwickelt Studienprotokolle leitliniengesteuert und in standardisiertem Format. Die Meldungs- und Randomisierungskriterien werden zentral festgelegt und Datener-

hebungsbögen entwickelt. 1983 wurde eine Studie zur Datenqualität durchgeführt, die die Notwendigkeit der Datenverifikation bestätigte. Bei der Dateneingabe werden sowohl einzelne als auch feldübergreifende Plausibilitätsprüfungen (range checks, cross checks) durchgeführt. Darüber hinaus findet eine Sichtkontrolle der eingegangenen Bögen statt. Fehlende Werte und fehlende Bögen ebenso wie Korrekturen inkonsistenter Werte werden per Telefon, Post oder E-Mail angefordert. Die Prüfärzte sind über Korrekturen ihrer Daten in der zentralen Datenbank zu informieren. Monitorbesuche vor Ort dienen der Bewertung der Protokollcompliance und der Prüfung der Daten (Originaldatenabgleich, Prüfung der Studienprozeduren z. B. bei der Rekrutierung oder Medikation, ethische Belange). Bei der EORTC arbeiten der Studienkoordinator, der Datenmanager und der Statistiker im Team zusammen.

Roos et al. 2003

In einer randomisierten klinischen Studie zur Untersuchung neuropathischer Schmerzen bei Knochenmetastasen wurden Audits einschließlich Source Data Verification zur Qualitätssicherung eingesetzt. Nach einem initialen Audit der Daten der ersten 90 Patienten wurde das weitere Audit nach folgendem Verfahren durchgeführt:

1. Wenn es keine Protokollverstöße bei den letzten fünf Patienten eines Zentrums gab, wurde zufällig einer der fünf Patienten für das Audit ausgewählt.
2. Jeder festgestellte Protokollverstoß eines Zentrums führt zu einem Audit der nächsten fünf Patienten dieses Zentrums.

Rouff/Child 2003

CPCRA ist ein community based program für AIDS-Forschung mit 15 Forschungsstandorten, die klinische Studien durchführen. Qualitätsverbesserungsmaßnahmen sind u. a. Patientenerinnerungsschreiben, Ausbildung und Schulungen für Personal und Patienten, Datenchecks, Peer Support. Die Organisation hat die CQI-Prinzipien (Continuous Quality Improvement) eingeführt. Dies wurde bei klinischen Studien bisher weniger angewendet. Im Gegensatz dazu werden z. T. Quality Con-

trol (QC) Maßnahmen angewendet, d. h. die Probleme werden gelistet und bearbeitet, nachdem sie aufgetreten sind.

Das Problem der QC-Maßnahmen ist, dass sie nicht die Qualität sichern, sondern erst im Nachhinein greifen. Daher hat sich CPCRA für CQI entschieden. Hierzu wurden verschiedene Treffen durchgeführt und z. B. Standards für Patient-lost-to-FU entwickelt. 1995 haben alle beteiligten Zentren Qualitätsmanagement-Handbücher entwickelt. Zur Beurteilung der performance wurde vor allem peer review and support eingesetzt, außerdem gibt es ein Komitee, das auch Sanktionen verhängen kann. Vierteljährliche Berichte beschreiben die Datenqualität (studienspezifisch und -übergreifend) und müssen schriftliche Verbesserungspläne enthalten (welche Maßnahmen werden bis wann ergriffen und was soll dabei konkret herauskommen). Für Schlüsselvariablen wurden Standards definiert, es wurden Terminverwaltungssysteme aufgebaut und das Berichtswesen verbessert. Prüfpläne wurden genauer formuliert und monatliche FAQ-Sheets erstellt sowie monatliche Telefonkonferenzen durchgeführt.

CQI wurde auf zwei Ebenen implementiert: in der Komitee-Struktur und vor Ort. Das Performance Assessment Committee (PAC) prüft vierteljährlich die Performance vor Ort anhand von Qualitätsindikatoren (Benchmark 75 % der Gesamtqualität), geht die Berichte mit den Betroffenen vor Ort durch, diskutiert sie per Telefonkonferenz und bietet den Zentren Unterstützung an (z. B. bei Schulungen, mit Plänen etc.). Was auf dieser Ebene nicht gelöst werden kann, wird an das Performance Oversight Committee verwiesen, das auch sanktionieren kann. Auf der Ebene der lokalen Zentren wurden QI-Handbücher erstellt und implementiert.

Ergebnisse des CQI: Datensammlung effektiver, Entwicklung einer Vorlage zum Identifizieren und Lösen von Problemen, mehr Kommunikation, vor Ort mehr Kontrolle über den Prozess, mehr Teamwork, bessere Performance. Veränderungen bezogen sich auf statistical center reports (z. B. wurden dadurch Auffälligkeiten eher identifiziert), story boards (Verbesserungspläne im plan-do-check-act-Zyklus), mehr Treffen, Reviews für Akten und Daten

und Verlagerung der Verantwortung auf die, die direkt an den Schlüsselmaßnahmen beteiligt sind. Es wurden drei Maßzahlen überprüft: unberichtete unerwünschte Ereignisse, unberichtete Outcomes, nicht eligible und trotzdem aufgenommene Patienten. Dabei zeigten sich substantielle Verbesserungen über einen Zeitraum von 6,5 Jahren.

Schmidt et al. 1995

Betrug und Fehlverhalten bei klinischen Studien kann durch eine Kombination von biometrischen Methoden und On-site Audits aufgedeckt werden. Die Autoren schildern mehrere Fälle aus ihrer eigenen Arbeit. In einem Fall wurden signifikante Inkonsistenzen zwischen den CRFs und den Originaldaten (z. B. Laborbögen) gefunden. Einfache deskriptive Statistik erwies sich als sinnvoll. Als geeignete Methoden des On-site Audits erwiesen sich:

- Interviews
- Review objektiver Daten, wie z. B. EKG, Laborbögen, Nachweise über Verwendung und Verbleib der Studienmedikation, In-Haus-Dokumentation, Schriftwechsel mit konsultierten Ärzten

Die folgenden biometrischen Methoden wurden eingesetzt:

- Vergleich und Untersuchung von Laborwerten und Blutdruckwerten
- Überprüfung der Einhaltung der geplanten Visiten
- Vergleich von Inzidenz, Schweregrad und Muster unerwünschter Ereignisse zwischen Zentren
- Berechnung der Streuung und von Ausreißern
- Pharmakokinetische Methoden (Untersuchung von Wirkstoffkonzentrationen im Blut)

Während die hier vorgestellten Methoden verhältnismäßig einfache statistische Verfahren nutzen, hat die ABPI Fraud Statistics Working Party im Jahre 1993 eine umfangreiche Liste biometrischer Verfahren zur Aufdeckung von Betrug und Fehlverhalten bei klinischen Studien zusammengestellt und dabei auch komplexere Verfahren miteinbezogen, wie z. B. Cluster-Analysen und Untersuchungen von Kovarianz-Matrizen.

Von August 1990 bis Dezember 1994 wurden von den Autoren 234 On-site Audits durchgeführt, die meisten davon in Europa und Südafrika. In den zufällig oder aufgrund ihrer hohen Patientenzahl ausgewählten Zentren wurde in einem Fall ein betrügender Prüfarzt entdeckt (0,43 %). Bei weiteren acht Verdachtsfällen konnte in drei Fällen der Betrug bestätigt werden. Inzidenzschätzungen für Betrug in Europa bewegen sich im Bereich von 0,4 % bis 7 % und Audits der FDA im Zeitraum von 1977 bis 1990 zeigten eine Inzidenz von 1,7 %. Ein häufigeres und rigoroseres Monitoring einschließlich einer frühzeitigen Source Data Verification verringert die Wahrscheinlichkeit betrügerischen Verhaltens. Biometrische Methoden können als Screening-Instrument zur Auswahl verdächtiger Zentren für ein Audit benutzt werden.

Svolba 1999

In seiner Dissertation beschäftigt sich der Autor mit statistischen Verfahren zur Qualitätslenkung in klinischen Studien (Regressionsanalyse, Breakpoint-Analyse, One-step-ahead-forecast-error, Shewhart-charts) für normalverteilte Daten und Ereignisse, die eine Binomial- oder Poisson-Verteilung aufweisen. Daneben stellt er Methoden zur Überwachung von Entwicklungen über die Zeit vor. Die Theorie wird an Beispielen aus einer Melanom-Studie verdeutlicht. Für die vorgestellten Verfahren hat der Autor SAS-Makros entwickelt, deren Quellcode in Auszügen in dieser Arbeit enthalten ist.

Svolba/Bauer 1999

Das Papier behandelt die Anwendung statistischer Methoden zur Prozess- und Qualitätskontrolle für klinische Studien. Ein wichtiger Anwendungsbereich dieser Methoden ist die Prüfung, ob ein bestimmter Prozess sich stabil verhält (z. B. anwendbar auf Rekrutierung/Population, Fehlerrate pro CRF etc.), die Methoden sollen Änderungen im laufenden Prozess aufdecken. Bei relevanten Änderungen muss schnell gehandelt werden. Desweiteren prüft die Prozesskontrolle, ob eine Handlung die beabsichtigten Folgen hat (z. B. hat eine neue Rekrutierungsstrategie die Rekrutierungszahl erhöht). Die Methoden lassen sich

ohne große Zusatzkosten für Studien implementieren.

Es werden keine zusätzlichen Daten benötigt. Mit Hilfe einer Zeitachse lassen sich Subgruppen bilden, die den Produktionschargen entsprechen. Dabei sind die Stichprobengrößen verschieden. Zur Überwachung können Patientendaten, aber auch Metadaten (z. B. Zeit zwischen Visite und CRF-Eingang, Fehlerraten) verwendet werden. Trends in der Wirksamkeit weisen nicht notwendigerweise auf instabile Prozesse hin. Wiederholte Messungen werden gemittelt. Das Risk set per calendar time kann definiert werden als Anzahl der Patienten im Intervall oder als kumulative Zeit, die Patienten im Intervall verbringen. Rasche Fehlerkorrekturen sind seltener möglich als in der Produktion.

Messwerte und Ereignisdaten (Event Data) sind zur Überwachung gut geeignet. Für Event Data muss die number of units at risk ermittelt werden. Die Zeitvariable für Kalenderintervalle liegt in der Mitte des Intervalls, für Regressionsanalysen wird das Datum verwendet. In Control Charts (Shewart Charts) werden die Stichproben für ein Intervall durch den Mittelwert, die Spannweite und die Standardabweichung beschrieben. Event-Daten werden mit p-Charts (Proportionen), c-charts (Häufigkeiten) und u-Charts (Raten) aufgetragen.

Zur praktischen Anwendung der Methoden statistischer Qualitätskontrolle bei klinischen Studien haben die Autoren einen Satz SAS-Makros entwickelt. Diese Makros übernehmen auch Aufgaben des Datenmanagements und gehen über die in SAS/QC verfügbaren Verfahren hinaus. Die Anwendung der SAS-Makros auf eine zweiarmige Langzeitstudie erfolgte im Rahmen einer Doktorarbeit des Erstautors.

Vantongelen et. al. 1989

Die EORTC hat eine formalisierte Prozedur entwickelt, um die Genauigkeit der Datensammlung und die Effektivität der Datenübertragung zu beurteilen. Hierzu werden die Zentren besucht, SDV durchgeführt, der Prozess der Datensammlung beobachtet und die Gründe für Abweichungen notiert. Die Vollständigkeit und Klarheit der eingesetzten Protokolle wird geprüft.

Für die Studie wurden drei klinische Studien ausgewählt, es wurden pro Studie fünf Zentren besucht und mit Checklisten alle Items auf den CRF geprüft. Die Akten wurden zufällig ausgewählt. In zehn von 15 Zentren gab es administratives Personal, in den anderen gaben die Ärzte die Daten selbst ein. Die Datenqualität wurde kodiert nach

- Art und Bedeutung des Fehlers (korrekt, kleinere Abweichung, wichtige Abweichung, fehlend, nur auf CRF vorhanden) und
- nach den Fehler-Ursachen (administrativer Fehler, Akte falsch interpretiert, Akte falsch, Information in Akte nicht vorhanden, Übertragungsfehler, CRF unklar formuliert).

Nach dem Besuch wurde ein Bericht erstellt. Elf der 15 Zentren haben > 90 % korrekte Daten. Die Schwankungen rühren vor allem von den Daten her, zu denen es keine Source Data gibt (z. B. weil Patient teilweise vor Ort weiterbehandelt wird). Die mediane Fehlerrate betrug 2,8 %, Hauptgrund sind Übertragungsfehler. Die Datenfehler-Raten waren nicht wesentlich verschieden in Zentren mit und ohne Dokumentationsassistenten. Weitere Einflüsse auf die Fehlerrate waren: Strukturen im Datenmanagement vor Ort (Bedeutung gut ausgebildeter und erfahrener Data Manager, effizienter Organisation) statt „wandernder“ Data Manager, die Daten verschiedenster Studien erheben und die Einzelstudien nicht gut genug kennen; Qualität der Daten in den Akten; schlechte Ansprechbarkeit der Ärzte; verspätetes Ausfüllen der CRF. Das Problem der Datenqualität in den Akten ist besonders für die Toxizitäten relevant, da diese selten in den Akten so dokumentiert werden, wie es für die Studien nötig ist. Am schlechtesten war der Performance Status dokumentiert. Außerdem gab es Unterschiede in der Interpretation der WHO Toxicity Gradings und der Prüfplan war nicht immer klar. Datenmanagement vor Ort ist nicht der Schlüssel zur Datenqualität, sondern das Wissen vor Ort und die gute Organisation mit eingebauten internen Kontrollen. Die Ärzte sollten unterstützt werden z. B. durch Erinnerung, Zusammenfassung und Anmerkungen in der Akte.

Weiss et al. 1993

Publik gewordene Betrugsfälle in klinischen Studien haben den Eindruck erweckt, als sei dies ein weit verbreitetes Phänomen. Bisher liegen nur Daten von Shapiro und Charrow vor, die die Ergebnisse von FDA-Audits untersucht haben und einen Anteil von 7 % für ernste Probleme gefunden haben. Das NCI führt Studien unter der Aufsicht von Studiengruppen durch, da die FDA hier keine Aufsichtsfunktion hat. 1978 wurde ein Onsite-Monitoring für alle Studiengruppen verpflichtend gemacht. Eine der Studiengruppen ist die Cancer and Leukemia Group B (CALGB), deren Monitor-Ergebnisse hier berichtet werden. Dem Data Monitoring Committee der CALGB gehören ca. 15 erfahrene Personen (Ärzte, Study Nurses, Datenmanager) an, die den Großteil des Monitorings durchführen. Etwa 20 % der Monitorbesuche werden vom NCI auditiert. Mindestens einmal alle drei Jahre werden die Zentren besucht und eine Stichprobe der bisher eingebrachten Patientendaten gemonitort. Das Zentrum erhält die Patientenliste 30 Tage vor dem Besuch. Die Monitorteam umfassen zwei bis acht Personen, es arbeiten jeweils ein Arzt und ein Datenmanager zusammen und führen den Originaldatenabgleich sowie die Prüfung weiterer protokollrelevanter Aspekte (z. B. Einwilligung) mit Hilfe von Checklisten durch. Protokollverletzungen werden in triviale (werden nicht dokumentiert), kleinere (Daten sind trotzdem verwendbar) und größere (Daten sind möglicherweise oder de facto nicht verwendbar) unterteilt. Der Teamleiter fasst einen schriftlichen Monitorbericht mit Angabe der Protokollverletzungen sowie einer Einschätzung der Datenqualität. Ist die Qualität im Zentrum zufriedenstellend, erfolgt der nächste Besuch zum Regeltermin. Ist die Qualität nicht zufriedenstellend, wird ein Folgebesuch in den nächsten 6 bis 12 Monaten empfohlen. Wird das Zentrum ein zweites Mal als nicht zufriedenstellend eingestuft oder treten ernste Probleme auf, kann es für die Studien geschlossen werden. Im Durchschnitt werden etwa 15 % bis 20 % der Patienten gemonitort.

Seit dem Beginn des Monitorings 1982 wurden alle Zentren viermal regulär besucht und dabei 3.787 Patienten gemonitort. Der Anteil

größerer Protokollverletzungen bzgl. ethischer Aspekte (z. B. kein Ethikvotum vor Einbringung des 1. Patienten) lag dabei zwischen 11,9 % und 49,6 %. Der Anteil von Problemen mit Einverständniserklärungen lag zwischen 2,6 % und 25,1 %. Die Monitore überprüften auch, ob die Ein-/Ausschlusskriterien beachtet (zwischen 89,3 % und 95,5 % waren tatsächlich eligibel) und ob Tumorresponse und Medikamentendosis korrekt bestimmt wurden (die Fehlerrate bzgl. größerer Protokollverletzungen in der Dosierung lag durchgängig bei etwa 11 %). Ebenfalls überprüft wurden die Unterlagen, die für die zentrale Referenzbegutachtung von Gewebeproben einzureichen waren. Bei allen untersuchten Problemfeldern zeigt sich, dass sich die Werte über die Zeit verbesserten und in der Regel die Hauptzentren weniger Protokollverletzungen aufwiesen als die assoziierten Zentren. Es wurde ein Fall von Fälschung entdeckt, in dem Daten von 25 Patienten so verändert worden waren, dass der Eindruck perfekter Protokollcompliance entstand. In einem Fall wurden 50 Patienten, die nicht eligibel waren, durch Irrtum eines Prüf arztes in eine Studie eingebracht. Die Rate von Fehlverhalten liegt also bei zwei von 691 gemonitornten Zentren (0,28 %). Über die vier Monitorzyklen hinweg schieden drei Hauptzentren und 96 assoziierte Zentren aus, zumindest teilweise wegen schlechter Beurteilung nach dem Monitoring.

Zu wissenschaftlichem Fehlverhalten in klinischen Studien liegen wenige Daten vor. Die einzigen Daten, die auf Monitoring basieren, sind die von Shapiro und Charrow, die schwerwiegende Fehler in 7 % bis 12 % von der FDA auditierten Zentren fanden. Sie konnten zeigen, dass die Anzahl über die Zeit etwas abnahm. Durch die von der CALGB implementierten Qualitätssicherungsmaßnahmen (z. B. Prüfung durch Datenkoordinatoren in den Datenzentren) werden viele Fehler früh entdeckt. Monitoring wirkt nicht nur qualitätslenkend, sondern sorgt auch in den Zentren für einen Anreiz zur Verbesserung, da diese sonst ggf. von Studien ausgeschlossen werden. Die Southwest Oncology Group, die ebenfalls NCI-Gelder erhält, hat bei ihrem bisherigen Monitoring noch keine Fälschungsfälle entdeckt. Eine andere Gruppe hatte drei Fälle von Fehl-

verhalten, einer davon war der Prüf arzt, der von der CALGB als Fälscher entdeckt wurde. Insgesamt kann zumindest für die NCI-Studien gezeigt werden, dass Fälschung und Fehlverhalten Einzelfälle sind. Durch das Monitoring hat sich die Compliance bzgl. Einreichung von Materialien zur pathologischen und radiologischen Referenzbegutachtung verbessert. Die CALGB betrachtet das Monitoring als erfolgreich sowohl zur Aufdeckung von Fälschungen als auch zur Aufrechterhaltung der Motivation bzgl. Protokollcompliance und Ausschluss schlechter Zentren.

Weiss 1998

Das NCI ist der weltgrößte Sponsor von Krebstherapiestudien. Die von ihm finanzierten elf Forschergruppen (cooperative groups) setzen sich aus Hauptzentren und assoziierten Zentren zusammen. Diese Gruppen haben unterschiedliche Schwerpunkte und führen ggf. auch gruppenübergreifende Studien durch. Jede Gruppe hat ein zentrales Verwaltungsbüro und (je) ein Zentrum für Datenmanagement und Statistik. Für die Datenqualität ist neben Prüf arzten und Datenmanagern vor Ort das Datenzentrum der Gruppe verantwortlich, in dem Patienten gemeldet und die Daten gesammelt, überprüft und Rückfragen generiert und bearbeitet werden. Auf der dritten Ebene werden die Daten von der Studienleitung auf Protokollcompliance geprüft, zusätzlich sind ggf. zentralisierte Referenzbegutachtungen implementiert. Diese vielfachen Qualitätslenkungsmaßnahmen sorgen für zuverlässige Daten. Sie beruhen jedoch darauf, dass das Personal vor Ort ehrlich, sorgfältig und genau arbeitet und dokumentiert. Daher werden in regelmäßigen Abständen Audits (Monitorbesuche) durchgeführt, wie sie die FDA schon seit längerem implementiert hat und wie sie vom NCI seit 1978 gefordert werden. Der Artikel beschreibt die Methodik des Monitoring in der Cancer and Leukemia Group B (CALGB, zu den Ergebnissen s. Weiss et al. 1993). Das Data Audit Committee (DAC) hat einen hauptamtlichen Vorsitzenden und 20 weitere Mitglieder, je zur Hälfte Prüf arzte und Datenmanager. Diese arbeiten ehrenamtlich ohne weitere Vergütung für die Monitorbesuche. Ein Monitorteam besteht aus zwei bis acht

Personen, je zur Hälfte Prüfarzte und Datenmanager. Die Teamgröße wechselt je nach Größe des zu besuchenden Zentrums. Der Teamleiter ist immer ein DAC-Mitglied, weitere Monitore können bei Bedarf hinzugezogen werden.

Das NCI hat Monitor-Richtlinien herausgegeben, die von den Gruppen je nach Bedarf ergänzt wurden. Die Richtlinien der CALGB umfassen z. B.:

- maximales Monitorintervall 36 Monate
- alle Zentren können jederzeit besucht werden
- alle Zentren werden besucht, wenn sie auch nur einen Patienten gemeldet haben
- es werden primär, aber nicht ausschließlich, die seit dem letzten Besuch gemeldeten Patienten gemonitort
- Zentren mit Bewertung „nicht zufriedenstellend“ werden nach sechs bis zwölf Monaten erneut besucht
- alle Ergebnisse werden den verschiedenen Kommissionen und der Gruppenleitung berichtet
- alle Studienprotokolle können gemonitort werden

Während eines Monitorbesuchs werden geprüft:

- das Vorliegen eines Ethikvotums vor Einbringen des ersten Patienten
- die Compliance bzgl. jährlicher Reviews, Berichten unerwünschter Ereignisse und Review von Protokoll-Amendments
- Umgang, Lagerung und Ausgabe der Studienmedikation
- Prüfung, ob der Inhalt der Einverständniserklärungen richtig verstanden wurde
- Prüfung der Protokollcompliance bzgl. Einverständniserklärungen (korrekt datiert und unterschrieben, ggf. übersetzt), Ein/Ausschlusskriterien, Randomisierung, Behandlung nach Protokoll, Erfassung und Meldung von Toxizitäten, Tumorresponse-Beurteilung und Datenqualität (Vollständigkeit und Zeitnähe der Dokumentation)
- Originaldatenabgleich

Der Hauptprüfer im Zentrum wird drei bis fünf Monate vor einem Monitorbesuch informiert. Mindestens 10 % der gemeldeten Pa-

tienten (bei wichtigen Studien bis zu 40 %) werden für das Monitoring ausgewählt, davon wird mindestens einer unangekündigt gemonitort. Der Arzt des Monitorteam prüft die Krankenakten, der Datenmanager die ausgefüllten CRFs. Beide arbeiten anhand standardisierter Checklisten. Der Besuch dauert ein bis zwei Tage, es werden etwa acht bis zehn Patientenakten geprüft und die Ergebnisse auf Formularen festgehalten, die die Grundlage für den schriftlichen Monitorbericht darstellen. Am Ende des Besuchs findet eine Besprechung mit dem Prüfpersonal vor Ort statt.

Der Monitorbericht wird mit Hilfe eines von der NCI verteilten Standardprogramms erstellt und dem NCI, dem leitenden Prüfarzt vor Ort und den verschiedenen Komitees der CALGB zugestellt. Protokollabweichungen werden tabelliert, im Text dargestellt und als gering (Daten verwendbar und gültig) bzw. schwerwiegend (Daten fragwürdig bzw. nicht auswertbar) bewertet. Das Zentrum erhält eine Gesamtbewertung in einer der folgenden Kategorien: zufriedenstellend, zufriedenstellend aber Follow-Up nötig, nicht zufriedenstellend. Nicht zufriedenstellende Zentren werden in sechs bis zwölf Monaten wieder besucht. Das NCI kann eine Modifikation der Bewertung (meist hin zum Schlechteren) und wiederholte Besuche fordern. Die Ergebnisse der ersten vier Monitorzyklen sind in Weiss et al. 1993 berichtet. Es zeigte sich eine deutliche Verbesserung der Datenqualität. In einer anderen Studiengruppe wurde ein Fälschungsfall aufgedeckt, das CALGB konnte zwei Fälle von Fehlverhalten entdecken. Seit 1984 ist dies jedoch nicht mehr aufgetreten. Das Monitoring dient nicht nur der Verifikation der Daten, sondern hat auch erzieherische Wirkung auf die Prüfarzte und Datenmanager. Wenn ein Prüfarzt weiß, dass Kollegen ihm über die Schulter schauen, verbessert sich die Qualität und davon profitiert die Studiengruppe ebenso wie die Öffentlichkeit, die sich darauf verlassen kann, dass berichtete Studienergebnisse verlässlich sind.

