

# **APS-Weißbuch Patientensicherheit**

**Sicherheit in der Gesundheitsversorgung:  
neu denken, gezielt verbessern**

**Matthias Schrappe**

**Herausgegeben vom Aktionsbündis Patientensicherheit e.V.  
Gefördert durch den Verband der Ersatzkassen e.V.  
Mit Geleitworten von Jens Spahn, Don Berwick und Peter Durkin**

Version 1.2 vom 30.06.2018

Es handelt sich hier um die ursprüngliche Fassung des Gutachtens einschließlich des Vorwortes des Autors, jedoch ohne die anderen Vor- und Geleitworte. Inhaltlich ist der Text mit der Buchveröffentlichung identisch.

Kapitel 3:

## **3. Erhebungsmethodik und Epidemiologie Unerwünschter Ereignisse**

### **Gliederung**

3. Erhebungsmethodik und Epidemiologie Unerwünschter Ereignisse	245
3.1. Einführung	246
3.2. Terminologie und Nomenklatur	253
3.3. Die kritische Rolle der Erhebungsmethodik	263
3.3.1. Einführung: vier Schritte	263
3.3.2. Zielorientierung	267
3.3.3. Das Erkenntnisinteresse der Erhebung	270
3.3.3.1. Wissenschaftliche Perspektive ( <i>scientific pathway</i> )	275
3.3.3.2. Klinisch-epidemiologische Erhebung	279
3.3.3.3. Monitoring durch Indikatoren	282
3.3.3.3.1. Einführung	282
3.3.3.3.2. Patientensicherheitsindikatoren	288
3.3.3.3.3. PSI-Sets	293
3.3.3.4. Generierende Verfahren: CIRS und Co.	299
3.3.4. Messinstrumente	302
3.3.4.1. Allgemeines	302
3.3.4.2. Direkte Beobachtung	305
3.3.4.3. Befragungen	306
3.3.4.3.1. <i>Patient-Reported Outcome Measures</i>	306
3.3.4.3.2. Erhebungen der Patientensicherheitskultur	307
3.3.4.4. Trigger-gestützte Instrumente	313
3.3.4.5. <i>Big Data</i> : Die Analyse unstrukturierter Daten	315
3.3.5. Datenquellen	316
3.3.6. Zusammenfassung und erste Empfehlung	319
3.4. Der zentrale Begriff der Haftungslücke ( <i>litigation gap</i> )	323
3.5. Zurechenbarkeit und Vermeidbarkeit	328
3.6. Neuere Studien unter besonderer Beachtung der Situation in Deutschland	332
3.6.1. Allgemeine Vorbemerkungen	332
3.6.2. Internationale Studien seit 2007	338
3.6.2.1. Trigger-gestützte Studien	338
3.6.2.2. Systematische Reviews	341
3.6.2.3. Mortalität	342
3.6.2.4. Interventionsstudien	344
3.6.2.5. Spezifische Datenquellen	346
3.6.2.6. <i>Patient Reported Outcomes</i>	347
3.6.3. Übertragbarkeit auf Deutschland und deutsche Studien	348
3.7. Zusammenfassung	351

*"Forth, gaming is universal."*

Lucien Leape, in: *"The Checklist Conundrum"*, NEJM  
370, 2014, 106

### **3. Erhebungsmethodik und Epidemiologie Unerwünschter Ereignisse**

### III

Dieses **dritte Kapitel** widmet sich epidemiologischen Fragestellungen: wie häufig sind Unerwünschte Ereignisse, wie oft sind sie vermeidbar, welche Bedeutung kommt ihnen hinsichtlich der (vermeidbaren) Morbidität und Mortalität zu? Insbesondere wird nochmals auf die Frage eingegangen, welche Daten für Deutschland vorliegen und wie es um die Übertragbarkeit internationaler Daten auf Deutschland bestellt ist.