Williams 1997

Der Artikel ist ein Kommentar zu weiteren Vorträgen in einer Session. In den vergangenen Jahren ist das Vertrauen in klinische Stu-

dien durch Fälle von Fehlverhalten gesunken. Verdächtige Daten müssen jedoch nicht immer Ausweis von Betrug sein. Es muss zwischen Betrug, Irrtum, Inkompetenz, Missverständnissen und Bias klar unterschieden werden. Maßnahmen zur Vermeidung systematischer und Berücksichtigung zufälliger Fehler sollten schon in der Planungsphase einsetzen. Protokolle sollten nicht zu komplex sein und nicht zu viele Daten enthalten. Definitionen sollten so einfach und so klar wie möglich sein. Die Auswahl der Prüfarzte sollte sorgfältig erfolgen (Erfahrung, Patientengut, ausreichende Ressourcen, Vermeidung von Interessenkonflikten). In der Durchführung hilft der Einsatz von (verblindeten) Endpunkt-Komitees und zentralen Referenzbeurteilungen, gutes CRF-Design, zentrale Randomisation, Schulung und Zertifizierung des Personals, frühzeitige und regelmäßige Datenprüfung inklusive Vergleich zwischen den Zentren, Onsite-Monitoring, Qualitätssicherungs-Audits. Onsite-Monitoring hat eine hohe Rate von falsch positiven Ergebnissen (viele Abweichungen, aber wenig Fälschungen). Die Validierung von Daten sollte entsprechend ihrer Bedeutung bzw. des Einflusses von Fehlern auf die Ergebnisse durchgeführt werden. Vor der Auswertung sollte ein statistischer Analyseplan erstellt werden. Bei Verdacht auf Fälschung sollte rasch und deutlich gehandelt werden.

Witte et al. 2000

Fehlverhalten und Betrug sind gesellschaftliche Phänomene, die auch in Wissenschaft und Forschung verbreitet sind. Im Rahmen von klinischen Studien findet man immer wieder Datensätze, deren Plausibilität zweifelhaft ist. Neben anderen Gründen kommt ein Fehlverhalten des Prüfarztes als Ursache in Frage. In diesem Fall lassen sich unbewusstes, bewusstes und betrügerisches Verhalten unterscheiden. Unbewusstes Fehlverhalten entsteht durch Unkenntnis und/oder mangelnde Qualifikation. Zu letzterem zählen auch Nachlässigkeit und ein mangelndes Qualitätsbewusstsein. Bewusstes Fehlverhalten reicht von der Interpolation eines nicht vorhandenen Befundes bis zur Erfindung kompletter Patientenunterlagen oder Forschungsergebnisse und

geht fließend in Betrug über. Betrug ist die wissentliche Täuschung eines anderen mit dem Ziel, dem Getäuschten Schaden zuzufügen, und ist strafrechtlich relevant. Über Ausmaß und Hintergründe für Fehlverhalten oder Betrug gibt es keine verlässlichen Angaben. Eine hohe Dunkelziffer ist zu erwarten.

Der Nachweis eines Fehlverhaltens oder Betruges ist schwierig. Eine zeitnahe Datenerfassung, Plausibilitätskontrollen mittels deskriptiver Statistik und der regelmäßige Abgleich mit den Originaldaten sind wichtige Hilfsmittel zur Erkennung von Manipulationen.

Zur Prävention von Fehlverhalten und Betrug sind eine sorgfältige Auswahl, Aufklärung und Training der Prüfarzte von großer Bedeutung.

Zettelmeyer et al. 2005

Die Anforderungen bezüglich Sicherheit und Qualität sind für Therapieoptimierungsstudien mit der 12. AMG-Novellierung und der GCP-Verordnung gestiegen. Damit stellt sich die Frage nach dem notwendigen Umfang eines Onsite-Monitorings bei solchen Studien. Die Deutsche Hodgkin Studiengruppe (DHSG) führt zurzeit in Zusammenarbeit mit der Arbeitsgruppe Qualitätsmanagement des Kompetenznetzes Maligne Lymphome probeweise ein stichprobenartiges Onsite-Monitoring bei drei Studien durch. Es wurden zufällig 50 Prüfzentren mit zirka 100 Patienten ausgewählt, wobei die drei Studien und die Typen der Zentren gleichmäßig berücksichtigt wurden. Zwischenergebnisse von 21 Prüfzentren mit insgesamt 44 Patienten liegen bereits vor. Bei der Bestimmung des Hauptzielkriteriums und der Dokumentation unerwünschter Ereignisse und schwerwiegender Nebenwirkungen gab es keine Beanstandungen. Bei etwa der Hälfte der Patienten fehlten beim Staging gemäß Studienprotokoll vorgeschriebene Untersuchungen aus bildgebenden Verfahren oder Laboruntersuchungen. 25 Dokumentationsbögen (7,6 % aller begutachteten Bögen) wiesen Mängel – im Wesentlichen bei der Dokumentation der Therapiedurchführung – auf, die anhand der Originalakten korrigiert werden konnten.

Die Zwischenergebnisse zeigen die Notwendigkeit eines Onsite-Monitorings zur Siche-

zung der Datenqualität. Nach Abschluss des Probemonitorings werden aus den Ergebnissen der Umfang des Datenabgleichs und die Besuchsfrequenz des Onsite-Monitorings abgeleitet.

2.4.7.4 Datenqualität in der medizinischen Versorgung

Aronsky/Haug 2000

Die Umwandlung von papierbasierten in computerunterstützte Leitlinien kann Doppelerfassung von Daten vermeiden und Ausfüllaufwand senken. Hierzu müssen aber die Daten im rechnerunterstützten Teil des Krankenhausinformationssystems (KIS) entsprechend vorhanden sein, z. B. indem die benötigten Daten mit einer strukturierten Maske erfasst werden. Eine andere Möglichkeit ist, die benötigten Daten direkt aus der elektronischen Patientenakte (EPA) abzufragen. Die vorliegende Studie bewertet die Datenqualität der Routinedaten, die in den Pneumonia Severity Index (PSI) eingehen, die entsprechende Leitlinie ist relativ komplex.

Das LDS Hospital hat 520 Betten. Die EPA wird durch handgeschriebene Einweisungsnotizen und andere papierbasierte Formulare ergänzt, die offizielle Akte hat also einen EPA und einen Papier-Anteil. Es wurden Daten aller erwachsenen Patienten mit Entlassungsdiagnose community-acquired pneumonia zwischen Juni 1995 und November 1996 verwendet. Die Datenerfassung erfolgt durch die Pflegekräfte in der Notfallabteilung, teils kodiert und teils als Freitext. Laborwerte werden aus dem Laborsystem direkt übernommen. Der Arztbericht und eventuell vorliegende Röntgenberichte werden als Freitext eingegeben.

Die PSI-Guideline ist ein Schweregrad-System, das die Mortalitätsabschätzung ermöglichen soll. Sie hat einen zweischrittigen Algorithmus, in einem ersten Schritt werden elf Variablen aus der Anamnese und körperlichen Untersuchung abgefragt, in einem zweiten weitere Variablen aus dem Labor und der Röntgenuntersuchung. Der Risikoscore dient der Ermittlung einer Risikoklasse, die u. a. für die Krankenhaus-Einweisung relevant ist. Für jeden Patienten wurde der PSI Score einmal auf der Basis von Daten aus dem KIS (per

Abfrage) ermittelt und einmal Daten aus der Akte (EPA und Papier). Als Referenz galten alle verfügbaren Daten. 24 Patienten dienten als Re-Evaluations-Stichprobe. Die Differenzen wurden für den Risikoscore und die Risikoklasse ermittelt. Für die benötigten Variablen wurde die Datenqualität gemessen als

- Verfügbarkeit (availability = Anzahl vorhanden)
- Übereinstimmung (Concordance = Anzahl identisch mit dem Referenzstandard).

Dabei wurde zwischen erforderlichen und optionalen Variablen getrennt. Fehler wurden nach Datenformat (Freitext, kodiert, Labor, Vitalzeichen) und nach prozessbedingten und algorithmusbedingten Fehlern unterschieden.

Die Risikoklassen waren zu 86,7 % identisch, die Risikoscores zu 72,1 %. Abweichungen gingen in beide Richtungen. Die durchschnittliche Verfügbarkeit betrug 0,991 für die erforderlichen und 0,788 für die optionalen Variablen. Die Übereinstimmung betrug 0,982. 64 % der Fehler traten in Freitextvariablen auf. 77 % der Fehler waren auf Fehler in der Akte zurückzuführen, 23 % der Fehler waren Systemfehler.

Barnett et al. 1978

Die Beurteilung der Versorgungsqualität bei ambulanten Patienten wirft besondere Probleme auf, da die Leistungserbringer weitgehend unabhängig arbeiten, und die Daten- bzw. Aktenhaltung und Dokumentation nicht standardisiert sind und es große Unterschiede im Gebrauch der Terminologie gibt. Folgende Definitionen werden gegeben:

- Qualitätsmessung (quality assessment) ist ein Versuch, die Qualität der Versorgung zu messen, der Fokus liegt dabei auf der Sammlung und Darstellung von Daten bzgl. vergangener Ereignisse.
- Qualitätssicherung (quality assurance) kombiniert die Qualitätsmessung mit der systematischen Anwendung von Korrekturmaßnahmen, um Verbesserungen zu erzielen.
- Feedback kann retrospektiv für eine Gruppe von Patienten oder parallel (concurrent) für einzelne aktuell behandelte Patienten erfolgen.

Der Artikel beschreibt ein Qualitätssicherungssystem, in dem die Daten laufend analysiert und entsprechende Feedback-Mechanismen zeitnah implementiert werden, so dass das Feedback situativ angemessen und früh genug für Korrekturen erfolgen kann. Von zentraler Bedeutung ist dabei das EDV-basierte medical information system COSTAR, das seit acht Jahren im Harvard Community Health Plan im Einsatz ist und papierbasierte Krankenakten weitgehend ersetzt hat. Behandlungsdaten werden anhand von vorkodierten Checklisten eingegeben, Labordaten online eingelesen. Es ist keine weitere Kodierung nötig, und die Daten, die dem Qualitätssicherungsprozess zugrunde liegen, sind die unmittelbar erhobenen Versorgungsdaten. Es wird ein definiertes Vokabular eingesetzt; die Daten können auch für Verwaltungszwecke verwendet werden.

Durch programmgesteuerte Plausibilitätsprüfungen kann eine Qualitätssicherung direkt bei der Dateneingabe erfolgen. Solche Programme sind für verschiedene Erkrankungen implementiert, als Beispiel dient hier die Behandlung von Streptokokken-Pharyngitis, bei dem z. B. ein Mahnschreiben generiert wird, wenn nicht vier Tage nach positivem Laborbefund eine Behandlung dokumentiert wird. Hierdurch wurde die Dokumentationsrate für die Behandlung deutlich erhöht.

Eine der Schwächen von medizinischen Qualitätssicherungsprogrammen war ihre Unfähigkeit, die Erziehungsschleife (educational loop) zu schließen, weil die Daten nicht unmittelbar für Qualitätssicherungsmaßnahmen einsetzbar waren oder nicht ausreichende Korrekturmaßnahmen implementiert wurden. Für ein effektives Feedback darf die Datenüberwachung keinen unannehmbar großen Teil der Systemressourcen beanspruchen (z. B. durch zusätzlichen Dokumentationsaufwand), die Daten müssen richtig und vollständig sein (z. B. durch Vermeidung unnötiger Datentransformationen, Eingabe direkt durch das Pflegepersonal) und das Feedback muss zeitnah erfolgen und ggf. Veränderungen auf institutioneller Ebene anstoßen. Durch das interaktive Computersystem kann die Qualitätssicherung implementiert und selbst in ihrem Ablauf überwacht werden.

Barnett/Winickoff 1990

Ein Feedback, das einen Vergleich mit anderen Ärzten umfasst, ist ein mächtiges Werkzeug der Qualitätssicherung in der medizinischen Versorgung. Sein Wert hängt ab von der Akzeptanz durch die betroffenen Ärzte. Zur Erreichung dieser Akzeptanz ist es wesentlich, dass der Vergleich in der Weise anonymisiert geschieht, dass die Identität eines einzelnen Arztes nur ihm selbst bekannt ist.

Zwei weitere kritische Faktoren sind bei Qualitätssicherungsprogrammen zu berücksichtigen. Zum einen muss der zurückgemeldete Fehler oder Abweichung vom Standard von dem verantwortlichen Arzt als relevant angesehen werden. Zum anderen muss das Feedback so frühzeitig erfolgen, dass Zeit für eine angemessene korrigierende Reaktion bleibt.

In vielen Fällen spielt eine computerbasierte Patientenakte eine Schlüsselrolle beim Monitoring von Patientendaten und der Rückmeldung von Abweichungen. Für die Effektivität eines solchen Systems halten die Autoren folgende Eigenschaften für essentiell:

- Die für die Qualitätssicherung benutzten Daten müssen mit den Routinedaten identisch sein. Eine zusätzliche Datenerhebung kann nicht sein.
- Die für die Qualitätssicherung benutzten Daten sollten soweit wie möglich kodiert sein.
- Es ist wünschenswert, dass die medizinischen Patientendaten zusammen mit Verwaltungs- und Abrechnungsdaten in einer gemeinsamen Datenbank gehalten werden.

Barrie/Marsh 1992

Es wurde die Datenqualität der Manchester Orthopaedic Database gegen die Daten der Hospital Activity Analysis überprüft. Diese Datenbank umfasst ein auf die Orthopädie spezialisiertes Textverarbeitungsprogramm und eine Forschungsdatenbank mit klinischen Daten. Es werden Schlüsselwörter zu Diagnosen, Prozeduren und Komplikationen aus den Arztdiktaten in die Datenbank eingegeben. Pro Einsatzort prüft ein Systemkoordinator regelmäßig die Schlüsselwörter, kontrolliert die Einführung neuer Begriffe und korrigiert ggf. Fehler.

Ein Autor prüfte zu zwei verschiedenen Zeitpunkten 200 bzw. 121 zufällig ausgewählte Fallberichte und Akten.

- Für jede Visite wurde die ideale Anzahl von Schlüsselwörtern ermittelt und mit der tatsächlich eingegebenen Zahl verglichen (Vollständigkeit) und
- die Genauigkeit definiert als Anteil der vorhandenen und korrekten Schlüsselwörter von allen vorhandenen.

Es ergab sich eine Vollständigkeit von 61,7 % und eine Richtigkeit von 96,5 %. Die Richtigkeit der Schlüsselwörter für Diagnosen, Operationen und Komplikationen betrug 96,5 %, 97,8 % und 92,9 %.

Bei stationären Patienten wurden eher Schlüsselwörter eingegeben als bei ambulanten, aber es gab keine signifikanten Unterschiede in der Genauigkeit. Ab Dezember 1989 wurde die Vergabe von Schlüsselwörtern in den Besprechungen thematisiert. Danach verbesserte sich die Vollständigkeit und Genauigkeit. Bei neuen Schlüsselwörtern war die Datenqualität besser als beim Einfügen fehlender Schlüsselwörter. Der Vergleich mit der Hospital Activity Analysis zeigte, dass die Krankenhaus-Daten vollständiger, die Activity-Daten genauer waren. Die 38 % fehlenden Schlüsselwörter wurden häufig beim Diktat vergessen. Am größten war der Fehl-Anteil bei den Komplikationen. Die Autoren befürworteten den weiteren Einsatz von Schlüsselwörtern zur Kodierung. Nutzerschulungen verbessern die Datenqualität. Aufgrund der Studie wurde die Datenbank so geändert, dass statt Freitext eine Arztbriefschreibung automatisiert anhand der Schlüsselwörter erfolgt.

Brown et al. 2001

Mitarbeiter in der Primärversorgung verlassen sich immer mehr auf die Qualität der im rechnerunterstützten Teil der Krankenhausinformationssysteme (KIS) gespeicherten Daten. Die Autoren beschreiben einen neuen Ansatz zur Untersuchung der Datenqualität für alle Prozesse, die bei der Begegnung von Arzt und Patient ablaufen (z. B. Untersuchung, Interpretation der Befunde, therapeutische oder investigative Intervention). Data Quality Probes (DQPs) sind Anfragen, die bei einer perfekten Leistung des Arztes und einer perfekten

Datenhaltung keine Ergebnisse liefern. Ein nichtleeres Ergebnis weist auf einen Behandlungs- oder Datenfehler hin.

Eine DQP („Nur Patientinnen mit einer vorangegangenen Hysterektomie sollten eine bestimmte Formulierung von Östrogen erhalten“) wurde in einer Gemeinschaftspraxis mit drei Standorten in Großbritannien getestet. Die DQP lieferte als Ergebnis 105 Fehler aus 775 Patientinnen (13,6 %). Der größte Teil davon ließ sich auf Fehler der im KIS gespeicherten Daten zurückführen.

Brown/Warmington 2002

Um Qualitätsstandards zu erreichen und CQI (kontinuierliche Qualitätsverbesserung) zu ermöglichen, werden in Großbritannien zunehmend rechnerunterstützte Praxisinformationssysteme und KIS eingesetzt. Es gibt jedoch wenig methodische Forschung zur kontinuierlichen Qualitätsüberwachung dieser Daten.

Data Quality Probes (DQPs) basieren auf Abfragen in einem Praxisinformationssystem oder KIS, deren Ergebnis als Maß für die Systemperformance herangezogen werden kann. Im idealen System von Behandlung und Aufzeichnung würde keine von ihnen ein Ergebnis liefern, d. h., jedes Ergebnis einer solchen Query weist auf Fehler hin (z. B. Patient mit Penicillinallergie hat Penicillin erhalten).

Das System wurde in einer englischen Arztpraxis mit drei Behandlungsorten und einem gemeinsamen Praxissystem getestet. Es wurden beispielhaft DQPs entwickelt und angewendet, im Fehlerfall wurden die Daten und Papierakten retrospektiv überprüft. Die Fehlerraten pro DQP reichten von 0 bis 90,8 % (bei hohen Fehlerraten wurde in der Regel nicht dokumentiert statt schlecht behandelt, außerdem gab es Probleme, zwischen alten und aktuellen Diagnosen, Daten etc. zu unterscheiden, weil dies nicht mit aufgezeichnet wurde). Es gab aber auch Fehler in der Query-Formulierung und beim Retrieval. Inzwischen werden die Abfragen monatlich durchgeführt und werden gut angenommen.

Die Autoren empfehlen mehr Schulungen und standardisierte Protokolle für die Eingabe von Daten in Praxis- und Kliniksysteme.

Cook et al. 2003

Das Projekt IMPACT ist ein Datenbank-System, das die Versorgung von Intensivpatienten misst und beschreibt. Ärzte können ihre Behandlung sowie die Outcomes mit anderen Stationen vergleichen. Es werden in 96 Kliniken mehr als 100 Variablen erfasst, vierteljährlich in die zentrale Datenbank überspielt und Berichte generiert. Die Daten werden zu Mortalitätsberechnungen, aber auch als Basis für Entscheidungen über Behandlung und Policy verwendet.

In einer Pilotstudie wurde die Datenqualität geprüft durch Vergleich zwischen den in zwei Kliniken einmal manuell eingegebenen und einmal elektronisch abgefragten Daten und Daten, die von einer unabhängigen Person aus den Akten extrahiert wurden. Es wurden zwölf Items aus den Aufnahme-, Labor- und Entlassungsdaten ausgewählt und pro Klinik 45 Stichproben-Patienten geprüft. Die Extraktion erfolgte anhand des Leitlinien-Manuals der IMPACT-Gruppe, die Daten wurden in eine dem Datenerfassungsprogramm entsprechende Maske eingegeben. Für stetige Variablen wurde die Übereinstimmung in Ranges bestimmt, für andere als exakte Übereinstimmung. Kodierte Variablen wurden auf exakte (alle 4 Ziffern gleich) und ungefähre (erste 2 Ziffern gleich) Übereinstimmung geprüft. Die Übereinstimmung wurde pro Klinik bestimmt, um die verschiedenen Erhebungswege zu berücksichtigen.

Die Übereinstimmung lag je nach Item zwischen 52,3 % und 100 %. Statistisch signifikante Unterschiede zwischen elektronisch und manuell erfassten Items gab es nur für drei Variablen, zwei davon aus dem Labor-Bereich. Zwischen den Kliniken waren die Unterschiede für die Aufnahme-Items am größten. Die vier Entlassungs-Items zeigten gute Übereinstimmung. Für die kodierten Variablen waren die Ergebnisse bei Prüfung der ersten beiden Stellen besser als für alle vier Stellen.

Die Ergebnisse zeigen für die meisten Daten-Items eine Übereinstimmung von über 80 %. Dies entspricht den Daten einer Studie, die die Übereinstimmung zwischen verschiedenen Kodierern der ICNARC prüfte. Es lässt sich vermuten, dass zumindest für numerische

Items die elektronische Direktabfrage die Ergebnisse verbessert. Seit Durchführung der Pilotstudie sind die Kodierrichtlinien verbessert und standardisiert worden. Die Autoren empfehlen die Durchführung einer weiteren Validierungsstudie.

Courté-Wienecke 2003

Seit den 90er Jahren liegen in den USA Erfahrungen mit Instrumenten und Verfahren der externen Qualitätssicherung von Managed Care Organizations (MCO) vor. Führende Organisation auf diesem Gebiet ist das National Committee for Quality Assurance (NCQA), das mittels Bewertung von MCOs und ihrer Versorgungsqualität sowie der Berichterstattung über die Ergebnisse eine Verbesserung der Versorgungsqualität erreichen möchte. Die beiden komplementären Vorgehensweisen sind die Zertifizierung von MCOs sowie die Abbildung der Versorgungsqualität anhand eines standardisierten Katalogs von Struktur-, Prozess- und Ergebnisindikatoren (HEDIS). Die Anwendung dieser Information durch verschiedene Nutzergruppen und sich abzeichnende Auswirkungen der externen Qualitätssicherungsverfahren werden dargestellt. Weiterhin wird eine zukünftige Anwendung in Deutschland diskutiert, die bei einer Differenzierung der Krankenkassen und der Etablierung von Netzen aus Leistungserbringern sinnvoll werden kann.

De Lusignan et al. 2002

Die MediPlus-Datenbank enthält anonymisierte Daten von fast zwei Millionen Patienten und mehr als 53 Millionen Verschreibungen. Die Daten werden von mehr als 500 repräsentativ ausgewählten niedergelassenen Ärzten in Großbritannien elektronisch übermittelt. Zehn Marker für Datenqualität sind in der MediPlus-Datenbank implementiert. Sie dienen als Filter, um Daten unzureichender Qualität von der Datenbank fernzuhalten, und werden für jeden datenliefernden Arzt separat berechnet. Ein Feedback erhalten die Ärzte zum einen durch quartalsweise Rückmeldung der Marker-Werte für ihre versandten Daten und zum anderen durch einen halbjährlich erscheinenden Newsletter, in dem Kodierprob-

leme diskutiert werden. Ein Literaturreview lieferte keinen Aufschluss darüber, welche Marker zur Qualitätsverbesserung bei den niedergelassenen Ärzten zurückgemeldet werden sollten. Ein individuelles Feedback scheint wirkungsvoller zu sein als ein Gruppen-Feedback. Wenn Feedback einen positiven Einfluss auf die Datenqualität hat, müsste sich die Qualität umso mehr verbessern, je länger ein Arzt schon an dem Verfahren teilnimmt. Eine diesbezügliche Untersuchung zeigte bei vier der zehn Marker eine signifikante Verbesserung über die Zeit.

Die Ergebnisse dieser Untersuchung sind gemischt. Feedback führte bei einigen Markern zu einer Qualitätsverbesserung, bei anderen aber nicht. Ein derartiges Feedback ist aus sich heraus noch kein effektives Hilfsmittel zur Qualitätsverbesserung, aber es kann mit geringen Kosten zusammen mit anderen Mitteln eingesetzt werden.

De Lusignan 2005a

Der Autor hat mittels eines Literaturreviews die Frage untersucht, wie ein Feedback beschaffen sein sollte, um bei der Verbesserung der Datenqualität in der Primärversorgung hilfreich zu sein. Er fand folgende vier Punkte:

- Einbindung und Unterstützung lokaler klinischer Innovatoren, die am ehesten bereit sind, Änderungen zu übernehmen
- Modellierung des klinischen Kontextes, in dem Qualitätsverbesserung erforderlich ist
- Entwicklung eines Verständnisses für das Gesundheitssystem, seine Kultur und Management
- Identifikation und Ansprache technischer Fragen im Zusammenhang mit Computernutzung und Kodierung

De Lusignan 2005b

Die medizinische Primärversorgung in Großbritannien ist weitgehend computerunterstützt. Für strukturierte Daten wird seit kurzem einheitlich die SNOMED CT verwendet. Um Sekundäranalysen zu erlauben, müssen die Daten aus der EPA den Grad der Morbidität in der Bevölkerung richtig widerspiegeln. Der Artikel beschreibt eine Methode zur Qualitätsverbesserung durch Feedback im Rah-

men des Primary Care Data Quality Programme (PCDQ).

Für jedes Projekt werden in der Planungsphase Evidenzbasierte Auditkriterien, die benötigten Daten, der Auswertungsplan und die Feedback-Techniken festgelegt und eine entsprechende Projektdokumentation begonnen. Das Feedback erfolgt in Form einer educational intervention mit klinisch relevanten Daten gegenüber den Ärzten und Übersichtsstatistiken für Behörden und Forscher. Für jeden klinischen Bereich werden eigene Interventionen geplant unter Einbeziehung von Meinungsführern und denen, die Veränderungen früh annehmen. Die Ergebnisse der einzelnen Praxen werden miteinander verglichen, die Programme jedoch finden in einer nicht-beurteilenden Lernumgebung statt (halbtägige lokal geführte Datenqualitäts-Workshops).

Um effektiv zu sein, sollte Feedback kurz sein (kleine Datenmengen umfassen), erzieherisch (mit Lernzielen), eine geschützte Lernzeit haben (Datenqualitäts-Workshops), klinisch relevant, EBM-unterstützt, peer led. Weniger effektiv sind unaufgeforderte Briefe und Programme zu bestimmten Erkrankungen (z. B. COPD wegen Problemen bei der Diagnose der Bronchitis oder Osteoporose). Datenqualität ist definiert als fitness for purpose. In diesem Fall geht es um Daten, die das Disease Management bei chronischen Erkrankungen verbessern.

Der Prozess hat nach der Planungsphase sieben weitere Schritte: Prüfung der wichtigen Aspekte der Dateneingabe, Untersuchung der Datenextraktionsmethode, Migration der Daten in den Speicherbereich, Integration mit anderen Datenquellen, Datenbereinigung, Datenverarbeitung, Datenanalyse. Die Prüfung der Dateneingabe ist wichtig, da es viele Faktoren mit Einfluss auf die Dokumentation in einer EPA gibt (z. B. unterschiedlich vorkodierte Listen je nach EDV-System, so dass ein Bias entstehen kann). Mit Hilfe von Metadaten wird der Migrations- und Integrationsprozess überwacht. Bei der Datenbereinigung werden Werte außerhalb der Grenzbereiche gelöscht und Laborwerte ggf. in Standardeinheiten konvertiert. Es werden Häufigkeitstabellen erzeugt und Daten in sinnvolle Kategorien gruppiert. Vor den Auswertungen wird sicher-

gestellt, dass ein korrekter Bevölkerungs-Nenner verwendet wird (bekanntermaßen werden in Allgemeinarzt-Daten Prävalenzen eher überschätzt). Zur Prävalenzschätzung werden die Alters- und Geschlechtsprofile der Einzelpraxis mit dem englischen Durchschnitt verglichen. In den Auditkriterien und Auswertungsplänen sind die weiteren Analysen bzgl. Risikofaktoren, Komorbiditäten etc. festgelegt. Die Vergleichsdaten werden den Endnutzern zurückgemeldet, und sie selbst geben Rückmeldung bzgl. der Verständlichkeit, dem Impact und der Wahrscheinlichkeit für Veränderungen.

Der Artikel stellt ein Beispiel mit einer Analyse von 2,4 Mio. Patienten zum Thema KHK vor. In diesen Analysen zeigte sich, dass eine große Anzahl von Patienten mit Kammerflimmern keine optimale Antikoagulationstherapie erhält.

Um mehr Nutzen aus den Routinedaten zu ziehen, sollte die Dokumentation weiter standardisiert werden. Dies kann erreicht werden durch Datenqualitätsprogramme, Reduktion der Anzahl eingesetzter EDV-Systeme, mehr Standardisierung im Disease Management. Weitere Forschung ist notwendig.

Giuffrida et al. 1999

Der NHS (National Health Service) misst zunehmend die Versorgungsqualität. Hierzu werden u. a. die Einweisungsraten behandelter Patienten mit Asthma, Epilepsie und Diabetes als Maß für den Zugang zur Primärversorgung herangezogen. Die Interpretation dieser Raten hängt jedoch auch von anderen Faktoren wie z. B. der Qualität von Präventionsprogrammen sowie sozioökonomischen Faktoren ab. Rohe Einweisungsraten können bei fehlenden Prävalenz- und Inzidenzraten als Maß für die Größe eines Gesundheitsproblems herangezogen werden.

Für die Jahre 1989 bis 1995 wurden die Krankenhaus-Einweisungsraten für die drei o. g. Krankheiten in 90 Health Authority Areas herangezogen. Per multipler Regression wurde geprüft, wie viel von der Variation in alters- und geschlechtsstandardisierten Einweisungsraten durch die sozioökonomischen Charakteristika der Gegend, Bevölkerungsgesundheit und Möglichkeiten der Sekundärversorgung

erklärt werden kann. Für jede Krankheit wurden mit Hilfe der Regressionskoeffizienten vorhergesagte Einweisungsraten berechnet: (a) Gesundheitsvariablen b) Gesundheits- und sozioökonomische Variablen c) alle Variablen.

Es zeigten sich große Schwankungen zwischen den einzelnen Gegenden, von denen ein großer Teil durch sozioökonomische und Sekundärversorgungs-Faktoren erklärt werden kann. Wenn man diese Faktoren in ein Ranking der Gegenden mit einbezieht, gibt es starke Veränderungen im Ranking.

Die Ergebnisse zeigen den großen Unterschied zwischen Performance Indicators und Health Outcomes. Ein Qualitätsindikator sollte sich nur auf die Faktoren beziehen, die von den betroffenen Personen/Institutionen auch beeinflusst werden können. Es wird vermutet, dass eine gute Primärversorgung Krankenhaus-Einweisungen vermeiden kann, jedoch sollten die Einweisungsraten dabei auch für sozioökonomische Faktoren adjustiert werden. Es ist wichtig, mögliche Confounder außerhalb des Einflussbereichs der Versorgung zu identifizieren, bevor die Versorgung bewertet wird.

Goertzen/Stausberg 2004

Eine gute Datenqualität ist essentiell für die Nutzbarkeit medizinischer Dokumentation. Es ist naheliegend, mit der Qualitätssicherung so früh wie möglich zu beginnen. Im Falle rechnergestützter Dokumentationssysteme bedeutet dies, durch die Implementierung von Plausibilitätsregeln und Integritätsbedingungen den Anwender bei der Erfassung valider Daten zu unterstützen bzw. die Erfassung inkorrektur Daten zu unterbinden.

Gross/Fessler 2001

Aus der Industrie wurde das Konzept der kontinuierlichen Qualitätsverbesserung (Total Quality Management) in den Bereich der Medizin übertragen und adaptiert. Die Qualitätskontrolle überprüft nur das Endprodukt (z. B. im Laborbereich) und arbeitet mit „Belohnung“ (z. B. Zertifizierung) und „Bestrafung“ (z. B. Leistungskürzung). Die Qualitätssicherung soll die Mängel der Qualitätskontrolle beseitigen und greift mit Richtlinien oder Leit-

linien in die Prozesse ein. Das kontinuierliche Qualitätsmanagement geht über diese Ansätze hinaus, in dem es ein globales Konzept zur Qualitätsverbesserung vorsieht, das ständig aktualisiert und angepasst wird. Dabei werden zyklisch die folgenden Schritte durchlaufen:

1. Analyse des aktuellen Zustands
2. Festlegung der Ziele
3. Planung der Verbesserung
4. Evaluation der Ergebnisse

Leitlinien als Verhaltensempfehlung auf Basis gesicherter Daten können dieses Vorgehen unterstützen. Weitere unterstützende Strukturen:

- Qualitätszirkel
- Brainstorming
- Fallbesprechung
- Ärztenetze: Bildung verschiedener Qualitätszirkel mit einer größeren Gruppe von Ärzten

Hassey et al. 2001

In dieser Arbeit wird die Frage nach Messbarkeit der Datenvalidität und Brauchbarkeit in einem Praxis-EDV-System untersucht („Records are valid when all those events that constitute a medical event are correctly recorded and all the entries in the record truly signify an event“). Das ausgewählte Zentrum versorgt 13.500 Patienten und ist seit 1994 auf EDV umgestellt, es wurde eine Stichprobe von einer Woche retrospektiv ausgewählt. Die Patientenlisten wurden mit der lokalen Gesundheitsbehörde abgeglichen, die Terminverwaltung gegen die Patientenakten, die Visitendaten gegen Arztbriefe, Berichte u. ä., die Rezepte gegen die Daten der größten Apotheke am Ort, die Impfunterlagen gegen die Datenbank des lokalen NHS und die Diagnosekodes für 15 wichtige Diagnosen mit anderen diagnostischen Kriterien (Medikamente, Laborwerte etc.).

Vollständigkeit und Genauigkeit wurden als Sensitivität bzw. positiver prädiktiver Wert berechnet. Für die kodierten Diagnosen wurde dann eine Likelihood Ratio berechnet. Die TPFN (true positive to false negative) Ratio erlaubt die Schätzung, wie viele Fälle in der Datenbank fehlen, und unter Berücksichtigung der Prävalenz kann man sehen, ob die Gesamtzahl erwarteter Fälle kleiner ist als die

TPFN. Ist dies der Fall, muss man die Datenbank nicht mehr nach weiteren Fällen durchsuchen. Die Maßzahl $DBFind^{10000}$ ist der Anteil der falsch negativen an der Gesamtzahl der Fälle in der Datenbank.

Die Ergebnisse zeigen, dass die Hauptfunktionsbereiche gut abgedeckt sind und dass die Daten in der Praxis-EDV valide, vollständig und akkurat sind mit Ausnahme der Diagnose Adipositas. Die berichteten Krankheitsprävalenzen sind höher als die in der Krankheitsstatistik. Die Autoren empfehlen den Einsatz von Sensitivität, positivem prädiktivem Wert und Likelihood Ratio zur Beurteilung kodierter Diagnosen. Als Teil des Forschungsprojektes wurde ein Validierungs-Toolkit auf Excel-Basis entwickelt.

Haug et al. 1991

Im Rahmen eines Projektes zur Qualitätssicherung der Röntgen-Befundung wurden zwei Feedback-Varianten geplant:

- formale Variante: Qualitätsberichte mit Statistiken und Graphiken, die in wöchentlichen Review-Sitzungen mit allen beteiligten Radiologen diskutiert wurden.
- informelle Variante: Jeder Radiologe kann von seinem Computer aus einen individuellen Bericht über die Qualität seiner eigenen Befundung abrufen und die Werte mit denen seiner Abteilung vergleichen.

Erfahrungen mit dem Einsatz lagen zum Zeitpunkt der Erstellung des Artikels noch nicht vor.

Heller et al 2004

Hier geht es um ein Forschungsprojekt zur Qualitätssicherung in der stationären Versorgung auf Basis von GKV-Routinedaten mit Fokus auf Ermittlung der Ergebnisqualität. Vorteile der Nutzung von Routinedaten sind geringerer Erhebungsaufwand/geringe Kosten, Vollständigkeit der Daten, individueller Bezug und Langzeitbeobachtung, Validität der Daten. Die Qualitätsberichterstattung mit Routinedaten ist in anderen Ländern bereits etabliert. Sterblichkeit ist der am weitesten verbreitete Qualitätsindikator, hängt aber u. a. von der Schweregradverteilung der Patienten im Krankenhaus ab und muss daher adjustiert werden.

Im Projekt werden u. a. folgende Qualitätsindikatoren verwendet: Krankenhaus-Sterblichkeit, Verweildauer bzw. Anteil der Langlieger, Wiederaufnahmen, Re-Operation. Die Darstellung der Ergebnisse erfolgt am Beispiel der Krankenhaus-Sterblichkeit nach Colon-Carcinom-Operation.

Hogan/Wagner 1997

Das Paper ist ein Review der Literatur zur Datenqualität in elektronischen Patientenakten (EPA) und soll die Frage nach der Genauigkeit der EPA-Daten, Ursachen für Fehler, Einflussfaktoren für Fehler, Verbesserungsmöglichkeiten und Vergleich mit papierbasierter Akte beantworten.

Genauigkeit von Daten aus elektronischen Patientenakten ist wegen ihrer vielseitigen Verwendbarkeit von großer Bedeutung. Die zur Prüfung der Datengenauigkeit in Papierakten, Registern und klinischen Studien entwickelte Methodik kann auf die Datenanalyse für EPA angewendet werden. Als Schritte zur EPA ergeben sich Diagnosen, Beurteilungen des Kliniklers, Papierakte Dateneingabe, Diktate Schreibkraft – auf allen Stufen können Fehler auftreten. Zur Messung der Datenqualität können die

- Anzahl der korrekten Beobachtungen (correctness, Gegenmenge falsch positive) und die
- der insgesamt aufgezeichneten Beobachtungen (completeness, Gegenmenge falsch negative)

eingesetzt werden. Goldstandard ist der tatsächliche Zustand des Patienten.

Es wurden 26 Studien ausgewählt und dafür methodologische und beschreibende Scores berechnet. Im Allgemeinen wurden die EPA adäquat beschrieben, am häufigsten mit Anwendungsbereich und Inhalt, ggf. auch Namen, Hardware, Software. Methodologisch hatten viele Studien Fehler. Nur eine Studie berichtet Gründe für die Ungenauigkeit, dabei war der Patient die häufigste Fehlerquelle (durch ungenaue Angaben oder nicht abgesprochene Medikationsänderungen). Die Datenübertragung erfolgte teils von Papier, teils direkte Datenerfassung, teils durch Transkription von Diktaten, einmal automatisch von Patientenüberwachungs-Systemen. Positive

Resultate zur direkten Eingabe durch den Arzt waren wegen Powermangel oder Confounding nicht nachweisbar. Strukturierte Masken ebenso wie Monitoring und Feedback verbessern die Vollständigkeit der Diagnoseeingabe, Anzeige vorheriger Eingaben verbessert die Diagnosenkodierung. Ein Vergleich mit der Papierakte war nicht möglich, weil kein gemeinsamer Goldstandard eingesetzt wurde.

Die Qualität und Quantität gefundener Literatur sind eher unbefriedigend und die Ergebnisse vermutlich nicht repräsentativ. Gründe dafür können das geringe Vorkommen von EPA im akademischen Umfeld und mangelndes Problembewusstsein in der Medizinischen Informatik sein. Die meisten Correctness- und Completeness-Raten liegen über 80 %. Als gut bekannte EPA wurden bisher nicht untersucht. Zu wenige Fehlergründe wurden identifiziert. Es zeigt sich jedoch, dass Eingabefehler eine geringere Rolle spielen als vermutet, andere Faktoren wie der Umfang eines EPA-Systems hingegen größeren Einfluss haben. Zur Beeinflussung der Datenqualität werden CQI-Techniken und Methoden aus Studiendatenbanken empfohlen. Eine Direkteingabe durch den Arzt kann nicht allein aus Gründen der Datengenauigkeit gefordert werden. Theoretisch sollten EPA genauer sein (weil z. B. Plausibilitätsprüfungen eingebaut werden können und Erfassung standardisiert wird), ein direkter Vergleich ist jedoch nicht möglich.

Die Autoren empfehlen eine Standardisierung der Methoden bei Studien zur Datengenauigkeit: beide Maßzahlen sollten berichtet werden, eine unbiased sampling technique eingesetzt werden, Verwendung eines Goldstandards, der so nah wie möglich am wahren Zustand des Patienten liegt, Verblindung derer, die die Analyse durchführen, Beschreibung der EPA mit ihren Charakteristika.

Jick et al. 1991

In 58 Praxen, die von der Firma VAMP Health zur Verfügung gestellte Computer verwenden, wurden Diagnosen in der EDV mit Diagnosen in Arztbriefen an 291 Patienten verglichen, die Nichtsteroidale Antirheumatika (NSAR) erhielten. VAMP stellte 1.000 Praxen 1988 bis 1990 PCs zur Verfügung unter der Bedingung, dass

diese Daten im standardisierten Format erfassen (Diagnosen und Arzneimittel codiert), sie anonymisiert übermitteln und für Forschungszwecke zur Verfügung stellen. Im ersten Jahr werden die Praxen geschult und danach die Daten einmal überprüft. Gilt die Praxis als ‚up to standard‘, werden die Daten für die Verwendung in der Forschung zugelassen.

Für die untersuchten Patienten wurden die Arztbriefe eines Krankenhauses verwendet und überprüft, ob die dort angeführten Diagnosen im Praxissystem erfasst waren. Für neu aufgetretene Diagnosen wurde eingeteilt, ob die Krankheit eine Indikation zur Gabe von NSAR darstellt und ob sie möglicherweise medikamenteninduziert ist oder nicht. Bei up to standard-Praxen wurde nur ein Arztbrief untersucht.

Die Indikation für die erstmalige Verschreibung von NSAR war in 96 % der Fälle in der EDV erfasst. Von 2.491 Patienten lag für 1.191 mindestens ein Arztbrief vor. Überweisungsfälle und möglicherweise medikamenteninduzierte Krankheiten wurden gesondert untersucht. Bei 104 ambulanten Überweisungen mit potentiell medikamenteninduzierten Diagnosen lagen die Diagnosen in 87 % der Fälle vor.

Die Autoren schließen, dass Medikamentengaben in über 90 % der Fälle in der EDV dokumentiert waren und sich daher bei epidemiologischen Studien zur Medikamentensicherheit nur minimale Missklassifikationen ergäben. Allerdings hängt dies auch von der Art des Medikaments ab, da man nicht weiß, ob der Patient die verschriebenen Medikamente auch wirklich einnimmt. Bei den besonders interessierenden Erkrankungen war die Übereinstimmung der Daten hoch. Die VAMP-Daten werden daher für adäquat befunden.

Knap et al. 2004

Die Autoren beschreiben einen umfassenden Ansatz zur Verbesserung der Datenqualität und Verfügbarmachung klinischer Information. Mit einem formalen Modell werden Komponenten von Dokumentationsumgebungen und ihre Beziehungen zueinander beschrieben. Darauf aufbauend erlaubt ein Vorgehensmodell die systematische Planung kooperativer Dokumentationsumgebungen. Ziel ist es,

alle Daten nur ein einziges Mal (verteilt) zu erfassen und multipel zu verwenden.

Kuperman et al. 1998

Um die Genauigkeit von EPA zu verbessern, wurde eine Anwendung entwickelt, mit der Patienten einen Ausdruck ihrer Gesundheits-, Medikations- und Allergiedaten erhalten und auf dem Ausdruck korrigieren können. Da häufig Patientendaten von mehreren Kliniken, Praxen etc. erhoben werden, können die Daten einer Klinik den wahren Zustand des Patienten nicht immer abbilden. Für pap smear, Mammographie, Cholesterin und Pneumonie-Impfung wurde die Genauigkeit der Daten im Kliniksystem überprüft, indem die Ausdrücke in der Nacht vor der geplanten nächsten Visite erstellt und von den Patienten im Warteraum durchgesehen wurden. 80 Formulare wurden überprüft und an 17 Ärzte wurden Fragebogen bzgl. des Fragebogen-Systems verschickt. Die Übereinstimmungsraten lagen zwischen 44 % und 89 % wobei die Cholesterinwerte die besten Übereinstimmungen hatten. Es gab logistische Probleme mit den Formularbögen vor allem bei spanisch sprechenden und sehr kranken Patienten sowie bei Überfüllung. Daher kann die Prüfung nur unregelmäßig durchgeführt werden. Die Befragung der Ärzte ergab nicht genügend Daten. Inzwischen werden die Formulare auch zur Erfassung des Raucherstatus verwendet.

Leiner/Haux 1996

Dokumentation im Klinik-Umfeld unterstützt Behandlungsprozesse, erfüllt gesetzliche Anforderungen, erlaubt Beurteilung der Behandlung, unterstützt klinische Forschung, die Verwaltung der Institution und die Qualitätssicherung in der Behandlung.

In existierenden Systemen ergeben sich Probleme: Diagnosen werden mal als Freitext, mal kodiert erfasst, Computer sind nicht mobil zugänglich, Mediziner und Pflegekräfte werden zu Dateneingebnern für die Verwaltung, haben aber selbst kaum Nutzen von den Daten, die Daten sind nicht gut auswertbar, es werden statistische Fehler begangen bei der Auswertung.

Bei der Planung eines Dokumentationssystems müssen berücksichtigt werden: welche

Items sind notwendig, was kann vernachlässigt werden? Wird ein kontrolliertes Vokabular eingesetzt oder nicht? Wenn ja, wie sieht es aus? Wird immer dokumentiert oder nur in bestimmten Fällen? Wie können existierende Systeme integriert werden? Wie können Arbeitslast und Nutzen balanciert werden und welche Auswertungen sind wann wo sinnvoll?

Die Mehrfach-Verwendbarkeit der Daten sollte angestrebt werden, muss jedoch entsprechend definiert und organisiert sein. Dokumentationsaufgaben sollten klinikweit koordiniert werden.

Ein Dokumentationsprotokoll leitet Designanforderungen von den Systemzielen ab. Es impliziert vier Planungsschritte: Motivation und Zielformulierung, Aufgabenbeschreibung, Ableitung eines integrierten Systemdesigns, Beschreibung der Entwicklungs-, Einführungs- und Nutzungsbedingungen. In der Einführung werden Institution und Motivation beschrieben, dann werden Probleme und Ziele definiert. Hieran sollten alle Interessengruppen beteiligt werden und zu einem Konsens gelangen. Das existierende System wird anhand der Ziele überprüft. Die Anforderungen und Aufgaben für das neue System werden definiert und jeder Aufgabe eine Evaluierungsstichprobe zugewiesen (prolektiv oder retrolektiv). Es werden standardisierte Untersuchungs- und Dokumentationsmethoden eingeführt und motiviert (z. B. mit der Verwendbarkeit der Daten für Analysen und Präsentationen). Die Spezifikationen werden in ein Gesamtkonzept überführt (Datenspeicherung, Merkmalskatalog, Organisation des Dateneingabe-Prozesses (zeitnah, keine Doppelungen, gute Masken, Zeitfenster etc.), Maskenentwicklung, Kommunikation mit anderen EDV-Systemen, Sicherheitskonzept. Darüber hinaus muss das Gesamtprojekt adäquat organisiert sein (Verantwortlichkeit, Zeitschiene, Auswahl geeigneter Hard- und Software), und in der Produktivphase muss es möglich sein, Systemupdates durchzuführen.

Das Dokumentationsprotokoll ist in drei Projekten eines deutschen Universitätsklinikums durchgeführt worden, und es sind eine Reihe von Vor- und Nachteilen sichtbar geworden. Nach Ansicht der Autoren überwiegen die Vor-

teile, es ist jedoch viel Überzeugungsarbeit zu leisten.

Logan et al. 2001

Die größten Probleme bei der EPA ergeben sich in der Aufzeichnung von Visitedaten (Patientengeschichte, Untersuchungsergebnisse etc.). Die Qualität medizinischer Records muss definierbar und messbar sein. Das Paper entwickelt einen messbaren Goldstandard und eine Studienmethodologie, die den Vergleich von Aufzeichnungsmethoden ermöglicht.

Die Autoren empfehlen die Verwendung von Maßzahlen für Qualitätsstandards

- completeness (Vollständigkeit) und
- correctness (Richtigkeit).

Die Qualität einer Patientenakte muss jedoch auch nach ihrem Verwendungszweck beurteilt werden (wie granular muss sie sein etc.) Für die Aufzeichnung von Visitedaten werden Elemente definiert als

- korrekt (vorhanden im Goldstandard, vorhanden und korrekt in der Aufzeichnung – n1)
- inkorrekt (vorhanden im Goldstandard, vorhanden aber nicht korrekt in der Aufzeichnung – n2)
- fehlend (vorhanden im Goldstandard, nicht vorhanden in der Aufzeichnung – n3)
- zusätzlich (nicht vorhanden im Goldstandard, aber inkorrekt in der Aufzeichnung – n4).
- Vollständigkeit lässt sich dann berechnen als $(n1 + n2)/(n1 + n2 + n3)$
- Richtigkeit als $n1/(n1 + n2 + n4)$.

Jedes Einzelelement wird in zwei Dimensionen beurteilt. Um den Goldstandard zu bestimmen, müssen die individuellen, zählbaren Einzelitems in einer Visite bestimmt werden. Hierzu gibt es verschiedene Ansätze mit unterschiedlicher Granularität. Die Autoren empfehlen eine empirische Bestimmung durch Experten, die sich Videoaufzeichnungen von Visiten ansehen, und einen hohen Detaillierungsgrad. Durch mehrere Reviewer, die sich die Videos ansehen und dann Daten aufzeichnen, soll die Variabilität vermindert werden. Zur Beurteilung der Datenqualität soll dies besser geeignet sein als „Standardpatienten“.

In einer Pilotstudie wurden transkribierte Diktate mit einem computerbasierten CPR-System verglichen, das sowohl Freitexteingaben wie auch vorstrukturierte Formulare anbietet. Von vier Videos wurden zwei zu Schulungen und zwei für die Auswertung verwendet. Die Elemente, die in die Akte sollten, wurden aus den Videos extrahiert und dann wurden die Visiten entweder per Diktat oder mit dem Laptop erfasst. Dabei zeigten sich zwischen den beiden Methoden keine statistisch signifikanten Differenzen in Vollständigkeit und Korrektheit (allerdings war die Fallzahl sehr klein!)

Die beschriebene Methode ist für klar abgegrenzte Probleme in der Visite am ehesten geeignet. Die Experten sollten auf etwa demselben Wissensstand sein. Es wird nur die Aufzeichnung der Daten beurteilt, nicht der Inhalt der Akte oder die Qualität der Versorgung.

Lorenzoni et al. 1999

In einem italienischen Krankenhaus in Rom wurde die Frage untersucht, welchen Einfluss ein Trainingsprogramm auf die Qualität von Entlassungsunterlagen hat, die aus der Patientenakte extrahiert werden müssen. Im September 1994 wurde in diesem Krankenhaus ein dreistufiges Programm eingeführt, um den Prozess der Datenextraktion zu verbessern. Ein Trainingshandbuch wurde entworfen und zirkulierte unter den Mitarbeitern. Zusätzlich wurde ein Managementsystem für die Verwaltungsdaten implementiert, welches ein systematisches Monitoring der Datenqualität erlaubt, und monatliche Berichte wurden an die Mitarbeiter verschickt.

Für die Untersuchung wurden Daten aus dem dritten Quartal 1994 (1.442 Entlassungen), dem ersten Quartal 1995 (2.302 Entlassungen) und dem Zeitraum September bis Oktober 1997 (3.160 Entlassungen) untersucht. Aus jeder Zeitperiode wurde eine zufällig gezogene Stichprobe aus den Entlassungen (267, 295 bzw. 310 Fälle) daraufhin untersucht, inwieweit die Angaben in den Entlassungsunterlagen mit den Daten in der Patientenakte übereinstimmen. Für sechs Kategorien wurden die Fehlerraten pro Zeitperiode bestimmt:

- Falsche Auswahl der Hauptdiagnose: 8,5 %, 5,1 % bzw. 1,3 %

- Geringe Spezifität der Hauptdiagnose: 15,8 %, 14,3 % bzw. 1,6 %
- Unvollständige Angabe der Nebendiagnosen: 31,8 %, 30,8 % bzw. 27,2 %
- Falsche Auswahl der Hauptprozedur: 0,7 %, 4,7 % bzw. 0 %
- Geringe Spezifität der Hauptprozedur: 4,1 %, 0 % bzw. 0 %
- Unvollständige Angabe der anderen Prozeduren: 22 %, 3,7 % bzw. 2,6 %

Bei allen Fehlerkategorien zeigt sich nach drei Jahren eine deutliche Reduktion der Fehlerrate. Diese Reduktion ist in allen Kategorien außer bei der unvollständigen Angabe der Nebendiagnosen statistisch signifikant. Training und Feedback haben sich hier als wirksame Methoden zur Qualitätsverbesserung erwiesen.

Macaulay et al. 1996

Die vaskular-chirurgische Abteilung des Royal Infirmary Krankenhauses in Aberdeen bietet einen zentralisierten vaskulären und renalen Transplantationsdienst für 600.000 Einwohner Nord-Ost-Schottlands. Seit 1991 werden die Entlassungsbriefe der Patienten PC-gestützt unter Nutzung eines Thesaurus erstellt. Die Daten werden auf einem standardisierten Erfassungsbogen eingetragen und im Sekretariat am PC erfasst. Zwischen Januar und September 1994 wurde ein prospektiver Audit bei 637 Entlassungsbriefen durchgeführt. In 63 Briefen (10 %) wurden Fehler gefunden. Diese umfassten 11 (12 %) allgemeine Fehler, 21 (22 %) Fehler bei der Angabe der Diagnosen, 19 (20 %) Fehler bei der Angabe der Operation, 19 (20 %) Fehler in Freitextergänzungen, 14 (15 %) fehlende Komplikationen und 10 (11 %) Fehler im Zusammenhang mit der Einbestellung der Patienten zur Nachuntersuchung.

McDonald 1999

Das Nationale Qualitätssicherungskomitee (NCQA) hat für das Gesundheitssystem Qualitätsmaße ausgewählt, die sich auf wissenschaftlich fundierte Versorgungsprozesse beziehen und auf Daten basieren, die in existierenden EDV-Systemen bereits vorhanden sind. Primäre Datenerhebung und Durchsicht von Papierakten sollen vermieden werden. Als

Maßzahlen für Effektivität werden die Parameter empfohlen, die ohnehin von Ärzten bereits erfasst werden. NCQA empfiehlt die Verwendung von HL7 und darüber hinaus bessere Patienten-Identifikatoren (eindeutig), die Verwendung standardisierter Codesysteme für die Erfassung kodierter klinischer Daten (unter Berücksichtigung des Versionenproblems, z. B. bei der ICD) sowie standardisierte Prozeduren und technische Unterstützung bei Datenschutz und Datensicherheit. Der Autor rät von möglichst detaillierter ICD-9-Kodierung ab.

Mikkelsen/Aasly 2001

Hier geht es um einen retrospektiven Vergleich von Patientendaten in EPA und papierbasierten Akten. Da die Einführung von EPA sehr aufwendig ist, kann ein allmählicher Übergang von Papier zur EPA sinnvoll sein. Dafür müssen aber beide Akten parallel geführt und gepflegt werden. Hier steigt das Risiko der Inkonsistenz zwischen beiden. Die Autoren prüfen den Grad der Übereinstimmung zwischen parallel geführten EPA und Papierakten. Hierzu wurden Daten einer neurologischen Abteilung einer norwegischen Universitätsklinik verglichen, wo seit 1997 ein EDV-System für Aktennotizen existiert. Aufgrund systeminterner und organisatorischer Probleme wird dieses System nicht ständig genutzt.

152 Patienten aus der Zeit von November 1997 bis April 1999 wurden zufällig ausgewählt, für 90 von ihnen müssten beide Akten vorliegen. Alle verfügbaren Dokumente (auch aus drei anderen Abteilungen zu Vergleichszwecken) wurden herangezogen. War beides da, wurde überprüft, ob die Daten gleich waren oder unabhängig voneinander geändert worden waren. Änderungen wurden als Korrektur (von Schreibfehlern), Hinzufügung, Veränderung (des Sinns), Löschung oder unspezifiziert klassifiziert und die Information als klinisch, administrativ oder andere. Um mögliche Fehlerfolgen abzuschätzen, wurde die Bedeutung klassifiziert als möglicherweise von Bedeutung, vielleicht von Bedeutung und Bedeutung nicht bestimmbar.

Es wurden für die eingeschlossenen 90 insgesamt 299 Dokumente gefunden, davon waren

260 in der EDV und 295 papierbasiert. Eine Akte wurde ausgeschlossen. 256 fanden sich in beiden Systemen (13 % fehlend in der EDV, am häufigsten Entlassungsnotizen – schließt man diese aus, fehlen 4 %). 9,1 % der Papierdokumente enthielten Änderungen, von denen nur zwei der 37 Änderungen in die EDV übertragen worden waren. 6,6 % der 256 Akten in beiden Systemen waren inkonsistent. In den Vergleichsabteilungen waren mehr Daten in der EDV, der Anteil inkonsistenter Dokumente war aber vergleichbar.

Der beobachtete Inkonsistenzgrad liegt insgesamt bei etwas über 5 %, das meiste davon ist auf fehlende Dokumente zurückzuführen. Falls die Anzahl fehlender Papierakten (1 %) höher ist als vor Einführung der EDV, könnte dies darauf hinweisen, dass nach EDV-Einführung die Papierakten nicht mehr so gut gepflegt werden.

Da die Bedeutung fehlender und falscher Daten vom Verwendungszweck abhängt, können keine Rückschlüsse auf die Reliabilität gezogen werden. Die Datenqualität sollte regelmäßig geprüft werden.

Moczygemba/Hewitt 2001

Durch die elektronische Patientenakte sind verschiedene Initiativen entstanden, die outcome data verwenden, um Verbesserungen zu überwachen. Datenqualitätsmanagement umfasst Data application, warehousing, analysis, collection. Die Datenqualität hat zehn Eigenschaften: accessibility, consistency, currency, granularity, precision, accuracy, comprehensiveness, definition, relevancy, timeliness. Es sollte ein IT-Verantwortlicher eingestellt werden, der die vier Domänen ansprechen kann: Grund der Datensammlung (application), Speicherung (warehousing), Datenverarbeitung (analysis) und wer die Daten erhebt (collection). Endnutzer sollten stärker einbezogen werden, besonders die, die die Daten direkt am Patienten erheben. Dabei entsteht das Problem, dass die Arbeitslast für die Ärzte durch die Notwendigkeit der DV-Einarbeitung steigt. Nach ca. sechs Monaten fällt dies aber wieder ab und die Vorteile (z. B. Entscheidungsunterstützung) werden auch für den Anwender sichtbar. Es muss festgelegt werden, welche Daten erhoben werden, wie sie zu

standardisieren sind (z. B. bzgl. der Terminologie) und wie die Qualität verbessert werden kann (outcomes überwachen, hatte Patient wirklich einen Nutzen etc.). Kommunikation (z. B. zur Vermeidung von Wechselwirkungen) ist von großer Bedeutung. Klinische Informationssysteme können Datensammlung und Retrieval maßgeblich beschleunigen.

Pflüglmayer 2001

Das Gesundheitssystem in Österreich befindet sich im Wandel. Die Umsetzung eines Qualitätsmanagements und die Existenz einer angemessenen informations- und kommunikationstechnischen Infrastruktur sind wichtige Voraussetzungen, damit das Gesundheitssystem auch in Zukunft seinen Anforderungen gerecht werden kann. Der Autor beschreibt in seiner Dissertation zum einen Aspekte des Qualitätsmanagements im Krankenhaus unter Berücksichtigung der derzeit (2001) geltenden Rahmenbedingungen in Österreich. Daneben erläutert er die qualitätsverbessernden Potentiale, die Informations- und Kommunikationstechnologien bieten können.

Neue Konzepte integrativer Versorgung sowie die Zahl schwierig zu behandelnder Krankheiten und chronischer Erkrankungen machen eine fach- und institutionenübergreifende Zusammenarbeit notwendig. Die dabei vorhandenen Bedürfnisse nach Datenaustausch und Kommunikation werden durch eine vertikale oder horizontale Vernetzung unterstützt. Der Aufbau von medizinisch relevantem Wissen (z. B. in Krankheitsregistern) kann durch Vernetzung quantitativ und qualitativ verbessert werden.

Die Verwendung computergestützter Dokumentationsmethoden und eine adäquate Strukturierung der erfassten Datenobjekte können dazu beitragen, eine im Vergleich mit konventioneller Dokumentation höhere Qualität bei der Erfassung und Auswertung medizinischer Daten zu erreichen.

Porcheret et al. 2004

In Großbritannien soll im Jahre 2005 die nationale elektronische Patientenakte implementiert werden. Um dieses Ziel zu erreichen, ist eine Qualitätsverbesserung der vorhandenen elektronischen Patientenakten notwendig. Das

dafür vorgesehene Verfahren wurde in einem Netzwerk niedergelassener Ärzte getestet. Die Prävalenz von zwölf ausgewählten Krankheiten wurde aus den erfassten Daten der teilnehmenden Ärzte anhand der vorgenommenen Kodierungen berechnet und mit den entsprechenden Daten verglichen, die in der 4. Nationalen Studie zur Morbiditätsstatistik erhoben wurden. Die Ärzte wurden regelmäßig über den Stand ihrer Datenqualität unterrichtet und umfassend in allen Aspekten der Datenerfassung und Kodierung geschult. Das Verfahren erwies sich als effektiv zur Verbesserung der Datenqualität. Alle teilnehmenden Ärzte fanden Zeit für die Schulungssitzungen und hielten oder verbesserten ihren persönlichen Qualitätsgrad.

Pringle et al. 1995

Aus bisherigen Studien von Daten in Praxis-systemen lässt sich schließen, dass akute Krankheiten bis zu 3/4 unterdokumentiert sind, dass jedoch für einige chronische Erkrankungen hochwertige Daten vorliegen. Lebensstil-Daten und sozioökonomische Daten sind in Praxissystemen kaum bis gar nicht dokumentiert. Die vorliegende Studie untersucht die Vollständigkeit und Genauigkeit von Daten über 18 Monate aus vier Praxen, die dasselbe Praxis-EDV-System verwenden. Es wurde die Variabilität der Diagnosen für zehn Diagnosen geprüft (z. B. Diabetes), die Vollständigkeit von Lebensstil und sozioökonomischen Daten sowie die Vollständigkeit der Aufzeichnungen zu Visiten (pro Arzt wurden 50 Visiten aufgezeichnet und die Items verglichen)

Für die Diagnosen ergab sich: je subjektiver die Diagnose, desto inkonsistenter wurde sie aufgezeichnet. Von den Lebensstilfaktoren war Rauchen mit 52,1 % am besten dokumentiert. Die Papierakten enthielten mehr Visiten-daten als das EDV-System (mit Ausnahme der Diagnosen, Rezepte und Überweisungen). Die vier Praxen haben ausreichend vollständig und genau dokumentiert, allerdings sind die Papierakten immer noch ausführlicher.

Prins et al. 2000

Performance indicators sind systematisch entwickelte Messgrößen zur Bewertung von Gesundheitsentscheidungen, -dienstleistungen

und Outcomes. Um diese Indikatoren zu quantifizieren, sind zuverlässige Daten zu Patienten, Versorgung und Outcome notwendig. Die Autoren prüfen die Verfügbarkeit (availability) und Genauigkeit (accuracy) von Daten in rechnerunterstützten KIS. Fragestellungen: welche Indikatoren sollen zur Beurteilung ausgewählt werden, sind die benötigten Daten im KIS vorhanden und ist die Datenqualität ausreichend?

An einer pädiatrischen Abteilung wurde eine Fallstudie durchgeführt, der Versorgungsprozess wurde modelliert. Nach Validierung des Ablaufdiagramms identifizierten Pädiater performance indicators. Diese wurden bewertet nach der RUMBA-Regel (relevance, understandability, measurability, to be influenced by behaviour, acceptability). Die Indikatoren wurden auf die Daten des vorhergegangenen Kalenderjahres von Patienten mit vermuteter Meningitis als Einweisungs- bzw. Meningitis als Entlassungsdiagnose angewendet, es wurden Daten aus dem KIS und papierbasierte Akten herangezogen. Die Accuracy der KIS-Daten wurde anhand der Papierakte bewertet und dort, wo es keine Papierakten gab, nach der Erhebung der Daten an der Datenquelle.

- Accuracy = welcher Anteil der erfassten Daten ist wahr
 - gute Accuracy = > 95 %
 - reasonable = zwischen 80 und 95 %
 - schlecht = < 80 %

14 Indikatoren wurden ausgewählt, davon beziehen sich neun auf den Prozess und vier auf das Outcome. Die Indikatoren erfassen das Zusammenpassen von Intervention und Indikation, die Zeitnähe und das Outcome. Im Allgemeinen war die Accuracy für demographische Daten, Testergebnisse und Ereignisdaten gut. Die Qualität der Daten zum Einweisungsgrund und zur Diagnose war schlecht. Zeit und Datum der Einweisung waren nur papierbasiert dokumentiert. Das Datum der Anamnese und Untersuchung war nicht dokumentiert, und Daten von anderen Einrichtungen bei Überweisungen waren nicht erhältlich. Von 14 Indikatoren ließen sich fünf nicht zuverlässig quantifizieren, weil Daten nicht spezifisch genug oder nicht in standardisierter Terminologie dokumentiert waren.

Krankenhausbasierte Register müssen weiter auf- bzw. ausgebaut werden.

Ricketts et al. 1993

Die Autoren beschreiben eine Untersuchung der in der Manchester Orthopaedic Database (siehe Barrie/Marsh 1992) gespeicherten Schlüsselworte auf Vollständigkeit und Richtigkeit. Je 100 stationäre Fälle aus zwei Kreiskrankenhäusern wurden auf die gleiche Art untersucht wie von Barrie/Marsh 1992 berichtet. Die Vollständigkeit und Richtigkeit der Schlüsselworte lag in dem einen Krankenhaus bei 45 % bzw. 57 %, während das zweite Krankenhaus deutlich bessere Werte von 75 % bzw. 88 % erreichte. Dies wurde darauf zurückgeführt, dass das zweite Krankenhaus einen Systemkoordinator benannt hatte, der die korrekte Datenerhebung kontrollierte. In beiden Krankenhäusern erreichte die Beschreibung von Diagnosen und Operationen eine höhere Qualität als die Beschreibung der Komplikationen.

Roos et al. 1989

In dieser Arbeit wird der Ansatz beschrieben, zur Untersuchung der Datenqualität zwei verschiedene Datenquellen, die sich auf dieselben Ereignisse beziehen, miteinander zu verknüpfen. Die Datenquellen waren Daten aus den Entlassungsunterlagen von Krankenhäusern und Zahlungsanforderungen der Ärzte in Manitoba aus dem Zeitraum April 1979 bis März 1984. Die Analyse umfasste elf verschiedene chirurgische Eingriffe an Erwachsenen mit einem Mindestalter von 25 Jahren. Untersucht wurden die Fragen:

- Wie gut können die beiden Datenquellen miteinander verknüpft werden?
- Wie gut ist die Übereinstimmung bei Prozeduren und Diagnosen bei gegebener Verknüpfung?

Die Verknüpfung konnte bei fünf der elf chirurgischen Prozeduren in mehr als 95 % der Fälle hergestellt werden. Bei weiteren drei Prozeduren waren es mehr als 90 %. Die restlichen drei Prozeduren erreichten nur Werte von 77 % und 84 %. Die Übereinstimmung bei den Diagnosen erreichte im Mittel 75 % bei Benutzung der dreistelligen ICD-9-CM-Kodes und variierte erheblich mit der Art der Operation.

Savage 1999

Das Paper schlägt ein Rahmenkonzept vor, mit dem Routinedaten durch Metadaten charakterisiert werden können. Diese Metadaten können zur Analyse der Datenstrukturen und zum Aufdecken von Datenfehlern eingesetzt werden.

Routinedaten aus der Versorgung werden zunehmend für verschiedene Zwecke eingesetzt. Häufig werden Datenbanken aus existierenden heraus generiert, dabei gehen jedoch Strukturinformationen verloren. Dies wird besonders bei institutionsübergreifenden Datenanalysen zum Problem. Durch die Hinzufügung von Metadaten kann dies abgefangen werden. Zum einen sind dies technische Metadaten, die sich auf die Datentransformationen und -transferprozesse beziehen, zum anderen Metadaten, die den Datenerhebungsprozess und die Originalstrukturen und -charakteristika beschreiben. Letztere sind von besonderer Bedeutung zum Aufdecken von Datenfehlern. Der Ansatz des Papers betont nicht die Datenstrukturen, sondern die Prozesse und ihre Beziehungen zu den Daten.

Das Rahmenkonzept umfasst folgende Prozesse: den Prozess der Datengewinnung (Data Acquisition, Daten erheben, generieren, speichern), der Transformation (beim Laden in die zentrale Datenbank) und der Analyse (Prozesse, die die Daten der zentralen Datenbank verwenden). Datenfehler sind ungültige Werte (z. B. Fehleingaben), gültige Werte, die jedoch von der Wirklichkeit abweichen (z. B. Fehlklassifikationen), Inkonsistenzen und fehlende Werte. Bei Daten können einer oder mehrere dieser Fehlertypen vorliegen.

Prozesse generieren Datenelemente, die wieder in andere Prozesse eingehen. Diese Prozesse und Elemente haben bestimmte Attribute (wer, wann, was, warum etc. für Prozesse und strukturelle, Definitions- und Bedingungsattribute für Elemente). Sowohl Prozesse als auch Elemente können fehlerbehaftet sein, es wird zwischen Prozessfehlern und Elementfehlern unterschieden.

Die Autoren empfehlen den Einsatz von Metadaten von vornherein. Die Prozesse in der Datengewinnung und -verarbeitung müssen stärker berücksichtigt werden.

Slaughter/Willner 1996

Kaiser Permanente ist eine health maintenance organisation in Kalifornien, die mehrere Kliniken etc. betreibt. Als 1991 Datenprobleme bzgl. der Erfassung brusterhaltender Operationen beim Mammakarzinom festgestellt wurden, wurden in Zusammenarbeit mit einem Kodierzentrum neue Leitlinien zu Dokumentation und Kodierung entwickelt. Auch bei der Durchführung dreier risikoadjustierter Outcome-Studien zeigten sich Probleme mit der Datenqualität, die auf Unterschiede in Terminologie und Dokumentation zurückführbar waren. Durch ein Pilotprojekt unter Leitung eines regionalen Kodierungs-Koordinators, in dem ein Arzt als Reviewer am Kodierungsprozess teilnahm, ließen sich die Kodierung und damit die finanziellen Einnahmen deutlich verbessern. Daraufhin wurden durchgehend Ärzte als Reviewer in der Kodierung eingesetzt. Auch im Bereich der Geburtshilfe sowie für das Problem der peri- und postoperativen Blutverluste wurden Leitlinien zur Dokumentation entwickelt. Zurzeit wird eine Mischform aus Papierakte und EPA eingesetzt. In allen Bereichen wurde die Zusammenarbeit zwischen Medizinern und Medizininformatikern verstärkt.

Stausberg 1997

Vorgestellt wird eine retrospektive Studie zu Auswirkungen verschiedener Qualitätskontrollmethoden auf die Qualität der gesetzlich vorgeschriebenen medizinischen Dokumentation (kodierte Diagnosen und Prozeduren). Es wurden Daten vom Uniklinikum Essen aus dem Operativen Zentrum II (elektronische Patientendaten) benutzt.

Die untersuchten Methoden umfassten Vorträge, Schulungen, Qualitätszirkel, Abteilungsstatistik und zentrales Feedback.

Untersucht wurden die Benutzerakzeptanz (z. B. Teilnehmerzahl bei Schulungen) und die Qualität der medizinischen Dokumentation. Als Ergebnis zeigt sich, dass Information und Schulungen der Benutzer einen geringeren Einfluss auf die Datenqualität haben als abteilungsinterne und zentrale Kontrollmaßnahmen.

Stausberg/Albrecht 1996

In einem Projekt soll die Qualität der Basisdokumentation im stationären Bereich durch inhaltliche Schulung und Fortbildung sichergestellt werden. Dazu wurden zwei Qualitätszirkel eingerichtet mit für die Basisdokumentation zuständigen Oberärzten als Teilnehmern, die als Multiplikatoren dienen sollen. Die Erwartungen der Teilnehmer sind eine Vermittlung von Wissen (speziell zur Klassifikation und Kodierung) und die Möglichkeit zum Erfahrungsaustausch.

Stausberg/Albrecht 1997

Begleitend zur Einführung der Basisdokumentation nach dem Gesundheitsstrukturgesetz (GSG) wurden verschiedene Verfahren zur Sicherstellung einer hohen Datenqualität untersucht und bewertet. Die Daten stammen aus dem Uniklinikum Essen (Operatives Zentrum II). Untersucht wurden die Vollständigkeit und Zeitnähe der Diagnosendokumentation, die Validität der Operationsdokumentation sowie die Treffsicherheit der Basisdokumentation zur Ermittlung der Pauschalentgelte.

Ergebnisse:

1. Lehr-, Fortbildungs- und Informationsmaßnahmen: Nur das Thema ICD-10 stieß auf Interesse der Ärzte; die anderen Themen (Anforderungen an Dokumentation, Klassifikation) nicht. Das Pilotprojekt „Qualitätszirkel Dokumentation“ wurde wegen erheblicher Informationsdefizite der Teilnehmer vorzeitig beendet. Eine Stichprobe der Daten von einer Woche zeigt, dass nur in 32 % der Fälle die Aufnahme Diagnosen dokumentiert sind.
2. Abteilungsinterne Qualitätslenkung: Die Erweiterung der EDV-Systeme um Funktionen zur Auswertung der Dokumentationsqualität (Mängellisten, abteilungsinternes Controlling) steigerte innerhalb eines Monats den Anteil entlassener Patienten mit dokumentierter Entlassungsdiagnose auf ca. 75 % aller entlassenen Patienten.
3. Abteilungsübergreifende Qualitätslenkung: Eine wöchentliche Überprüfung der Basisdokumentation durch Mitarbeiter des Instituts für Medizinische Informatik, Biometrie und Epidemiologie auf Vollständig-

keit und Validität und zeitnahe Rückmeldung der Ergebnisse wurden eingeführt. Innerhalb einiger Wochen wurde ein Anteil von 90 % initial dokumentierter Diagnosen erreicht. Ähnliche Qualitätsverbesserungen zeigten sich bei der Validität der Operationsdokumentation.

Stausberg/Jöckel 2001

Die medizinische Qualitätssicherung erfolgt in deutschen Krankenhäusern weitgehend unabhängig von den flächendeckenden Informationssystemen zur Routedokumentation. Informationssysteme und Qualitätsmanagementsysteme beinhalten die gleichen Elemente und könnten daher miteinander verzahnt werden. Dies ließe u. a. folgende Vorteile erwarten: eine hohe Motivation und Akzeptanz bei den Mitarbeitern, vollständige und valide Daten, die Möglichkeit der zeitnahen Erfassung und Analyse sowie eine Reduktion des Erfassungsaufwands.

Im Operativen Zentrum II (OZ II) des Universitätsklinikums Essen wurde ein Modell des integrierten Qualitätsmanagements konzipiert und umgesetzt. Im OZ II wird ein Informations- und Kommunikationssystem eingesetzt, welches alle Abteilungen und Mitarbeiter unterstützt. Über die Basisdokumentation hinaus wurde die medizinische und operative Dokumentation an die Anforderungen der einzelnen Abteilungen angepasst. Für das Qualitätsmanagement und die Forschung wurde parallel ein Data Warehouse aufgebaut.

Das PDCA-Modell von Deming (Plan-Do-Check-Act) wurde durch das Konzept eines Qualitätsmonitorings auf der Basis der Routedokumentation ergänzt. Ausgangspunkt ist die Routedokumentation, die einen Basisatz an Behandlungsmerkmalen enthält. Die Qualität der Behandlung wird zeitnah und kontinuierlich an die Fachabteilungen zurückgemeldet. Die Interpretation der Kenngrößen erfolgt durch die Fachabteilungen in Zusammenarbeit mit der Medizinischen Informatik und der Biometrie. Offensichtliche Qualitätsdefizite können direkt behoben und Verbesserungen eingeführt werden. Die Verbesserungen fließen in die Routedokumentation ein. Der Erfolg wird wiederum in den Kennzahlen

sichtbar. Hierdurch wird eine kontinuierliche Qualitätsverbesserung erzielt.

Seit 1997 werden den drei operativen Fachabteilungen des OZ II monatlich Kenngrößen aus dem Data Warehouse zur Verfügung gestellt. Die Kenngrößen umfassen die Struktur-, Prozess- und Ergebnisqualität. Diagnosen und Eingriffe werden mit den gesetzlich vorgeschriebenen Klassifikationen kodiert. Die vergleichende Analyse wichtiger Behandlungsgruppen zeigt, dass das Modell relevante Kenngrößen zur Versorgungsqualität zur Verfügung stellt.

Zur Qualitätssicherung der Dokumentation werden informationstechnische und organisatorische Maßnahmen eingesetzt. Informationsveranstaltungen und Schulungen werden zum OZ II-System, zur Kodierung und zu den Dokumentationsanforderungen abgehalten. Die Integration und multiple Verwendung der Daten spielt eine wichtige Rolle für die Datenqualität. Daneben wird die Dokumentation in den Behandlungsablauf eingebunden. Bei dem Einsatz organisatorischer Verfahren lassen sich drei Ebenen unterscheiden: der erfassende Arzt, der ärztliche Dokumentationsbeauftragte der Abteilung (Oberarzt) und das Institut für Medizinische Informatik, Biometrie und Epidemiologie (IMIBE), welches für eine zentrale Qualitätslenkung zuständig ist und wöchentlich die Vollständigkeit und Validität der Basis- und Operationsdokumentation am OZ II analysiert und Ergebnisse an die Dokumentationsbeauftragten zurückmeldet.

Das eingeführte Konzept hat sich in der Praxis bewährt. Die Integration der Dokumentation für die Qualitätssicherung in die Routinedokumentation hat entscheidende Vorteile gegenüber einer isolierten Lösung. Der Erfassungsaufwand wird reduziert und die erfassten Daten können multipel verwendet werden. Dies steigert die Akzeptanz bei den erfassenden Mitarbeitern und trägt positiv zur Datenqualität bei.

Stausberg et al. 2003b

Es wurden eine Studie und ein Literaturreview zum Vergleich papierbasierter mit elektronischer Patientendokumentation durchgeführt. Datenbasis waren kodierte Diagnosen und Prozeduren von Patienten aus der chirurgi-

schen Abteilung eines nicht-universitären Krankenhauses. Der Vergleich der Daten erfolgte patientenweise nach Nachkodierung der Einträge in der Papierakte.

Die Papierakte zeigt leichte Vorteile bei der Qualität der Diagnosenkodierung und potentielle Nachteile bei der Qualität und Quantität der Prozedurenkodierung.

Aus dem Literaturreview ergeben sich u. a. die Empfehlung, den Patienten selbst als Goldstandard zu benutzen (statt der Papierakte), sowie Vorschläge zu Qualitätskriterien für den Vergleich und für die Studienmethodologie.

Stausberg et al. 2003c

Ziel der vorgestellten Studie ist es, die Qualität der Basisdokumentation in einem Klinischen Arbeitsplatzsystem (KAS) mit einer Nachdokumentation aus der konventionellen Krankenakte zu vergleichen. Die Stichprobe umfasste 244 Behandlungsfälle von September 2001 einer Abteilung für Allgemein-, Unfall- und Gefäßchirurgie. Diagnosen und Prozeduren wurden aus den Krankenakten ermittelt und mit ICD-10-SGB-V 2.0 bzw. OPS-301 2.0 kodiert (Vorgehen 1). Parallel wurden die Daten der Stichprobe aus dem KAS extrahiert (Vorgehen 2). Für beide Datenbestände wurden die Australian Refined-DRGs ermittelt.

Die mittlere Anzahl unterschiedlicher Diagnosen je Fall betrug 3,7 vs. 3,9 (Vorgehen 1 vs. Vorgehen 2) und die mittlere Anzahl Prozeduren je Fall 3,1 vs. 3,9. In 5,3 % bzw. 11,5 % der Fälle wurden nicht näher bezeichnete Erkrankungen als Hauptdiagnose angegeben. In 152 Fällen (62 %) stimmten die ermittelten DRGs bei beiden Vorgehen überein. Der Unterschied im Case Mix Index betrug 1,5 %. Aus den Ergebnissen lässt sich ableiten, dass eine vollständige und umfassende Sicht auf einen Behandlungsfall nur bei Betrachtung beider Dokumentationen zu erlangen ist.

Stausberg et al. 2003d

Klinische Arbeitsplatzsysteme (KAS) können im Rahmen des Qualitätsmanagements vielfältig eingesetzt werden: bei der Messung von Qualität, zur Qualitätsverbesserung und zur Steuerung von Maßnahmen. In dem vorgestellten Projekt geht es darum, Informationsmanagement und Qualitätsmanagement zu

integrieren, um die Prävention und Behandlung des Dekubitus zu optimieren.

Am Universitätsklinikum Essen wurde eine interdisziplinäre Arbeitsgruppe eingerichtet, an der die Pflege, der ärztliche Dienst, das IT-Zentrum und die Medizinische Informatik beteiligt sind. Für operationelle Aufgaben wurde eine bestehende interdisziplinäre Wundsprechstunde einbezogen. Die Dokumentation eines Dekubitus erfolgte im System *medico//s* der Firma Siemens. Die Arbeitsgruppe legte die Erhebungsmerkmale, das Vorgehen bei der Dokumentation, Interventionskriterien und ein Schulungskonzept fest. Der Qualitätssicherung der Dokumentation kam eine wichtige Rolle zu. Die Daten sollten rechtzeitig erfasst, vollständig, widerspruchsfrei und valide sein. Im November 2002 begann eine Testphase mit fünf Stationen, ab März 2003 wurde eine flächendeckende Vollerhebung auf allen 78 Stationen des Universitätsklinikums durchgeführt.

Die Erfassung durch das Pflegepersonal umfasste die Risikoeinstufung und den Status bei Aufnahme, den Status und die eingeleitete Therapie bei Auftreten eines neuen Dekubitus sowie den Status bei Entlassung. Aufgrund der Erfahrungen in der Testphase wurde die Erfassung bei der Entlassung aufgegeben.

Zur Prüfung der Validität wurde ein Stichprobenverfahren implementiert. Unter Annahme von 6 % Aufnahmen mit Dekubitus und 2 % Neuerkrankungen auf der Station wurde ein Stichprobenumfang von ca. 3.000 Fällen berechnet. Aus praktischen Erwägungen wurde die Stichprobe auf eine zufällig ausgewählte Station pro Werktag festgelegt, die von zwei erfahrenen Pflegekräften aufgesucht wurde, um ohne Kenntnis der in *medico//s* gespeicherten Daten den aktuellen Status der Patienten zu erheben.

Die Vollständigkeit der Dokumentation bei Aufnahme stieg von 31 % im November 2002 auf 73 % im Februar 2003. Im Februar 2003 wurde bei 255 Aufnahmen ein Dekubitus dokumentiert, zum gleichen Zeitraum in der Kontrollgruppe bei zwei von 95 untersuchten Patienten.

In diesem Projekt gab es zwei wesentliche methodische Probleme:

- Die Messung von Qualität führt bereits zu

einer Qualitätsverbesserung. Damit lässt sich die Wirkung des Projekts nicht sicher von den Wirkungen prophylaktischer oder therapeutischer Maßnahmen trennen.

- Eine Dokumentation ohne Überprüfung liefert keine nutzbaren Daten. Deshalb hat sich die Projektarbeit auf die Qualitätssicherung der Dokumentation konzentriert.

Die Ergebnisse der Testphase zeigen, dass die Dokumentationsqualität noch nicht ausreichend ist. Deshalb ist die Einführung eines Stufenplans zur Qualitätssicherung der Dokumentation dringend erforderlich. Maßnahmen sollen bei der einzelnen Pflegekraft, der Stationsleitung und der Klinikpflegedienstleitung ansetzen. Eine Rückkopplung der Validierungsergebnisse an die Stationen ist in einem Abstand von zwei Tagen zum Audit geplant.

Thiru et al. 2005

Die Autoren führten einen Literaturreview durch, um Angaben zur Messung der Datenqualität der elektronischen Patientenakte zu finden. Gesucht wurde in allen großen bibliographischen Datenbanken unter den englischsprachigen Publikationen aus den Jahren 1980 bis 2001. Es wurden 37 Quellen gefunden, die direkt über die Messung der Datenqualität berichteten, und weitere 15 Quellen, die die Qualität elektronischer Patientenakten im Vergleich mit einem Referenzstandard beschrieben.

Als Referenzstandard wurden in den meisten Fällen Routinedaten oder nationale Statistiken genutzt. In wenigen Fällen wurden Surveys herangezogen und in Einzelfällen die Patienten direkt untersucht. Die Reliabilität der Daten wurde durch einen Ratenvergleich gemessen. Messungen der Sensitivität hingen sehr von den untersuchten Datenobjekten ab. Der positive prädiktive Wert erreichte stets hohe Werte, die eine gute Validität bedeuten.

VdAK et al. o. J.

Der Medizinische Dienst der Krankenversicherung (MDK) führt im Auftrag der am Pflegegesetzverfahren beteiligten Krankenkassen und gegebenenfalls ihrer (Landes-)Verbände die Fehlbelegungsprüfung bei den Krankenhäusern durch. Der Nachweis des Umfangs

der Fehlbelegungen wird vom MDK durch eine Stichprobenprüfung der Entlassungen aus den letzten zwölf Kalendermonaten vor der Ankündigung der Prüfung geführt. Die Spezifikation der Stichprobenziehung und des für eine ausreichende Prognosesicherheit erforderlichen Stichprobenumfangs wird durch die Spitzenverbände auf Vorschlag des Medizinischen Dienstes der Spitzenverbände der Krankenkassen (MDS) bekannt gegeben.

Aus der Stichprobe sollen die primäre und sekundäre Fehlbelegung in der Grundgesamtheit geschätzt werden. Dazu werden von jedem Patienten die Angabe über die Fehlaufnahme (ja/nein) und die Anzahl fehlbelegter Tage bezogen auf alle Belegungstage des Patienten benötigt. Aus der Grundgesamtheit wird eine einfache, ungeschichtete Zufallsstichprobe gezogen. Für die primäre Fehlbelegung müssen der erwartete Anteil von Fehl-aufnahmen und die relative Breite des Konfidenzintervalls vorgegeben werden. Für die sekundäre Fehlbelegung müssen die erwartete Standardabweichung der Rate (Fehlbelegte Tage/Belegungstage) und die absolute Breite des Konfidenzintervalls vorgegeben werden. Desweiteren wird angenommen, dass 2 % der zur Prüfung vorgesehenen Patientenakten nicht verfügbar sind. Es werden die statistischen Formeln genannt, mit denen sich aus den o. g. Daten die benötigte Stichprobengröße für die Fehlbelegungsprüfung berechnen lässt. Weiterhin werden die Formeln angegeben, mit denen sich aus der Stichprobe die Punktschätzer und 95 % – Konfidenzintervalle für die primäre und sekundäre Fehlbelegung berechnen lassen.

Ward et al. 2004

Das Krankenhaus, an dem der Erstautor beschäftigt ist, setzt auf seiner Intensivstation ein kommerzielles umfassendes rechnerunterstütztes KIS ein, welches fast alle relevanten Aspekte der Patientenversorgung auf der Intensivstation abdeckt. Da die Daten des KIS auch für Forschungszwecke genutzt werden sollten, wurde die Qualität dieser Daten untersucht. Als besonders wichtig wurden die Länge des Aufenthaltes auf der Intensivstation, Vitalzeichen, Tage mit mechanischer Beatmung, Medikationen und Diagnosen aus den

Daten ausgewählt. Sie wurden mit den gleichen Daten verglichen, die parallel in anderen Datenquellen (z. B. Abrechnungsdatenbank, direkte Untersuchung vor Ort) erhoben wurden. Bei Medikationen und Diagnosen zeigten sich signifikante Unterschiede zwischen dem KIS und den anderen Datenquellen. Bei den übrigen untersuchten Variablen ergab sich eine gute Übereinstimmung. Verglichen mit anderen weitverbreiteten Datenquellen, ist ein kommerzielles KIS eine akzeptable Quelle für die Patientendaten einer Intensivstation.

Welzer et al. 2005

Die Datenqualität diagnostischer Prozesse bildet den Ausgangspunkt, um die Bedeutung der syntaktischen und semantischen Komponente von Datenqualität zu untersuchen (häufig sind beide verknüpft). Qualitätsindikatoren: Accuracy, integrity, consistency, timeliness, completeness. Prozessbezogene Probleme entstehen häufig, wenn Daten bei einem falschen Schritt im Geschäftsprozess eingegeben werden oder die Daten nicht richtig verstanden wurden – Bedeutung der Qualität statt reiner Geschwindigkeit bei der Erfassung und der richtigen Systemarchitektur. Besonders in Data Warehouses ist die Datenqualität sehr wichtig. Wenn die Daten nicht vorher entsprechend bereinigt und aufgearbeitet werden, können sie nicht verwendet werden (besser keine Daten als falsche Daten). In medizinischen Umgebungen müssen sowohl der Geschäftsteil des Systems als auch der medizinische Teil berücksichtigt werden, und da die Entscheidungen schwerwiegende Konsequenzen haben, ist eine korrekte Modellierung besonders wichtig (Sicherung der Datenqualität in allen Subsystemen und weitere Datenvalidierungen im Gesamtsystem). Die Autoren empfehlen die Wiederverwendung von semantisch überprüften Datenstrukturen und Objekten, wobei für die Wiederverwendung der neue Kontext geprüft werden muss – die Struktur muss erlauben, multiple Objekte der gleichen Funktionalität aber aus verschiedenen Kontexten abzuspeichern.

Wood/Youatt 1979

Der Artikel beschreibt ein EDV-System, das ein Forschungsprojekt zur Methadonsubstitu-

tion unterstützt. Die Daten werden bei Eingabe plausibilitätsgeprüft, monatliche Zusammenfassungen und die Untersuchung von ausgegebenen Datenmatrizen durch externe Experten dienen der weiteren Fehlerfindung. Die monatliche Patientendatenmatrix ist dabei ein zentrales Tool mit einem einheitlichen Datenformat und der gleichen Zeitbasis für alle Variablen. Sie ermöglicht Vergleich und Korrelation der verschiedenen Variablen. Die Datenstruktur ist auf Kosteneffizienz sowie auf die Vereinfachung der Dateneingabe, Retrieval und Analyse hin ausgelegt.

Durch Koordination und Interaktion von Nicht-EDV-Prozeduren mit dem programmierten Datenmanagementsystem können relativ viele Datenprobleme abgefangen werden. Die Daten werden in der Klinik erhoben und zunächst temporär gespeichert. Bei den automatischen Plausibilitätsprüfungen entdeckte Fehler werden in einer eigenen Datei abgespeichert, und das System generiert eine Fehlermeldung. Nach dem Editieren werden die Daten geprüft und, wenn sie fehlerfrei sind, permanent abgespeichert.

Einige Fehler können so direkt beim Eingeben abgefangen werden. Weitere Fehler werden durch Feedback in Form von Ausdrucken entdeckt. Programme erzeugen einen monatlichen Zusammenfassungsbericht, auf dem mit einem Tag pro Zeile die Daten für jeweils einen Patienten für den jeweiligen Monat aufgelistet sind. So lassen sich Daten vergleichen und Muster erkennen, und das Review der Berichte ist Teil der Feedbackschleife. Die anderen monatlichen Berichte beziehen sich auf die Klinikleistung, Zeitbasis ist hier der Patientenmonat (= Monate seit Einbringung in das Programm), und es werden Kaplan-Meier-Kurven generiert. Weitere Datenretrieval-Programme erzeugen die Datenmatrizen. Bei der Generierung der Matrizen wird für Daten, die vordefinierte Grenzwerte verletzen, eine Fehlermeldung generiert.

2.4.7.5 Datenqualität in Registern

Allgaier et al. 2005

Das Deutsche Register für Stammzelltransplantation (DRST) dokumentiert seit 1998 die wesentlichen Parameter und Ergebnisse der

in Deutschland durchgeführten Transplantationen von Blutstammzellen. Die Daten werden auch an die internationalen Register der European Group for Blood and Marrow Transplantation (EBMT) und das Centre for International Blood and Marrow Transplant Research (CIBMTR) weitergegeben. Die Datenerfassung kann über Papierformulare erfolgen, die an das DRST-Sekretariat zur Datenerfassung geschickt werden, oder über ein von EBMT übernommenes internetbasiertes Remote-Data-Entry-System (ProMISe). Altdaten wurden aus verschiedenen Formaten in die SQL-Datenbank übernommen. Das DRST verfügt über ein Programm, das umfangreiche Plausibilitätsprüfungen durchführt. Das Prüfungsergebnis wird den Zentren zur Bearbeitung zugeschickt. Jeweils am Jahresanfang werden von allen Zentren kumulierte Daten des Vorjahres abgefragt (DRST-Survey).

Die Plausibilitätsverletzungen werden nach Schweregrad in verschiedene Kategorien eingeteilt:

1. vorhandene, aber widersprüchliche Daten
2. fehlende obligate Datenfelder
3. fehlende wichtige Datenfelder und ausstehende Aktualisierungen des klinischen Verlaufs
4. fehlende selten verwendete Datenfelder

Durch die Rückmeldung an die Zentren konnten im Zeitraum von November 2002 bis April 2004 die Fehler der Kategorie 1 zu 74 % und der Kategorie 2 zu 86 % eliminiert werden. In den anderen beiden Kategorien wurden nur 34 % der Fehler behoben. Die Vollständigkeit der Daten in der ProMISe-Datenbank lässt sich durch Vergleich mit den Daten des kumulativen DRST-Surveys ermitteln. Die Vollständigkeit lag für die Jahre 1998 bis 2002 zwischen 82 % und 98 %. Die Vervollständigung der Daten ist dadurch für die Zentren bedeutsamer geworden, dass die Dokumentationsqualität ein wesentliches Kriterium für die seit Januar 2005 im Internet einsehbare Akkreditierung ist.

Arts et al. 2001

Im Rahmen der National Intensive Care Evaluation wurde in den Niederlanden ein Register entwickelt. Hierzu war ein Modell zum

Messen der Datenqualität notwendig, und es wurde ein Rahmenkonzept zur Organisation der Datensammlung und Qualitätssicherung entwickelt.

Zum Messen der Datenqualität werden

- Ziele des Registers beschrieben,
- zu prüfende Datenitems ausgewählt,
- Aspekte der Datenqualität definiert,
- Methoden zur Qualitätsmessung ausgewählt,
- Kriterien definiert,
- Qualitätsmessungen durchgeführt
- gemessene Datenqualität quantifiziert.

Die Prozeduren werden nach zentraler Registerstelle und datensammelnden Stellen unterschieden. Weitere Unterteilungen sind Prozeduren zum Vermeiden, Aufdecken und Korrigieren von Datenfehlern. Das NICE-Register hat die meisten Prozeduren bereits implementiert. Mittlere Fehlerrate beim SDV liegt bei 6,1 % inkorrekte und 2,7 % fehlende Daten. Auf die Berechnung standardisierter Mortalitätsraten für die einzelnen Intensivstationen scheinen diese Fehler keinen Einfluss zu haben.

Arts et al. 2002a

Die Studie wurde im Rahmen der National Intensive Care Evaluation (NICE) durchgeführt und sollte Gründe für unzureichende Datenqualität identifizieren und ein Rahmenkonzept für Datenqualitätsicherungsmaßnahmen aufstellen. Zunächst wurde eine Literatur-Recherche durchgeführt nach Typen und Gründen von Datenfehlern und Prozeduren, die die Fehlerzahl im Register senken können. Es wurden Datenbanken aus Registern und Studien einbezogen.

Es wurden zwei Intensivstationen untersucht, von denen eine die Daten in elektronischer Form bekommt, bei der anderen erfolgt eine Datenerfassung. Für je 20 Patienten fand ein Abgleich statt zwischen dem zentralen Register, der lokalen Datenbank und den Goldstandard-Daten, die aus der Papierakte re-extrahiert worden waren. Dabei wurde auf falsche und unvollständige Items geprüft. Durch strukturierte Interviews mit den verantwortlichen Ärzten wurde der Prozess der Datenerhebung mit möglichen Fehlerquellen deutlicher. Das von Arts et al. erarbeitete Rahmen-

konzept umfasst die wesentlichen Prozesse sowohl für automatische wie manuelle Datenerhebung.

Datenqualität muss aus der Perspektive der Datennutzung bewertet werden. Datenqualitäts-Attribute aus der Literatur waren häufig unklar oder gar nicht definiert, am häufigsten werden accuracy und completeness genannt.

- accuracy = das Ausmaß, in dem Registerdaten der Wahrheit entsprechen
- completeness = das Ausmaß, in dem die Daten registriert wurden, die tatsächlich hätten registriert werden müssen.

In der Literatur werden Fehler unterschiedlich klassifiziert, Van der Putten et al unterscheidet zwischen Interpretationsfehlern, Dokumentationsfehlern und Kodierfehlern. Andere Autoren unterscheiden zwischen zufälligen und systematischen Fehlern, Knatterud et al. nennt Bias als eigenen Fehlertyp. Als Fehlerquellen für systematische Fehler kommen Programmierfehler, unklar definierte Merkmale und Verletzungen des Datenerhebungsprotokolls vor, für zufällige Fehler werden Fehler bei der Transkription, Tippfehler und schlecht lesbare Handschriften als Ursachen genannt. Am häufigsten werden ungenaue Datentranskriptionen und Programmierfehler genannt, auch unzureichende Schulungen und fehlende Guidelines kommen vor.

Die Fehlerquellen in NICE wurden nach zentralen und lokalen Fehlerquellen unterteilt Whitney et al. unterscheiden zwischen Quality assurance (vor Beginn der Datensammlung), z. B. vorherige Schulungen) und quality control (während und nach Datensammlung). Nach Knatterud et al. muss Prävention (Auswahl und Schulungen Personal), Fehlerfindung (Monitoring) und Fehlerkorrektur die gesamte Laufzeit einer Studie hindurch stattfinden. Um die Datensammlung für Register zu standardisieren, sollten Merkmale klar definiert sein und Leitlinien für die Erhebung vorliegen. Viele Autoren empfehlen zentrale Schulungen des beteiligten Personals. Eventuell kann das CRF-Design die Datenqualität verbessern, die Datensammlung sollte möglichst zeitnah erfolgen. Wo möglich, sollten Daten direkt aus EDV-Systemen (z. B. Labor) importiert werden. Doppeleingabe, Plausibilitätsprüfungen und Sichtkontrolle können zur

Fehlerfindung eingesetzt werden. Darüber hinaus sind site visits und Audits möglich. Die Prozeduren im Rahmenkonzept werden getrennt nach zentralen und lokalen Fehlerquellen aufgelistet (weil die meisten Register solche Anteile haben).

Je nach Nutzungszweck müssen Registerdaten unterschiedliche Anforderungen erfüllen (z. B. Vollständigkeit für Berechnung von Inzidenzen). Daher ist eine klare Definition und Beschreibung dessen, was als Datenqualität verstanden wird, notwendig. Standardisierte Definitionen der Datenqualität erlauben einen Vergleich verschiedener Register. Obwohl auch die Patientenakte fehlerbehaftet ist, war sie hier die bestmögliche Datenquelle. Bei automatischer Datenübertragung sind Datenfehler meist Resultat von Programmierfehlern, bei manueller Dateneingabe eher zufällig. Nach Ansicht der Autoren ist das Rahmenkonzept auch für klinische Studien geeignet.

Arts et al. 2002b

Das Dutch National Intensive Care Evaluation (NICE) Register sammelt seit 1996 Daten, um die Effektivität der Intensivpflege zu verbessern. Nach der Implementierung qualitätsverbessernder Maßnahmen im Register wurde die Datenqualität von Scores zur Bewertung des Schweregrads der Erkrankung analysiert. Von neun Intensivstationen wurden jeweils 20 Fälle zufällig ausgewählt. Die zu untersuchenden Daten wurden aus den Originaldaten extrahiert und mit den gespeicherten Registerdaten verglichen.

Der mittlere Prozentsatz unrichtiger Angaben pro Krankenhaus lag bei $6,1 \pm 4,4$ %, der Anteil unvollständiger Daten bei $2,7 \pm 4,4$ %. Die mittlere Differenz zwischen den Registerdaten und den extrahierten Daten lag bei 0,2 Punkten für den APACHE II und bei 0,4 Punkten für den SAPS II. Die mittlere Differenz der vorhergesagten Mortalität gemäß APACHE II bzw. SAPS II betrug 0,4 % bzw. 0,02 %. Diese Ergebnisse zeigen eine gute Datenqualität des NICE-Registers und könnten durch die Qualitätssicherungsmaßnahmen erklärt werden.

Arts et al. 2003

Das holländische Register zur National Intensive Care Evaluation (NICE) basiert auf Daten,

die aus den jeweiligen Patientenakten extrahiert werden. Hierzu gehören auch Scores wie APACHE und SAPS, aus denen der Schweregrad der Krankheit berechnet wird. Der APACHE Score ist schwierig zu berechnen und z. T. ungenau erklärt. Um die Datenqualität im Register zu verbessern, hat NICE die Definitionen ins Internet gestellt und pro Intensivstation müssen zwei Ärzte an einer zentralen Schulung teilnehmen.

Es wurden drei Beispielakten ausgewählt und von zwei Experten bewertet, dann wurden die Scores als Goldstandard berechnet. Bei vier zentralen Schulungen wurden diese Akten vor und nach der Schulung bewertet sowie von einer dezentral geschulten Kontrollgruppe (auch zweimal hintereinander). Die Genauigkeit der Daten wurde bewertet für den gesamten Registerdatensatz, das APACHE subset und das SAPS subset. Datenfehler wurden kategorisiert als unvollständige Daten, Nichteinhaltung von Definitionen und andere Fehler wie z. B. Schreibfehler. Die Nichteinhaltung von Datendefinitionen sank nach der Schulung am stärksten. SAPS-Scores verbesserten sich ebenfalls in der Qualität, APACHE-Scores jedoch nicht. In der Kontrollgruppe waren die Daten bei der ersten und zweiten Bewertung etwa gleich in der Qualität, insgesamt war die Kontrollgruppe so gut wie die Schulungsgruppe nach dem Training. Es lässt sich daher postulieren, dass zentrale wie dezentrale Schulungen etwa gleich wirksam sind.

In einer Studie zur Datenqualität im NICE-Register war die Datenqualität besser. Allerdings werden hier die Daten von den behandelnden Ärzten erhoben, auch war das Studiendesign anders. Die Variabilität im APACHE-Score lässt sich auch in anderen Studien zeigen und sollte bei der Auswertung von Registerdaten berücksichtigt werden.

Benjamin/Watkins 1992

Die Identifizierung von Krankheitsfällen und die Aufnahme der entsprechenden Patienten sind wesentliche Aufgaben beim Aufbau eines Registers. Aus den verschiedensten Quellen werden Daten extrahiert und – oftmals kodiert – in ein Register übernommen. Aufmerksamkeit muss der Qualität und Korrektheit der

Daten sowie dem Prozess der Datenextraktion gewidmet werden.

Die folgenden Vorkehrungen sollten für eine effektive Nutzung eines Registers getroffen werden:

- Benutzer, die ein Register für einen anderen als den ursprünglich geplanten Zweck nutzen wollen, sollten über die Datenquellen unterrichtet werden. Fehlende Kenntnis über die Datenbasis sowie unzureichende Aktualität der Daten kann zu fehlerhaften Auswertungen führen.
- Forscher sollten eine statistische Analyse der missing values vornehmen. Fehlende Angaben aufgrund unvollständiger Sichtung der Datenquellen oder einer nicht durchgeführten Suche nach den fehlenden Werten in anderen Datenquellen sollten vermieden werden, da sie die Reliabilität und Brauchbarkeit der Registerdaten beeinträchtigen.
- Ein Data Dictionary sollte eine komplette Beschreibung aller Datenelemente enthalten und allen Nutzern eines Registers zur Verfügung stehen.
- Alle Nutzer sollten die Historie eines Registers kennen. Hierzu gehören alle Faktoren, die die Konsistenz, Vergleichbarkeit und Korrektheit der Daten beeinflussen können.

Daraus lassen sich für die Betreiber eines Registers Maßnahmen zur Qualitätsverbesserung ableiten:

- Medizinisches Personal und andere Nutzer eines Registers sollten kontinuierlich in den Prozess der Modifikation eingebunden werden.
- Das Data Dictionary sollte laufend aktualisiert und die Nutzer angehalten werden, es auch zu nutzen.
- Alle Modifikationen eines Registers sollten mit Datum, Begründung und Beschreibung der Änderung in einem Protokoll festgehalten werden und den Nutzern zugänglich sein.
- Der Prozess der Datenextraktion sollte genau kontrolliert werden. Eine Stichprobe von Registerdaten sollte regelmäßig durch eine zweite Person erneut extrahiert werden und die Ergebnisse an die datenerfassenden Personen zurückgemeldet werden.

- Ein Bewusstsein für die Bedeutung der Datenqualität sollte bei allen Beteiligten geweckt werden.
- Eine gute Schulung zu Beginn und kontinuierlich ist für das Personal eines Registers wichtig.

CDC 1999

Im National Program of Cancer Registries (NPCR) in den USA wird die Qualitätssicherung auf der Ebene der Ergebnisse und Prozesse betrieben. Auf der Ergebnisebene wird verlangt, dass innerhalb von zwölf Monaten nach Abschluss des Diagnosejahres 97 % der Fälle einen vorgeschriebenen Satz von Standardtests passieren. Nach 24 Monaten muss dies für 99 % der Fälle gelten. Zur Untersuchung der Prozessqualität werden Audits durchgeführt, bei denen die Daten ein zweites Mal extrahiert werden, und alle Informationen aus den Originaldokumenten gesammelt und gepflegt. Schriftlich niedergelegte Anweisungen zur Qualitätssicherung ergänzen das Verfahren.

Cnattningius et al 1990

In Geburtenregistern werden für eine bestimmte Population von Geburten zusätzliche Informationen abgespeichert. Die Datenqualität dieser Register ist bisher wenig untersucht worden. Schweden hat seit 1973 ein Geburtenregister, und 1977 sowie 1988 sind Qualitätssicherungsstudien durchgeführt worden. In das Register gingen Daten ein, die auf standardisierten Formularen erfasst, auf einem Formular zusammengefasst und bei der Dateneingabe im National Board of Health kontrolliert wurden. Nach dem Transfer zum Register wurden die Daten erneut geprüft und korrigiert (1. Modell). Ab 1982 wurde das Zusammenfassungsformular nicht mehr verwendet und die Daten direkt von den anderen Formularen eingegeben. Außerdem wurden mehr Checkboxen verwendet (2. Modell). Zur Qualitätskontrolle wurden Daten aus den Jahren 1974 und 1986 per SDV überprüft und bewertet. Darüber hinaus wurde die Konsistenz der Häufigkeit von Diagnosen überprüft. Verschiedene Fehlerquellen ließen sich identifizieren: Daten werden nicht von der Patientenakte übertragen (hat sich zwischen

den beiden Modellen nicht geändert), Fehler bei der Dateneingabe, Fehler durch schlechtes Formular-Design und Verwendung schlecht lesbarer Kopien, Checkboxen haben die Eingabe verbessert (statt Codierung), Fehler sind in den Quelldaten bereits vorhanden. Die Variabilität in der Diagnosenhäufigkeit lässt sich durch Unterschiede in den Kollektiven, verschiedene Verwendungstraditionen etc. erklären.

Das schwedische Geburtenregister ist für Analysen mit „harten“ Daten gut geeignet, jedoch sind z. B. die Diagnosen mit Vorsicht zu behandeln. Hier muss das behandelnde Zentrum als Faktor mit berücksichtigt werden. Die einzelnen Datensätze sind nicht voll zuverlässig, es können Fehlklassifikationen auftreten.

Gissler et al. 1995

Nach einer Qualitätsprüfung wurde das finnische Geburtenregister 1990 restrukturiert – auf den Formularen wurden offene Fragen mit Codes durch Checkboxen ersetzt und zur Vermeidung der Unterrepräsentation von Totgeborenen und perinatal Verstorbenen die Daten aus dem zentralen Bevölkerungsregister ergänzt. Für 1991 wurde eine erneute Validierung der Daten anhand von 870 Krankenakten durchgeführt und für die interessierenden Variablen ein Datenqualitäts-Index berechnet (Anteil der Fälle mit übereinstimmenden Daten, bei stetigen Variablen zusätzlich ein Index mit erlaubter kleiner Fehlerrate). Die Daten wurden per Chi-Quadrat-Test mit der Validitätsprüfung von 1987 verglichen. Die Validität entspricht in etwa der der 1987er Daten, allerdings lässt sich zeigen, dass Änderungen im Fragebogen einen großen Einfluss auf die Datenvalidität für einzelne Variablen haben (in beide Richtungen). Die Einführung von Checkboxen war sinnvoll, muss jedoch durch klare Definitionen und Anleitungen ergänzt werden.

Gissler et al. 1996

Trotz Liberalisierung in den 70er Jahren des 20. Jahrhunderts hatte Finnland in den letzten Jahren eine der niedrigsten Abtreibungsraten unter den Industrieländern. Vergleiche sind jedoch wegen möglicherweise invalider Daten problematisch. Daher wird geprüft, ob die

niedrige Rate eher ein Problem unvollständiger Registrierung ist. 1977 wurden die seit 1950 gesammelten Abtreibungsdaten computerisiert. Der ausführende Arzt muss pro Fall ein Formular ausfüllen und dies innerhalb von vier Wochen an das Register schicken. Bei fehlenden Daten wird nachgefragt.

Für Oktober/November 1993 wurde eine Stichprobe von 482 in 18 Kliniken durchgeführten Abtreibungen gezogen. Vor Ort wurden die Krankenhaus-Register, Operationslisten und andere Datenquellen identifiziert und mit den Registerdaten verglichen. Die Validität wurde anhand einer kleineren Stichprobe überprüft.

Zum Vergleich diente ein Datenqualitäts-Index:

- $\geq 99\%$ Übereinstimmung = gut
- $90\text{--}98,9\%$ Übereinstimmung = zufriedenstellend
- $\leq 90\%$ Übereinstimmung = ungenügend

Die Vollständigkeit lag bei 99% . 95% der Informationen entsprach den Patientenakten (über alle Variablen), sieben Items hatten eine gute, elf eine zufriedenstellende und vier eine ungenügende Übereinstimmung. Für Variablen, die nicht in jedem Fall auszufüllen waren, war die Datenqualität insgesamt schlechter.

Die sinkenden Abtreibungszahlen sind nicht auf mangelnde Datenqualität zurückzuführen. Mögliche Faktoren sind z. B. kostenfreie Familienplanungs-Services, Sexualekundeunterricht in den Schulen, weite Verfügbarkeit und Verwendung von Kontrazeptiva.

Gissler/Shelley 2002

Es wird über die erneute Prüfung der Daten des finnischen Geburtenregisters berichtet, diesmal bzgl. eines Datenvergleichs zwischen der Erst-, Zweit- und Drittgeburt von Müttern mit Vergleich der Daten zur Fruchtbarkeitsgeschichte, weil sich bei diesen Fragen die Formulierung geändert hat.

Es zeigten sich Übereinstimmungen von $98,5\%$, wobei der Anteil der Übereinstimmungen bei der dritten Geburt geringer wird, auch die Datenqualität wird schlechter. Dies könnte an den neu formulierten Fragen liegen, weil für einige die Informationen zwischen durch nicht erhoben wurden, diese Variablen später jedoch wieder eingeführt wurden.

Goldberg et al. 1980

Register sind ein gutes Hilfsmittel, um die Verbreitung von Krankheiten oder Krankheitsmuster zu studieren. Ihre Brauchbarkeit hängt von der Qualität und Quantität der gesammelten Daten ab. Nach der Herkunft der gesammelten Daten lassen sich Register einteilen in:

- Lokales Krankenhausregister: Hier werden nur Daten von Patienten gesammelt, die dieses Krankenhaus aufgesucht haben.
- Zentrales Register: Hier werden Daten von einer ausgewählten Gruppe von Krankenhäusern zusammengetragen.
- Bevölkerungsbasiertes Register: Hier werden aus einer bekannten Bevölkerung Daten von Patienten gesammelt, die an einer bestimmten Krankheit leiden – unabhängig davon, ob sie in einem Krankenhaus behandelt wurden.

Register können verschiedenen Verwendungszwecken dienen. Die wichtigsten sind:

- Identifikation von Individuen, die einer bestimmten Bedingung genügen
- Unmittelbarer Schutz des Individuums durch die Bereitstellung von Informationen, die im Notfall lebenswichtig sein können
- Überwachung der medikamentösen Behandlung bei Langzeit-Therapien
- Bearbeitung epidemiologischer Fragestellungen, speziell Bestimmung von Inzidenz und Prävalenz für eine definierte Population
- Planung, Durchführung und Evaluation von medizinischen Leistungen
- Evaluation von therapeutischen Maßnahmen
- Bearbeitung von Forschungsfragen
- Unterstützung bei der Ausbildung von Ärzten

Für die Qualität der Registerdaten sind die Vollständigkeit und Validität von entscheidender Bedeutung. Die Vollständigkeit kann definiert werden als der Anteil der Fälle im Register bezogen auf alle Fälle in der Zielpopulation. Validität kann hier definiert werden als der Prozentsatz der Fälle im Register, die gemäß den gespeicherten Daten eine bestimmte Eigenschaft aufweisen sollen und diese Eigenschaft tatsächlich besitzen. In der Praxis be-

deutet letzteres die Übereinstimmung zwischen den Registerdaten und einer zusätzlichen unabhängigen Quelle, die dieselbe Eigenschaft objektiv misst.

Zur Bestimmung der Vollständigkeit können u. a. die folgenden Methoden eingesetzt werden:

- Rückgriff auf Todesbescheinigungen: Vollständigkeit wird hier definiert als der Anteil der registrierten Fälle, der *nicht* zuerst über die Todesbescheinigung identifiziert wurde. Diese Methode ist bei Krankheiten mit einer geringen Mortalitätsrate nicht mehr sinnvoll.
- Unabhängige Fallbestimmung: Die Fallzahl im Register wird mit der Fallzahl aus einer unabhängigen Erhebung verglichen. Eine Variante dieser Technik besteht darin, aus einer Stichprobe Schätzungen für die Vollständigkeit abzuleiten, und diese Schätzungen auf das gesamte Register zu extrapolieren.
- Rückgriff auf historische Daten: Eine bereits bekannte Inzidenz- oder Prävalenzrate wird dazu benutzt, für die Registerdaten eine erwartete Zahl von Fällen zu berechnen. Die Differenz zwischen der erwarteten Fallzahl und der tatsächlich im Register vorgefundenen Fallzahl ist ein Maß für die Vollständigkeit.
- Simulation: Hier wird die Vollständigkeit nicht direkt gemessen. Stattdessen werden, basierend auf den Registerdaten, Muster unvollständiger Datenerfassung und ihre Auswirkungen auf abhängige Variablen simuliert.

Die Validität kann ebenfalls auf verschiedene Weise bestimmt werden:

- Rückgriff auf diagnostische Kriterien: Hier wird der Anteil der Registerfälle bestimmt, der strenge diagnostische Kriterien erfüllt (z. B. histologische Bestätigung bei Krebs). Diese Methode fokussiert auf die Validität von Diagnosen und ist für die Prüfung anderer Variablen (z. B. Geschlecht, Alter) nicht geeignet.
- Nochmalige Ableitung der Registerdaten: Die für das Register zu erfassenden Daten werden ein zweites Mal aus den Originaldaten abgeleitet und mit den bereits vorhandenen Registerdaten verglichen. Die

Daten des zweiten Ableitungsprozesses werden dabei als korrekt angenommen.

- Prüfung der internen Konsistenz: Diese Methode identifiziert unlogische Fälle. Allerdings werden unkorrekte aber logische Fälle nicht aufgedeckt.

In einer Fallstudie wurden die Daten des Illinois Trauma Register (ITR) untersucht. Das ITR besteht seit 1971 und erfasst die Daten aller mit traumatischen Verletzungen in 50 Krankenhäuser in Illinois eingelieferten Patienten. Die Fallstudie sollte Daten für drei Qualitätsmerkmale der Registerdaten liefern:

- Vollständigkeit der Fallberichterstattung der Krankenhäuser an das Register
- Übermäßige Berichterstattung („over-reporting“) der Krankenhäuser, d. h. Fälle aus dem Register, die in den Krankenakten nicht wiederzufinden sind
- Validität der Registerdaten verglichen mit den Originalakten in den Krankenhäusern

Die Untersuchung basierte auf je einer zehnprozentigen Zufallsstichprobe aus den Krankenakten und den Registerdaten. Die detaillierten Studienergebnisse haben die Autoren anderenorts bereits publiziert und stellen hier nur ausgewählte Ergebnisse für den Schweregrad der Verletzung vor. Die Vollständigkeit nahm monoton mit dem Schweregrad von 42 % auf 33 % ab.

Håkansson et al. 2001

Seit 1992 hat Schweden ein nationales Katarakt-Register, u. a. um die Einhaltung einer 1991 beschlossenen Wartezeit-Garantie für Patienten, die auf eine Katarakt-Operation warten, zu prüfen und Effekt und Effizienz des Zugangs zu und der Durchführung von Katarakt-Operationen zu überwachen. Zwischen 1992 und 1998 wurden etwa 95 % der durchgeführten Operationen erfasst. Die Autoren führen Studien an, in denen die Zuverlässigkeit von Register- oder Surveydaten nicht explizit überprüft wurden. Das Register erfasst einen Datensatz von zwölf Variablen (ursprünglich 20), von denen zehn seit Registerbeginn erfasst werden. Pro Operation wird ein Formular ausgefüllt. Die teilnehmenden Chirurgen erhalten ein Handbuch, die Daten werden entweder von Hand eingegeben oder aus

E-Mails direkt eingelesen. Die Plausibilitätsprüfungen in der Datenbank sind streng programmiert. Einmal pro Jahr wird ein Bericht herausgegeben und es gibt ein Treffen der Beteiligten, um Ergebnisse und Probleme zu diskutieren. (Feedback).

Von 15 teilnehmenden Kliniken wurden fünf ausgewählt, davon eine Uniklinik. Die Plausibilitätsprüfungen des Eingabeprogramms wurden verifiziert und die Formulardaten für einen zufällig ausgewählten Monat (September 1998) mit den Patientenakten und den Daten in der Datenbank verglichen. 574 Operationen (ca. 10 % des Gesamtaufkommens für dieses Jahr) wurden geprüft. Alle Patienten, die registriert waren, waren in den vorgeschriebenen OP-Büchern auffindbar. Fehlende und falsche Werte wurden als Fehler betrachtet. Häufig füllten Ärzte nur einen Teil des Formulars aus, die Schreib- oder Pflegekräfte den Rest.

Für 96,2 % der Operationen waren die Daten registriert (Variation zwischen den Kliniken aber hoch). Durch die zweistellige Eingabe des Geburtsjahrs wurden Patienten, die vor 1900 geboren waren, mit dem Jahr 1900 registriert, es ergab sich hieraus nur ein minimaler Fehler. Beim Vergleich mit den Patientenakten ergaben sich Differenzen in 4,57 % der Fälle, die höchsten Abweichungen waren in allen Kliniken bei drei Variablen. Die Autoren führen die Fehlerraten für die verschiedenen Variablen an. Beim Vergleich zwischen den Registerdaten und den Formulardaten ergab sich ein Verlust von 1,57 % der Operationen. Von den 0,20 % unterschiedlichen Variablen war das meiste auf Eingabefehler zurückzuführen.

Die folgenden Fehlertypen wurden gefunden: Schreibfehler, Lesefehler, Dateneingabe von inkorrekten Quellen, Beurteilungs- oder Interpretationsfehler. Teilweise sind Probleme auf unklare Variablendefinition zurückzuführen, teilweise auf Fehler bei der Übertragung vom Freitext in der Akte auf eine Auswahlliste von Diagnosen. Das Ausfüllen der Formulare durch mehrere Personen könnte eine weitere Fehlerquelle sein. Die Validität der Software ist sehr hoch. Die geringe Anzahl an Fehlern bei der Datenerfassung ist darauf zurückzuführen, dass diese von einer einzelnen gut ge-

schulter Person durchgeführt wird. Drei Variablen sind nicht zuverlässig erfasst worden. Hier sollten Fehlerkorrekturmaßnahmen eingeführt werden.

Hedicke et al. 2005

Das vom Bundesministerium für Bildung und Forschung geförderte Kompetenznetzwerk CAPNETZ erforscht die ambulant erworbene Pneumonie (CAP) des Erwachsenenalters. CAPNETZ umfasst derzeit acht lokale klinische Zentren, die in Kooperation mit den niedergelassenen Ärzten ambulant betreute CAP-Patienten identifizieren und in die CAP-Studie aufnehmen. Die niedergelassenen Lübecker Allgemeinmediziner und Internisten wurden zu Studienbeginn um Mitwirkung gebeten. Der Anteil ambulanter Fälle am Fallaufkommen war trotzdem geringer als erwartet. Daraufhin wurden für das Jahr 2004 aus dem bestehenden Praxisnetz zehn Meldepraxen rekrutiert und das Meldeverfahren verändert:

- „Active reporting“: Die Praxen meldeten wöchentlich die Zahl ihrer CAP-Patienten (auch Null) per Fax und erhielten dafür eine Zusatzvergütung.
- „Study nurse“: Jede Praxis benannte eine für die wöchentlichen Meldungen zuständige Mitarbeiterin.
- Kontrollen: Wurde die Nullmeldung zweimal hintereinander vergessen, erfolgte eine Rückfrage.
- „Incentives“: Neben der Zusatzvergütung wurden auch andere Anreize geboten.

Neun angesprochene Praxen nahmen am neuen Meldeverfahren teil. Um den Erfolg des „active reporting“ einzuschätzen, wurden die Meldezahlen der einzelnen Praxen für die Jahre 2003 und 2004 miteinander verglichen. Trotz geringer Fallzahlen und kurzer Beobachtungszeit zeigte sich ein deutlicher Effekt. 80 % der Praxen meldeten eine höhere Zahl von CAP-Fällen während des „active reporting“, 20 % eine geringere. Das „active reporting“ setzt nicht nur auf die Erhöhung der Meldebereitschaft, sondern durch die Wiedererkennung der Studie im Praxisalltag und die Integration der Meldung in die Praxisroutine auf eine systematische Verbesserung der Fallfassung.

Hemminki et al. 1992

Die Daten für das Finnische Geburtenregister werden vom lokalen Krankenhauspersonal im Rahmen seiner Routinetätigkeiten erfasst. In einer Studie sollte das Bedürfnis des Personals nach einem Feedback und dessen Einfluss auf die Datenqualität untersucht werden. Per Interview wurden Angaben zur Einstellung gegenüber dem Register und zum Bedürfnis nach einem Feedback erfragt. Eine zufällig stratifiziert ausgewählte Hälfte der Krankenhäuser, die Daten für das Register liefern, bekam ein „Feedback-Paket“. Dieses Paket enthielt Auswertungen der Registerdaten, die auch einen Vergleich der Häuser erlaubten. Die Registerdaten der Krankenhäuser, die ein Feedback-Paket erhalten hatten, wurden vor und nach der Aussendung der Pakete mit den Registerdaten der restlichen datenliefernden Krankenhäuser verglichen.

Der größte Teil des ärztlichen und pflegerischen Personals hatte eine negative Einstellung zu dem Register. Ein Feedback wurde von einer Mehrheit gewünscht. Die Datenqualität der Krankenhäuser, die ein Feedback-Paket erhalten hatten, verbesserte sich. Im Gegensatz dazu verschlechterte sich die Datenqualität in den anderen Krankenhäusern.

Horbar et al. 1995

Das Vermont-Oxford Studiennetzwerk ist eine Forschergruppe von Neonatologen, die eine Datenbank für Neugeborene mit einem Geburtsgewicht zwischen 501 und 1.500 Gramm betreibt. Die Datenbank liefert Basisdaten für randomisierte Studien, dient als Forschungsbasis und liefert Qualitätsmanagement-Berichte für die teilnehmenden Zentren. Um Aussagen über die Reliabilität der Datenbank und Fehlerquellen der Daten zu gewinnen, wurden bei 635 zufällig ausgewählten Fällen (15 % der geeigneten Fälle pro Zentrum) zehn Datenitems für die Datenbank ein zweites Mal aus der Patientenakte extrahiert und mit den in der Datenbank bereits gespeicherten Daten verglichen. Die ermittelten Widerspruchsdaten bewegten sich zwischen 1,3 % und 10,6 %. Insgesamt gab es 247 Widersprüche. 23 waren Tippfehler, der Rest (224) waren Übertragungs- oder Interpretationsfehler.

Jensen et al. 2002

Aus einer dänischen Region wurden Patientenakten mit Registerdaten für Brustkrebs verglichen. Als Quellen wurden herangezogen: das zentrale Personenregister (eineindeutige CPR-Nummer, wird bei Geburt vergeben), das dänische Krebsregister (bevölkerungsbasiert, seit 1987 Meldepflicht), das Register der dänischen Brustkrebsgruppe und das Register der lokalen onkologischen Abteilung.

Es wurden Patientinnen mit einer Mamma-Ca-Diagnose zwischen dem 01.01.1983 und dem 31.12.1989 ausgewählt (2.099 Fälle) und dafür die Krankenakten und Patientenakten beim niedergelassenen Arzt durchsucht. Nach Ausschluss der für Validierung nicht geeigneten Fälle blieben 1.749 Patientinnen in der Kohorte.

Da carcinoma in situ (CIS) nicht in den Inzidenzraten des nationalen Registers auftaucht, jedoch teilweise registriert wird, gab es hier eine leichte Überregistrierung. Bei den Seitenangaben war die Übereinstimmung hoch (98,5 % links, 97 % rechts), beidseitige Fälle waren im nationalen Register überregistriert. Beim Staging war die Übereinstimmung für lokale Tumore zwischen 87,5 und 97,2 %, bei regionaler Ausbreitung zwischen 87,3 und 91 %. Für die metastasierten Fälle war die Übereinstimmung wesentlich schlechter (65,3 bis 74,2 %). Bezüglich der Inzidenz betrug die Übereinstimmung des nationalen Registers mit der Kohorte 100 %. Bei der Therapie lag die Übereinstimmung zwischen 72,7 % (Chemotherapie) und 95,7 % (Operation). Die Datenqualität wurde bestimmt als

- Completeness = Grad der Fallermittlung im Register
- Correctness = Anteil der klinischen Information aus den Akten, die im Register korrekt wiedergegeben sind

Ein Grund für die Abweichungen bzgl. der metastasierten Fälle könnte sein, dass das Staging im Register nach Informationen erfolgt, die in den ersten vier Monaten nach Diagnose gegeben werden. Dass bei ca. 10 % der Patientinnen, deren nodaler Status in Wirklichkeit unbekannt war, No vergeben wurde, ist ein Problem, weil die Untersuchung der Lymphknoten bei Patientinnen mit infauster Prognose eher unterbleibt, so dass es hier zu einer

Missklassifikation kommt. Obwohl Änderungen im Behandlungsplan aufgrund der Meldeprozeduren häufig nicht nachgemeldet werden, war die Übereinstimmung relativ gut. Durch die multiplen Meldebögen wird allerdings nicht nur die Vollständigkeit erhöht, es entsteht auch die Gefahr einer Überregistrierung. Dass die dänische Brustkrebsgruppe seit 1977 Studien durchführt, hat vermutlich die Datenqualität des Registers positiv beeinflusst.

Katalinic 2005

Der Autor hielt am 16.02.2005 in Essen beim Workshop Medizinische Daten in Forschung und Versorgung einen Vortrag zur Qualitätssicherung von Daten in epidemiologischen Krebsregistern. Mit Verweis auf die International Agency for Research on Cancer werden folgende Ziele aufgestellt:

- Prozessqualität
- Vollständigkeit der Angaben (Meldung, Tumor)
- Plausibilität der Angaben (Meldung, Tumor)
- Richtigkeit der Kodierung
- Beste Tumordinformation
- Zeit Arztkontakt – Meldung
- Ergebnisqualität (Soll)
- Vollzähligkeit > 90%
- Histologisch gesichert > 90%
- Unbekannter Primärtumor < 5%
- DCO-Rate < 5%

Kuntoro et al. 1994

Die Arbeit stellt statistische Techniken der Qualitätssicherung für Register am Beispiel Indonesiens dar. Qualitätssicherung sollte von Beginn an ins System eingebaut werden (zuverlässige Datenquellen verwenden, Datenitems gut definieren), und sinnvolle Standards sollten gesetzt werden. Das Personal kann als Qualitätsinspektoren zur Fehlerfindung eingesetzt werden. Fehler können aufgrund von Fehlinterpretation der Akten, bei der Übertragung vom Papier, beim Programmieren, durch EDV-Fehler oder Management-Entscheidungen entstehen. Die Qualitätssicherungsschleife sollte geschlossen sein (Fehler und Fehlerursache beseitigen, Guidelines etc.). Register sollten von einem Qualitätssicherungs-Komi-

tee überwacht werden, und es sollte ein OP-Handbuch geben.

Die Vollständigkeit muss bzgl. der Patienten wie der erfassten Datenitems gegeben sein. Zur Vollständigkeitsprüfung für Patienten kann die Totenschein-Methode (Anzahl der DCO-Fälle), die independent case ascertainment method (Vergleich mit Daten aus unabhängigen Studien) oder der Vergleich der erwarteten mit den gemeldeten Fällen angewendet werden. Die Vollständigkeit der Items kann mit einer Shewhart-Kontroll-Chart visualisiert werden. Die Genauigkeit und Zuverlässigkeit kann anhand diagnostischer Kriterien (wie viele Fälle sind korrekt, welche falsch registriert worden – vor allem möglich bei histologischer Bestätigung der Diagnose), mit erneuter Aktenprüfung (einer Stichprobe, meist 10 % der Akten) oder mit internen Konsistenzprüfungen (EDV-basiert) analysiert werden. Die Zeitnähe lässt sich durch den Vergleich erwarteter vs. gemeldeter Fälle prüfen. Bei der Berechnung von Inzidenzraten muss ggf. eine Zeitverzögerung berücksichtigt werden. Ein umfassender Datenmanagement-Plan mit Checklisten, Doppeleingabe und Plausibilitätsprüfungen sollte implementiert werden, ggf. mit Pareto Charts.

Die Methoden wurden bei einem Brustkrebsregister in Indonesien angewendet. Als Schlussfolgerungen ergaben sich Probleme mit Adressangaben und die Notwendigkeit der Ausbildung von Ratern und Pre-Tests für Formulare.

Lin et al. 2004

Geburtsurkunden sind eine wichtige Quelle für bevölkerungsbezogene Daten. Um die Validität der Daten des taiwanesischen Geburtsregisters zu überprüfen, wurde bei einer Stichprobe von 2.279 Geburten die Daten aus einem städtischen Krankenhaus in Taipei mit den vorhandenen Registerdaten verglichen. Untersucht wurden Geschlecht, Geburtsreihenfolge, Schwangerschaftswoche und Geburtsgewicht. Die Krankenhausdaten wurden als Goldstandard benutzt. Die fehlenden Werte im Register erstreckten sich von 0 % bis 1,6 %. Die Übereinstimmung beider Quellen lag zwischen 99,0 % und 88,3 %. Unterschiede zwischen den beiden Datenquellen waren

in vielen Fällen durch unklare oder inkonsistente Definitionen bedingt.

Lu et al. 1995

Die Daten von 20.103 Einträgen eines taiwanesischen Geburtenregisters wurden bezüglich Vollständigkeit und Richtigkeit mit den zugehörigen Krankenhausunterlagen verglichen. Die Datenqualität wurde anhand der Fehlerrate klassifiziert:

- Schlecht: Fehlerrate > 10 %
- Akzeptabel: Fehlerrate zwischen 5 % und 10 %
- Gut: Fehlerrate < 5 %

Das Register umfasste 36 Items, von denen 24 Items eine gute, vier Items eine akzeptable und acht Items eine schlechte Datenqualität aufwiesen.

Malin et al. 2002

In den USA wurde von verschiedenen Gruppen (politische Entscheidungsträger, Berufsverbände etc.) gefordert, ein nationales System zur Überwachung der Qualität der Krebsbehandlung einzuführen. Es gibt in den USA ein hierarchisch organisiertes System von Krebsregistern, die unterschiedlichen Zwecken dienen. Die Validität von Krebsregisterdaten bezüglich der Qualität der Krebsbehandlung wurde bisher nicht ausreichend untersucht. Daher haben die Autoren eine entsprechende Untersuchung mit dem Kalifornischen Krebsregister durchgeführt. Untersucht wurden spezifische Daten zur initialen Behandlung von Brustkrebs (Brustbiopsie, Brustkrebsoperation, Lymphknoten-Dissektion, Strahlentherapie, Chemotherapie, Hormontherapie, Stadium gemäß Definition der American Joint Commission on Cancer). Die Registerdaten wurden mit Daten verglichen, die aus der Patientenakte extrahiert wurden. Dabei wurde die Patientenakte als Goldstandard benutzt. Zusätzlich wurde ein Qualitätsscore, basierend auf vier Evidenzbasierten Indikatoren, für die Registerdaten und die extrahierten Daten berechnet.

Die Daten einer Zufallsstichprobe von 304 Patientinnen mit einer ersten Brustkrebsdiagnose im Zeitraum von 1993 bis 1995 wurden untersucht. Die Genauigkeit der Registerdaten war für stationäre Therapien höher als für am-

bulante Therapien (Sensitivität im Bereich von 94,9 % bis 95,9 % vs. 9,8 % bis 72,2 %). Der aus den Registerdaten berechnete Qualitätsscore war im Mittel um elf Prozentpunkte niedriger als der aus den extrahierten Daten berechnete. Die Differenz stieg mit dem Brustkrebsstadium. Bei Patientinnen mit Stadium I betrug die Differenz fünf Prozentpunkte, bei Stadium II 16 Prozentpunkte und bei Stadium III 20 Prozentpunkte. Dies zeigte, dass der Schweregrad der Erkrankung und Behandlungssituation die Validität der Registerdaten beeinflussen.

Aus den Ergebnissen ziehen die Autoren den Schluss, dass Krebsregisterdaten nicht immer zur Beurteilung der Behandlungsqualität geeignet sind, sie aber ein Instrumentarium für die Sammlung von Daten über die Behandlungsqualität bereitstellen können.

Maruszewski et al. 2005

Die European Association for Cardio-Thoracic Surgery (EACTS) trägt derzeit die finanzielle Verantwortung für eine seit den 1990er Jahren bestehende Datenbank, in der europaweit Angaben zu Operationen bei angeborenen Herzfehlern gesammelt werden. Das Management der Datenbank entschied sich für einen Datenverifikationsversuch mittels Source Data Verification (SDV).

In fünf europäischen Zentren, die sich freiwillig gemeldet hatten, wurden im Zeitraum vom 01.06.2004 bis 31.07.2004 die Daten von insgesamt 32 % der Patienten und 34 % der Prozeduren aus dem Jahre 2003 von zwei Mitarbeitern der Datenbank geprüft. Die Prüfung umfasste folgende Datenitems:

- 30-Tage-Mortalität
- Länge des Aufenthaltes im Zentrum
- Geburtsdatum
- Datum der Aufnahme
- Datum der Operation
- Datum der Entlassung/des Todes
- Körpergewicht
- Fallkategorie
- Zeitdauer des kardiopulmonalen Bypasses
- Zeitdauer der Aortaklemme
- Zeitdauer des Kreislaufstillstands

Verglichen wurden die Mittelwerte der einzelnen Items vor und nach der Datenverifikation

mittels Welch's t-Test. Statistische Signifikanz wurde bei einem p-Wert < 0,05 angenommen. Es zeigten sich keine statistisch signifikanten Unterschiede zwischen den nicht-verifizierten und den verifizierten Daten.

Nickelsen et al. 2004

Die Datenbank der Danish Colorectal Cancer Group (DCCG) erfasst Patienten, die seit dem 1. Mai 2001 behandelt wurden. Um die Qualität der gespeicherten Daten zu untersuchen, wurden die Daten aus dem Zeitraum 1. Mai bis 31. Dezember 2001 analysiert. Zur Ermittlung der Vollständigkeit wurden die Daten mit denen des Nationalen Dänischen Patientenregisters verglichen. Zur Ermittlung der Validität wurde die initiale Registrierung aus den Krankenhäusern mit einer zweiten unabhängigen Registrierung basierend auf Fallberichten verglichen und der Prozentsatz der Übereinstimmung berechnet. Die Vollständigkeit betrug 86 %. Die Übereinstimmung betrug je nach diagnostischer Frage zwischen 84 % und 94 % bei einem Kappa-Wert zwischen 0,55 und 0,67.

Nielsen et al. 1996

Die Daten zweier Register (regionales Krankenhaus-basiertes Register und regionales Verschreibungsregister der Apotheken in Nord-Jütland) wurden verglichen für Patienten mit der Entlassungsdiagnose IDDM oder Typ 1-Diabetes. Als weitere Datenquellen wurden Patientenakten von Hausärzten und Krankenhäusern in der Region herangezogen. Patienten erhalten bei Geburt eine eindeutige CPR-Nummer, die in allen Registern als ID verwendet wird.

Die Vollständigkeit im Krankenhaus-Register wurde mit der capture-recapture-Methode geschätzt. Unter der Annahme, dass die Aufnahme in das eine Register die Aufnahme in das andere Register nicht beeinflusst und umgekehrt, sind die Daten hypergeometrisch verteilt.

- Wenn a die Anzahl der in beiden Registern erfassten Patienten ist und b bzw. c die in jeweils einem, dann kann die Anzahl eligibler Patienten durch $(a + b) * (a + c) / a$ geschätzt werden und die Zahl nicht identifizierter Fälle durch $(b + c / a)$.

- Die Vollständigkeit korrespondiert mit der Sensitivität (Anteil der Kranken, die tatsächlich registriert sind)
- Spezifität ist definiert als Anteil der Gesunden, die nicht im Register sind.
- Validität wird geschätzt als positiver prädiktiver Wert

Es wurden 1.621 Patienten identifiziert, die 1993 erstmalig Insulin erhielten und als IDDM im Krankenhaus-Register diagnostiziert waren. Der positive prädiktive Wert für das Krankenhaus-Register lag bei 96,3 % mit einer korrespondierenden Vollständigkeit von 91 %. Das Arzneimittelregister hatte einen positiven prädiktiven Wert von 91,8 % und eine Vollständigkeit von 96,2 %.

Die relativ hohe Qualität lässt sich dadurch erklären, dass die Krankheit chronisch ist, ohne Insulin zum Tode führt, häufig mindestens eine Einweisung nötig ist, das Krankheitsbild gut definiert ist, Insulin nicht für andere Zwecke verwendet wird und die Kosten in Dänemark zu 100 % übernommen werden. Wie gut die Datenqualität tatsächlich sein muss, hängt von der jeweiligen Fragestellung ab. Zum Vergleich zwischen Populationen sollten die Daten vollständig – oder im gleichen Maß unvollständig – sein. Bei prognostischen Studien darf die Diagnose keinen Einfluss auf die Prognose haben. Werden relative Risiken für verschiedene Populationen beschrieben, darf der Faktor keinen Einfluss auf die diagnostische Prozedur haben, da sonst missklassifizierte Daten einen Einfluss auf die Schätzer haben. Beide Register sind für verschiedene Typen epidemiologischer Studien geeignet. Allerdings sollten die Daten für die Insulinverschreibung an Kinder konsistenter erfasst werden. Da die Register in anderen dänischen Regionen analog aufgebaut sind, lässt sich die Eignung auch dieser Daten vermuten.

Parkin/Muir 1992

Im Rahmen der IARC-Publikation „Cancer Incidence in five continents Volume VI“ führen die Autoren die Aspekte Vergleichbarkeit und Qualität der Daten aus. Hierbei benennen sie folgende Indikatoren:

- Vollzähligkeit, z. B. abgeschätzt über
- Synonyme

- Verhältnis Mortalität/Inzidenz
- alterspezifische Inzidenzkurven
- Validität, z. B. abgeschätzt über
- Histologische Bestätigung
- Anteil DCO-Fälle
- Tumore mit unbekannter Lokalisation
- Patienten mit unbekanntem Alter

Schulman/Hahn 1993

Das kalifornische Geburtsdefektregister ist ein bevölkerungsbasiertes Register, welches die Daten von Kindern mit angeborenen Missbildungen sammelt. Geschulte Datensammler besuchen regelmäßig die Krankenhäuser. Sie identifizieren potentielle Fälle und prüfen, ob ihre Aufnahme in das Register erforderlich ist. Falls ja, extrahieren sie demographische und Diagnosedaten für das Register. Um die Qualität dieses Prozesses zu ermitteln, wurde eine zufällige Stichprobe von 30 Fällen pro Datensammler ein zweites Mal durch einen anderen Datensammler überprüft. Die Krankenhäuser wurden nach der Anzahl registrierter Fälle, die während eines Jahres auftraten, als „klein“, „mittel“ und „groß“ klassifiziert und die Stichprobe nach dieser Klassifikation stratifiziert. Die Untersuchung lieferte folgende Ergebnisse: Die Wahrscheinlichkeit, dass ein potentieller Fall übersehen wird, betrug 7 % in einem kleinen Krankenhaus, 4 % in einem mittleren und 1 % in einem großen. Die Wahrscheinlichkeit, dass ein Fall fälschlich nicht in das Register aufgenommen wird, betrug 3 % für kleine und mittlere Krankenhäuser und 1 % für große. Die Wahrscheinlichkeit, dass Registerdaten fehlerhaft extrahiert wurden, betrug 8 % bei kleinen und mittleren sowie 6 % bei großen Krankenhäusern.

Solomon et al. 1991

Das Michigan Department of Public Health (MDPH) ließ im Zeitraum von April bis August 1989 die 42 Register überprüfen, die zu jener Zeit teilweise oder vollständig vom MDPH gefördert wurden. Dies geschah vor dem Hintergrund, dass Kriterien zur Entscheidung über die Einrichtung neuer Register benötigt wurden. Ebenso fehlten Standards für den Aufbau genehmigter Register. Die Autoren definierten ein Register als „eine Datenbank identifizierbarer Personen, die einen klar

definierten Satz von Gesundheitsdaten und demographischen Daten enthält, die für einen spezifischen Zweck der öffentlichen Gesundheitspflege gesammelt werden“.

Bei der Prüfung, ob die Einrichtung eines neuen Registers gerechtfertigt ist, sollten folgende Punkte behandelt werden:

- Evaluation des vorgesehenen Zweckes
- Review der Funktion, Dauer und Anwendungsbereich des Registers
- Betrachtung existierender alternativer Datenquellen
- Untersuchung der praktischen Durchführbarkeit des Registers
- Wahrscheinlichkeit einer ausreichenden Finanzierung zu Beginn und auf lange Sicht

Für die erfolgreiche Entwicklung eines neuen Registers sind folgende Punkte maßgeblich:

- Erstellung eines Implementierungsplans
- Erstellung einer adäquaten Dokumentation
- Etablierung von Methoden zur Qualitätslenkung
- Festlegung der Falldefinition und der Methoden zur Auffindung der Fälle
- Festlegung der Datenelemente
- Festlegung der Methoden für die Datensammlung und -verarbeitung
- Festlegung der Verfahrensweise für den Zugriff auf Registerdaten
- Festlegung von Rahmenbedingungen für die Verbreitung der Registerdaten und Ergebnisse

Steil et al. 2004

In dieser Arbeit wird EuCliD vorgestellt, ein Fresenius-Register für Dialysepatienten basierend auf einer Lotus Domino-Server-basierten Datenbank. Pro Land ist eine Schlüsselperson für die Überwachung der einzelnen Dialysezentren benannt, die ebenfalls von einer Schlüsselperson verantwortet werden. An der Spitze steht der Leiter von Clinical Management Europe. Es werden regelmäßig Berichte zur Behandlungsqualität erstellt.

Die Datenbank hat fünf Module, vier mit Patientendaten (Anamnese, Follow-Up, Labor, Outcome z. B. Anzahl Krankenhaus-Tage) und eine mit Zentrumsdaten. Updates erfolgen

nur bei Datenänderungen. Zu bestimmten Zeitpunkten müssen aber alle Daten aktualisiert sein. Ein Bericht listet die Vollständigkeit der Daten und Fehlwerte werden nachgefordert. Analysen erfolgen mit SPSS.

In der Regel wird mit Benchmarking gearbeitet, allerdings ist der Benchmark-Wert länderspezifisch wegen Unterschieden im Versorgungssystem. Die meisten Unterschiede in den Daten lassen sich auch hierdurch erklären. Durch das Berichtswesen haben sich deutliche Qualitätsverbesserungen in den ersten 1,5 Jahren seit Einführung von EuCliD ergeben.

Teperi 1993

Das finnische Geburtenregister wurde 1987 implementiert. Zur Validierung der Registerdaten wurden drei Kontrollen herangezogen: Daten aus den Patientenakten (Stichprobe von 12 Tagen über stratifizierte Krankenhäuser, 760 Neugeborene), Daten aus einem Formularbogen, der an alle gynäkologischen Abteilungen verschickt wurde (basierend auf Stationsunterlagen) und Totenscheindaten aus dem zentralen statistischen Amt.

In den Registerdaten fehlten in 1/1.000 der Fälle Daten für mehr als 50 % der Variablen. Bei acht Variablen fehlten Daten in > 1 % der Fälle. Für die Verteilung des Geburtsgewichts zeigte sich eine Präferenz zu gerundeten Zahlen. Beim Vergleich mit den Daten aus den Patientenakten war die Übereinstimmung meistens gut (Übereinstimmung ≥ 95 %) bis zufriedenstellend (wenn bei stetigen Variablen kleine Ungenauigkeiten zugelassen wurden). Gründe für schlechte Datenqualität waren häufig technischer Art oder auf unklare Anweisungen zurückzuführen. 60 % bis 70 % der häufigen Diagnosen waren auch registriert. Die Fehler waren gleichmäßig verteilt, während sich bei fehlenden Werten eher Cluster bildeten (z. B. Zusammenhang mit Alter der Mutter). Beim Vergleich mit den Stationsunterlagen zeigten sich große Spannweiten in der Abweichung. Der Vergleich mit den Totenscheindaten (Todesursachen) ergab weniger Übereinstimmung, offensichtlich weil nach Obduktion der Totenschein genauer ist als die Krankenhaus-Information, die ins Register einging.

Der Vergleich mit re-abstrahierten Daten aus Krankenakten wird häufig durchgeführt. Hier ist jedoch die Frage, wie zuverlässig die Akten sind. Vermutlich wird die Fehlerhäufigkeit durch diesen Vergleich systematisch unterschätzt. Durch Zusammenführung über eine eindeutige ID können Daten aus verschiedenen Quellen verglichen werden. Auch bereits aggregierte Daten lassen sich verwenden, allerdings nur bedingt. Durch Untersuchung der Datenverteilung kann ein Einblick in die Datenqualität gewonnen werden. Durch Verwendung verschiedener Quellen konnte die Datenqualität im Register umfassender geprüft werden. Begrenzungen ergaben sich durch die kleine Stichprobengröße. Die Rate fehlender Daten und die Fehlerhäufigkeit waren geringer als bei vergleichbaren Studien. Allerdings war die Datenqualität für diagnostische Variablen eher schlecht.

Zur Verbesserung der Datenqualität werden klare Anweisungen, Schulungen, Dialog mit den Nutzern (Newsletter) und Sichtbarmachung des Datennutzens empfohlen. Allerdings sind Effekte bisher nicht berichtet worden. Zur langfristigen Entwicklung von nationalen Registern wird die Verschlankung (kein Duplizieren lokal registrierter Daten) und Investitionen in die Weiterentwicklung empfohlen.

Teppo et al. 1994

Das Finnische Krebsregister besteht seit 1952 und ist bevölkerungsbasiert. Das Register erhält seine Daten unabhängig von Krankenhäusern, pathologischen und hämatologischen Laboren und niedergelassenen Ärzten sowie von Autopsien und Todesbescheinigungen. Ein wachsender Anteil der Daten wird auf elektronischem Weg übermittelt. Eine kodierte Erfassung der übermittelten Daten erfolgt beim Betreiber des Registers.

Mit der Vollständigkeit der registrierten Krebsfälle sind verschiedene Probleme verbunden:

- Nicht diagnostizierte Tumore: Eine große Zahl von Tumoren wird erst bei einer Autopsie entdeckt. Eine höhere Autopsie-Rate führt damit zu einer höheren Anzahl nachträglich gefundener undiagnostizierter Tumore. Ein ähnlicher Effekt stellt sich bei der Einführung neuer Diagnoseverfahren und Reihenuntersuchungen ein.

- Unsicherheit der Tumordiagnose
- Fehlende oder verspätete Meldung diagnostizierter Tumore

Um die Vollständigkeit des Krebsregisters zu analysieren, wurden seine Einträge über die eindeutige Personenregistrierungsnummer, die jeder Einwohner Finnlands besitzt, für den Zeitraum 1985 – 1988 mit den Einträgen im nationalen Krankenhausentlassungsregister verknüpft. Alle im Entlassungsregister gemeldeten Fälle mit einer Tumordiagnose, die nicht im Krebsregister enthalten waren, wurden daraufhin überprüft, ob sie in das Krebsregister hätten aufgenommen werden müssen. Bei den meisten Tumorarten fehlten 0,1 % bis 6,6 % der Einträge im Krebsregister. Bei drei Tumorarten lag der Prozentsatz über 10 %.

Topp et al. 1997

Bei der Validierung eines Registers muss neben der Vollständigkeit auch die Datenqualität untersucht werden. Im Paper wird das Dänische Register für spastische Lähmungen (Cerebral Palsy) validiert, ein bevölkerungsbasiertes Register, das ca. die Hälfte der dänischen Bevölkerung erfasst. Da die Diagnose erst mit dem 4. Lebensjahr zuverlässig gestellt werden kann, sind epidemiologische Berichte um ca. fünf Jahre verspätet. Das Register soll die Prävalenzveränderungen dokumentieren, die wiederum als Qualitätskontrolle für die obstetrische und neonatologische Praxis dienen. Studien zu Risikofaktoren sind auf zuverlässige Daten angewiesen.

Das Register ist freiwillig und auf persönliche Kontakte mit den Behandlern angewiesen. Die Informationen werden aus den Patientenakten entnommen und enthalten demographische wie medizinische Daten. Zur Validierung der Vollständigkeit wurden die Geburtsjahre 1979 bis 1982 ausgewählt und die Registerdaten mit dem Nationalen Dänischen Patientenregister (DNPR) verglichen, das Informationen zu Entlassungsdiagnosen enthält. Mit Hilfe der so ermittelten fehlenden Fälle wurde die Vollständigkeit bestimmt. Zur Validierung der Datenqualität wurde ein Abgleich mit Daten durchgeführt, die für eine andere Studie direkt aus den Patientenakten entnommen worden waren. Hierfür wurde ein Kappa-Wert berechnet sowie die Sensitivität, die Spezifität und der

positive und negative prädiktive Wert. Die Vollständigkeit betrug 85 %, die Kappa-Werte je nach Variable zwischen 0,43 (UTI) und 0,98 (Geschlecht). Besonders problematisch ist die niedrige Übereinstimmung beim gestational age, wo sich ein systematischer Fehler zeigen ließ. Die Validität konnte wegen der Vergleichsdaten nur für Kinder mit Krankenhausaufenthalten geprüft werden. Die wirkliche Vollständigkeitsrate ist daher nicht bestimmbar, die Ergebnisse weisen jedoch nicht auf einen bedeutenden Bias hin. Es ergeben sich Hinweise darauf, dass die Diagnostik von CP problematisch ist und dass Zeitreihenanalysen dadurch erschwert werden, dass nicht immer der gleiche Neuropädiater ein Kind beurteilt. Die Ergebnisse fließen in die Verbesserung der Datenerhebung ein.

Vestberg et al. 1997

In Dänemark werden alle aus dem Krankenhaus entlassenen Patienten in regionalen Entlassungsregistern registriert. Von dort werden die Daten in ein zentrales Register überführt. Validität (bzgl. Erfüllung diagnostischer Kriterien) und Vollständigkeit eines krankenhausbasierten Registers für Leberzirrhose wurden geprüft. Zum Vergleich dienten Daten des regionalen Pathologieregisters und Krankenakten – durch die Verwendung zweier Referenzquellen konnten beide Maßzahlen berechnet werden. Für die Validität wurde der Anteil der Fälle mit Zirrhose (2 diagnostische Kriterien) an der Gesamtzahl der Fälle berechnet, die Vollständigkeit wurde geschätzt als Anteil der Fälle in beiden Registern geteilt durch die Anzahl bestätigter Fälle im Pathologie-Register. Für die Validität ergab sich ein Wert von 85,4 %, Fehlergründe waren eine falsche Diagnose, Kodierfehler und mehrdeutige Kodierung. Die Vollständigkeit wird mit 93,2 % geschätzt. Die Datenqualität wird als akzeptabel eingeschätzt.

2.4.7.6 Sekundärdatenanalyse

AGENS 2005

Die Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMPP) hat im Jahre 2005 Qualitätsstan-

dards für die Sekundärdatenanalyse in der Leitlinie „Gute Praxis Sekundärdatenanalyse“ (GPS) zusammengefasst. Neben einer Begriffsdefinition enthält sie Empfehlungen zu folgenden Punkten:

- Ethik
- Forschungsfrage
- Studienkonzept
- Datenaufbereitung
- Datenanalyse
- Qualitätssicherung
- Datenschutz
- Vertragliche Rahmenbedingungen
- Interpretation
- Kommunikation und Public Health

Qualitätssicherungsmaßnahmen sollen vor dem Beginn und während der Sekundärdatenanalyse durchgeführt werden. Gefordert werden die Prüfung der Datenintegrität, Plausibilitätskontrollen, die Festlegung personeller Zuständigkeiten und die Dokumentation der Datenaufbereitung.

Sorensen et al. 1996

Infolge der technologischen Entwicklung steht eine immer größer werdende Menge an individualisierten Gesundheitsdaten für die epidemiologische Forschung zur Verfügung. Es ist zu erwarten, dass in Zukunft die Zahl der Studien steigen wird, die auf Sekundärdaten basieren. Sekundärdaten sind Daten, die nicht für einen bestimmten Forschungszweck gesammelt wurden (z. B. Routinedaten aus der Patientenversorgung, Abrechnungsdaten). Sie haben den Vorteil, ohne großen Aufwand sofort verfügbar zu sein, und umfassen oftmals große Stichproben. Der Nachteil ist, dass ihre Sammlung und Qualität nicht unter dem Einfluss des Forschers stehen, sie oftmals nicht alle Aspekte der Forschungsfrage abdecken und manchmal nicht validierbar sind. Einige Faktoren beeinflussen den Wert von Sekundärdaten für die Forschung:

- Vollständigkeit der Registrierung von Individuen
- Genauigkeit und Grad der Vollständigkeit der registrierten Daten
- Umfang der Datenquelle
- Registrierungsperiode
- Datenzugriff, -verfügbarkeit und Kosten
- Datenformat

- Möglichkeiten, die Daten mit anderen Datenquellen zu verknüpfen (record linkage). Die benötigte Qualität der Sekundärdaten hängt von der Forschungsfrage ab.

Swart 2004

Die Adaption der Good Epidemiological Practice (GEP) an die Bedingungen der Sekundärdatenanalyse führt zum Konzept der GPS (Good Practice of Secondary Data Analysis) – erarbeitet von der AG AGENS der DGSMP. Besondere Bedeutung haben bei der Sekundärdatenanalyse die datenschutzrechtlichen Fragen. Da kein Einfluss auf die Datenerhebung genommen werden kann, werden spezifische Empfehlungen für die Prüfung von Validität und Vollständigkeit der Daten sowie für Transparenz der Datenentstehung und -verarbeitung sowohl beim Datenherren wie auch beim Sekundärnutzer gegeben.

Swart et al. 2005

Da die wissenschaftliche Nutzung von Sekundärdaten in den letzten Jahren stetig zugenommen hat, hat die Arbeitsgruppe „Erhebung und Nutzung von Sekundärdaten“ (AGENS) der DGSMP Qualitätsstandards für die Sekundärdatenanalyse formuliert. Die GEP-Empfehlungen wurden in einem mehrstufigen Konsensprozess an die Erfordernisse der Sekundärdatenanalyse angepasst. Die Leitlinie „Gute Praxis Sekundärdatenanalyse“ (GPS) wurde am 15. Januar 2005 verabschiedet. Neben Standards für die Durchführung von Sekundärdatenanalysen enthält sie Grundlagen für vertragliche Absprachen zwischen Forscher und Datenherren.

Swart/Ihle 2005

In diesem Buch werden Aspekte der Datenqualität aus Sicht der Datengrundlage, der Methodik und der Nutzung diskutiert. Im Beitrag zu personenbezogenen Abrechnungsdaten aus Arztpraxen werden auch konkrete Qualitätsindikatoren angegeben (S. 41–42):

- Dubletten (unter 0,01 %)
- Wertebereich für Alter (1 Tag–110 Jahre), Geschlecht (1–2 % fehlende Angaben in den Daten)
- Ausreißerprüfung (viele Diagnosen, Arzt-/Praxiskontakte oder Leistungen)

- Fehlende Werte (Geschlecht), hier fachgruppenspezifische Quote – bei Überschreitung wird die Geschlechtsangabe aller Patienten einer Praxis verworfen
- Diagnosenprüfung: Exoten (= in Mitteleuropa sehr selten), keine Übereinstimmung mit Geschlecht, Alter

2.4.7.7 Data Warehousing

Häussler 1998

Realistische Erwartungen und Forderungen der Initiatoren und Benutzer in Bezug auf die Datenqualität sind für den Erfolg eines Data Warehouse von großer Bedeutung. Datenqualität ist ein Prozess, der über die Korrektur vorhandener Daten hinausgeht und Qualität in der EDV und den Fachabteilungen langfristig sichern und messbar machen soll. Datenqualität kann auf verschiedene Arten analysiert werden. Bei einem intuitiven Ansatz wird projektabhängig entschieden, welche Qualitätskriterien zu messen sind. Mittels theoretischer Methoden wird nach methodischen Fehlern gesucht, die die Daten bei ihrer Entstehung verfälschen. Diese Ansätze betrachten das Problem aus der Sicht der Datenproduzenten und EDV-Systeme und sollten um einen empirischen Ansatz aus der Sicht der Datenkonsumenten ergänzt werden. Die Benutzer müssen die Kriterien bestimmen, die die Daten erfüllen müssen, um genutzt werden zu können. Gute Datenqualität bedeutet dann, dass die Daten von den Datenkonsumenten für die vorgesehenen Aufgaben benutzt werden können. Zur Messung der Datenqualität wird folgender Weg vorgeschlagen und mit Beispielen erläutert:

1. Definition einer Liste von Attributen, die die Datenqualität beschreiben
2. Gruppierung und Gewichtung der Attribute (z. B. von 1 = überhaupt nicht wichtig, bis 5 = sehr wichtig)
3. Bildung von Attributgruppen als Grundlage für die Messbarkeit der Datenqualität

Die Daten eines Data Warehouse können aus verschiedenen externen (z. B. von Marktforschungsunternehmen) und internen Quellen eines Unternehmens stammen. Die Qualität jeder einzelnen Quelle sollte methodisch erfasst und bewertet werden. Die Verantwortung

für die Sicherung der Datenqualität des gesamten Data Warehouse sollte einer Person oder Gruppe mit entsprechenden Vollmachten übertragen werden. Die Datenlieferanten sollten für die Qualität der von ihnen gelieferten Daten selbst verantwortlich sein. Dazu sollten Hilfen angeboten werden, um die Daten an den Schnittstellen zum Data Warehouse überwachen zu können. Wenn den Fachabteilungen Vorteile daraus entstehen, steigert dies die Motivation zur Verbesserung der Qualität der eigenen Daten. Der aktuelle Qualitätsstand (z. B. Anzahl als fehlerhaft zurückgewiesener Datensätze) sollte den Fachabteilungen bekannt gegeben werden. Denkbar ist auch eine interne Zertifizierung von Daten des Data Warehouse. Alle Aspekte der Qualitätssicherung und -verbesserung sollten dokumentiert werden. Zusammenfassend lassen sich folgende Schlüsselfaktoren für eine gute Datenqualität in einem Data Warehouse nennen:

- Detaillierte Analyse der Daten liefernden Geschäftsprozesse und EDV-Systeme
- Ausführliche Definition der Kennwerte und Qualitätsattribute
- Zuverlässige und performante Infrastruktur
- Klare Verteilung der Verantwortung für die Qualität der Daten liefernden Systeme
- Enge Zusammenarbeit mit den Fachabteilungen und offene Kommunikation mit den Datenkonsumenten
- Unterstützung des Top-Managements
- Flexibles und erweiterbares Design des Data Warehouse

Helfert 2001a

Der Autor stellt sein Dissertationsprojekt zur Spezifikation und Messung von Datenqualität in Data-Warehouse-Systemen vor.

Eine Data Warehouse-Datenbank ist von den operativen Systemen losgelöst, logisch zentralisiert, einheitlich und konsistent. Verschiedene Systeme dienen als Datenlieferanten. Eine Transformationskomponente extrahiert und transformiert die Daten, überführt sie in die Data Warehouse-Datenbank bzw. hält kleinere Teilausschnitte als Data Marts vor. Auf die Datenbestände kann der Endnutzer mit Werkzeugen zugreifen. Daneben gibt es ein Metadatenverwaltungssystem.

Nach der Definition von Bode sind Daten Informationen, deren sprachliche Repräsentationsform und materielle Träger auf eine maschinelle Verarbeitung gerichtet sind. Es lassen sich fünf Qualitätsdefinitionen unterscheiden: transzendent, produktbasiert, anwenderbezogen, prozessbezogen und wertbezogen. Die Datenqualität bestimmt sich hinsichtlich des Beitrags zur Erreichung der Ziele des Datenempfängers. Als Grundlage für die Arbeit dienen die auf Seite 5 der Arbeit grafisch dargestellten Qualitätskriterien.

Von TQM ausgehend ergeben sich als zentrale Anforderungen eines Datenqualitäts-Managementsystems: Verpflichtung des Managements (Unternehmensphilosophie und -kultur), Etablierung eines Qualitätsmanagement-Systems, Bereitstellung geeigneter Werkzeuge und Methoden. Zentrale Aufgabenbereiche sind die Qualitätsplanung, -lenkung, -sicherung sowie die kontinuierliche Verbesserung.

In Data-Warehouse-Systemen hat die Datenqualität eine besonders hohe Bedeutung. Besonders wichtig sind Qualitätsplanung und -lenkung. Die Dissertation entwickelt ein Datenqualitäts-Modell zur Spezifikation und Messung der Datenqualität in Data-Warehouse-Systemen unter besonderer Berücksichtigung der Zerlegungssystematik von Qualitätseigenschaften. Qualitätsmerkmale operationalisieren den Qualitätsbegriff, die Prozess- und Produktmerkmale werden durch Qualitätsindikatoren (Kenngrößen) messbar gemacht. Für Indikatoren ist das Basissystem entscheidend.

Dabei stellen sich folgende Fragen/Probleme: Anforderungen an ein Datenqualitäts-Modell, Defizite bisheriger Ansätze, Anforderungen der Endbenutzer an die Daten, Anwendungsszenario, Prüfung der Einhaltung von Datenqualitäts-Vorgaben, Generalisierbarkeit des Qualitätsmodells.

Produkt- und prozessbezogene Messverfahren können anhand der semiotischen Ebene in syntaktische (Vergleich mit syntaktischen Regeln), semantische (Inhalt der Daten) und pragmatische (Datenanwendung) unterschieden werden. Die Messung kann alle oder eine Stichprobe der Daten umfassen und dynamisch oder statisch erfolgen. Prozessbezogen

sind Verfahren, die z. B. Datentransformationen überprüfen. Häufig wird nur ein Ansatz zur Datenqualitäts-Messung berücksichtigt. Das Qualitätsmanagement sollte z. B. über ein Metamodell in die Metadatenverwaltung integriert werden.

Die Dissertation wird ein Datenqualitäts-Modell anhand eines Fallbeispiels entwickeln. Dieses Modell soll anschließend bezüglich seiner Verallgemeinerbarkeit bewertet werden.

Helfert 2001b

Ein Data Warehouse wird im Allgemeinen dazu genutzt, Analysen und Berichte zur Entscheidungsunterstützung zu erstellen. Um diese Aufgabe adäquat erfüllen zu können, ist eine hohe Datenqualität erforderlich und müssen Maßnahmen zur Sicherstellung der notwendigen Datenqualität erfolgen. Eine Erhebung innerhalb großer deutscher und Schweizer Firmen bestätigte dies und führte zur Entwicklung eines Methoden-basierten Datenqualitätsmanagements (DQM). In der Ökonomie werden mehrere Dimensionen von Information unterschieden:

- Semiotik (syntaktisch, semantisch, pragmatisch)
- Medium (ungebunden, an Menschen gebunden)
- Neuheit (objektiv, subjektiv)
- Wahrheit (Richtigkeit erforderlich, Richtigkeit irrelevant)
- Aktualität (statisch, prozessorientiert)

Wissen repräsentiert Objekte der realen Welt. Information wird hier als Wissen verstanden, das in einer Sprache ausgedrückt werden kann, und ist somit eine Teilmenge des Wissens. Daten wiederum sind diejenige Teilmenge der Information, die mit Maschinen verarbeitet werden kann. Das DQM ist eine Adaption des Konzepts des Total Quality Managements (TQM). Der Autor schlägt zur Planung und Messung der Datenqualität verschiedene Charakteristika vor (s. Tabelle).

Semiotic Level	Quality Characteristics	
	Quality of Design	Quality of Conformance
pragmatic	relevance, Completeness	timeliness, actuality, efficiency
semantic	precise data definitions, easy to understand, objective data definitions	interpretability, accuracy, consistent data values, complete data values, believability, reliability
syntax	consistent and adequate syntax	syntactical correctness, consistent representation, security, accessibility

Tabelle zu Helfert 2001 b: Charakteristika für die Planung und Messung der Datenqualität

Jarke/Vassiliou 1997

DWQ (Data Warehouse Quality) ist ein Kooperationsprojekt im Rahmen des ESPRIT-Programms der EU. Das Projekt adressiert drei Bereiche, in denen Qualitätsfaktoren eine entscheidende Bedeutung für ein Data Warehouse besitzen:

- Anreicherung der Semantik von Metadatenbanken mit formalen Modellen der Informationsqualität zur Unterstützung einer adaptiven und quantitativen Designoptimierung
- Anreicherung der Semantik von Modellen zur Beschreibung von Informationsquellen zur Unterstützung einer inkrementellen Änderungsweitergabe und Konfliktlösung
- Anreicherung der Semantik von Schema-Modellen, um die zeitliche und räumliche Verteilung der Daten ausnutzen zu können

Es sollen u. a. Methodologien und Werkzeuge entwickelt werden, die die Evolution und Optimierung von Data Warehouse-Anwendungen unter sich ändernden Qualitätszielen unterstützen. Ein Data Warehouse besteht aus mehreren Subsystemen mit dazwischengeschalteten Prozessen. Diese Komponenten müssen abgebildet werden auf das zu entwickelnde Qualitätsmodell. Eine nähere Untersuchung zeigt die Zusammenhänge zwischen

Qualitätsfaktoren und Aspekten eines Data Warehouse:

- **accessibility:** Datenquellen, Data Warehouse-Design, Data Warehouse-Prozesse
- **interpretability:** Data Warehouse-Design, Modelle und Sprachen, Query-Verarbeitung, Data Warehouse-Daten und -Prozesse
- **usefulness:** Update-Strategie, Data Warehouse-Evolution, Datenquellen, Data Warehouse-Design, Data Warehouse-Prozesse
- **believability:** Datenquellen, Data Warehouse-Design, Data Warehouse-Prozesse
- **validation:** Data Warehouse-Prozesse

Die Architektur eines Data Warehouse lässt sich aus verschiedenen Ebenen betrachten:

- Konzeptuelle Ebene (Geschäftsebene)
- Logische Ebene (Datenmodellierung)
- Physikalische Ebene (verteilter Informationsfluss)

Die Autoren stellen kurz die diesen Ebenen zugeordneten Fragestellungen vor, die mit Forschungsprojekten innerhalb des DWQ-Projektes bearbeitet werden sollen.

Jung/Winter 2000

Die Autoren beschreiben die Zwischenergebnisse des Kompetenzzentrums Data Warehousing Strategie am Institut für Wirtschaftsinformatik der Universität St. Gallen. Dort wird ein integriertes Vorgehensmodell entwickelt, das alle wesentlichen nicht-technischen Aspekte des Aufbaus und Betriebs eines Data Warehouse in Großunternehmen abdeckt.

Die Sicherung der Datenqualität ist ein wesentlicher Erfolgsfaktor von Data Warehouse-Projekten. Qualitätsmerkmale lassen sich u. a. in vier Kategorien einteilen:

- **Innere Datenqualität:**
 - Genauigkeit
 - Objektivität
 - Glaubwürdigkeit
 - Vertrauenswürdigkeit
- **Kontextabhängige Datenqualität:**
 - Relevanz
 - Vollständigkeit
 - Anwendungsbezug
 - Zeitlicher Bezug
 - Informationsgehalt
- **Darstellungsqualität:**
 - Interpretierbarkeit

- Widerspruchsfreiheit
- Knappe Darstellung
- Verständlichkeit
- **Zugangsqualität:**
 - Zugriffsmöglichkeit
 - Sicherheit

Zur Messung der Datenqualität kann ein aus drei Kategorien bestehendes Messsystem eingesetzt werden:

- **Subjektiv:** Erfassung subjektiver Qualitätsmerkmale durch Fragebogen
- **Objektiv (anwendungsunabhängig):** Erfassung quantifizierbarer, objektiver Variablen (z. B. Korrektheit); die anwendungsunabhängige Objektivierbarkeit ist manchmal fraglich (z. B. bei der Vollständigkeit)
- **Anwendungsabhängig:** Erfassung anwendungsspezifischer Variablen; Wissen über Daten, ihre Anwendungsbereiche und Formate ist hier erforderlich

Ein Teil der Datenqualitätsmessung kann automatisiert werden, wenn das Messverfahren in einer angereicherten Metadatenbank formal abgebildet wird. Über diesen Weg sind jedoch nicht alle Qualitätsziele operationalisierbar. Die im Bereich des Data Warehousing eingesetzten Werkzeuge zur Unterstützung des Datenqualitätsmanagements lassen sich verschiedenen Gruppen zuordnen:

- **Auditing-Werkzeuge:** Die Inhalte von Datenfeldern werden basierend auf Geschäftsregeln auf Widersprüche untersucht. Die angewandten Regeln müssen meist explizit definiert werden.
- **Werkzeuge zur Erkennung von Geschäftsregeln:** Aus Beziehungen, die zwischen Daten bestehen, werden Regeln generiert.
- **Werkzeuge zur Datenbereinigung:** Dieser Vorgang lässt sich in sieben Schritte zerlegen:
 1. Zerlegung der Daten in atomare Einheiten (Parsing)
 2. Standardisierung der Daten gemäß Regeln des Data Warehouse (z. B. standardisierte Texte)
 3. Vergleich der Daten mit anderen Datenquellen
 4. Entfernung von Duplikaten und Zusammenführung der Daten eines Objektes

5. Inhaltliche Anpassung der Daten (z. B. Währungsumrechnung)
6. Feststellung von Beziehungen zwischen einzelnen Objekten
7. Dokumentation der Datenbereinigung in den Metadaten
 - Werkzeuge zur Kontrolle bei der Dateneingabe
 - Metadatenmanagement-Werkzeuge
 - Migrationswerkzeuge

Winter et al. 2003

Folgen schlechter Datenqualität in Datawarehouse-Systemen sind laut Studie des TDWI 2001: zusätzlicher Zeitaufwand, Verlust an Vertrauen in das System, zusätzliche Folgekosten z. B. durch Mehrfachversand von Materialien, Kundenzufriedenheit, Ertragsverluste, Probleme bei der Erfüllung gesetzlicher Auflagen. Trotzdem hatten 48 % der befragten Unternehmen keine Strategie zur Datenqualitäts-Verbesserung.

- Wand/Wang 1996 fokussieren auf Informationssysteme: Inkonsistenzen zwischen Informationssystem und Wirklichkeit führen zu Datenqualitäts-Mängeln. Vier Datenqualitäts-Merkmale: Vollständigkeit, Eindeutigkeit, Bedeutung, Korrektheit.
- English 1999 unterscheidet Datendefinitions-, Architektur-, Datenwert- und Datenpräsentationsqualität, diesen Kategorien werden Merkmale zugeordnet.
- Wang/Strong 1996 nennen vier Kategorien: innere Datenqualität, kontextabhängige Datenqualität, Darstellungsqualität und Zugangsqualität mit je unterschiedlichen Qualitätsmerkmalen.
- Jarke et al. 1999 gliedern Datenqualität anhand Entwicklung und Verwaltung, Softwareimplementierung, Datennutzung und ordnen Datenqualitäts-Merkmale zu.
- Hinrichs 2001/2002 kategorisiert nach Glaubwürdigkeit, Nützlichkeit, Interpretierbarkeit, Schlüsselintegrität und ordnet 13 Datenqualitäts-Merkmale zu.

Die Autoren unterteilen den Qualitätsbegriff nach der anwenderbezogenen (externen), produktbezogenen (konzeptionellen) und herstellungsbezogenen (prozessorientierten) Ebene. Der externen Ebene entspricht der Informationsbedarf der Datenverwender, der konzeptionellen

Ebene die Datenschemata (Entitäten und Eigenschaften der Datenobjekte), und den Produktionsprozessen entspricht die Datenbereitstellung. Es wird zwischen Design- und Ausführungsqualität unterschieden. Nach Spezifikation wechselt die Sicht des Qualitätsmanagements von der Design-Qualität (Datenschema) zur Einhaltung der in den Spezifikationen festgelegten Anforderungen (Datenwerte).

Der Deming-Kreis besteht aus:

- Plan (Qualitätsplanung = Bedürfnisse und Erwartungen erfassen, in Vorgaben umsetzen, Qualitätsanforderungen festlegen – hierfür Qualitätsmerkmale auswählen, klassifizieren und gewichten)
- Do (Qualitätslenkung = Einhaltung von Spezifikationen und Beherrschung der Prozesse, Messung der Produkt- und Prozessqualität, Rückkopplung in Regelkreisen)
- Check (Qualitätssicherung = strukturelle Unterstützung der Qualitätslenkung, Risiken systematisch erkennen, aufdecken und bekämpfen – Risikoanalysen)
- Act (Qualitätsverbesserung = kontinuierliche Verbesserung – Verbesserungsprojekte)

Zur erfolgreichen Umsetzung muss das Management Datenqualität als Unternehmenskultur vorleben, ein Qualitätsmanagement-System muss etabliert werden, und geeignete Methoden und Werkzeuge sind zur Verfügung zu stellen.

Die Autoren stellen das Datenqualitäts-Management der Credit Suisse vor. Das Datenqualitäts-Modell ist in die Metadatenverwaltung integriert. Datenqualitäts-Regeln prüfen technische und sachliche Sachverhalte und führen zu Datenqualitäts-Aussagen. Bei Regelverletzungen wird ein Logfile geschrieben.

Wichtigste Datenqualitäts-Management-Aktivität ist die Spezifikation und Erfassung der Regelmengen (Regeln als SQL-Skripte). Pro Regel werden die Bewertung des Resultats und die Vorgehensweise nach der Prüfung festgelegt. Basis sind z. B. Integritätsregeln für Datenbanken.

Die Datenqualitäts-Messung erfolgt durch Anwendung der Regeln. Messresultate und falsche Werte (Logfile) werden gespeichert und