Hinter den Häufigkeitsdaten steht natürlich die Frage nach der Erhebungsmethodik: wie kommt man überhaupt zu verlässlichen Zahlen? Die grobe Größenordnung ist klar und wird in aller Welt verlässlich bestimmt, aber eine der großen Aufgaben der Zukunft liegt in der Verfeinerung der Methodik, so dass man auch kleinere Verbesserungen z.B. im Rahmen institutioneller Initiativen nachweisen kann. Zu einer Erfassung gehören vier Fragen: aus welcher Perspektive, mit welchem Erkenntnisinteresse (z.B. Wissenschaft, einfaches „Zählen“, Monitoring), mit welcher Messmethodik und mit welchen Datenquellen. Die Beantwortung dieser Fragen in der richtigen Reihenfolge ist entscheidend, nicht zuletzt deshalb, weil jede Erhebung auf den Kontext und den Gegenstand der Messung zurückwirkt.

Wegen der Wichtigkeit der Messmethodik ist es unerlässlich, noch einmal auf die den Erhebungen zugrundeliegende Terminologie (lineares Modell!) einzugehen, weiterhin auf den *litigation gap* (die Differenz zwischen Zahlen aus juristischen bzw. haftungsrechtlichen Registern und aus epidemiologischen Untersuchungen) und die beiden Begriffe der Zurechenbarkeit und Vermeidbarkeit von (Unerwünschten) Ereignissen.

Am Schluss wird dann ein Überblick über neuere Studien zur Häufigkeit (in Deutschland und im internationalen Rahmen) gegeben. Es wird kein erneuter Systematischer Review angefertigt, aber die Reviews von 2006/2008, die das APS damals angefertigt hat, werden aktualisiert. Die epidemiologischen Daten sind heute sehr viel robuster als in der Vergangenheit, weil nun serielle Untersuchungen (z.B. aus den Niederlanden) und Daten aus Interventionsstudien vorliegen, beide Studientypen verfügen über eine hohe Validität. In der Gesamtsicht ist eines klar: an der Größenordnung des Problems hat sich in den letzten 10 Jahren nichts geändert, die Angaben zur Häufigkeit (im Krankenhaus) von Unerwünschten Ereignissen (5-10%), Vermeidbaren Unerwünschten Ereignissen (2-4%), nach epidemiologischen Kriterien erhobene Behandlungsfehlern (1%) und vermeidbaren Todesfällen (0,1% der Krankenhauspatienten) stellen eine sehr konservative Schätzung dar. So sind z.B. Unerwünschte Ereignisse durch diagnostische Fehler, durch das Unterlassen von Maßnahmen oder Maßnahmen aus dem Bereich der Überversorgung deutlich unterrepräsentiert, und es fehlen die Unerwünschten Ereignisse aus der ambulanten Versorgung genauso wie aus Pflegeeinrichtungen.

### **3.1. Einführung**

Die Quantifizierung von Patientensicherheit (und Qualität) ist ein heiß diskutiertes und umstrittenes Feld, leicht verständlich, wenn man bedenkt, wie bedeutungsvoll diese Daten sein können, und wie wirkungsträchtig methodische Einzelfragen sind, die in die Erstellung dieser Daten einfließen. In Deutschland und etwas früher in den USA und anderen Ländern hat man zunächst mit der Erhebung von Daten zu Zwecken der internen Qualitätsentwicklung begonnen, bald ist man jedoch dazu übergegangen, diese Daten auch zur Strukturentwicklung und Steuerung zu verwenden (z.B. Qualitäts-orientierte Vergütung und Krankenhausplanung in Deutschland, *Value Based Purchasing* (VBP) in den USA). Dies ist nicht grundsätzlich vorwerfbar, stellt jedoch eine große Herausforderung dar, denn gut gemachte Erhebungsinstrumente werden problemorientiert und vor dem Hintergrund eines Grundkonzeptes entwickelt und neigen dazu, ihre Gültigkeit zu verlieren, wenn man Ziel und/oder Konzept verändert (Austin et al. 2016).

In dieser Situation muss man Anforderungen, die direkt aus methodischen Überlegungen abzuleiten sind, von solchen unterscheiden, die aus den Besonderheiten des Themas Patientensicherheit resultieren. Die methodischen Überlegungen gliedern sich in einen Vierklang:

- Zielorientierung: welches Problem soll gelöst werden,
- Erkenntnisinteresse: unter welcher Perspektive wird die Erhebung durchgeführt (z.B. wissenschaftliche Untersuchung oder Monitoring),
- Auswahl der Messinstrumente: z.B. Befragungen oder Trigger-Methoden,
- Auswahl der Datenquellen: z.B. Problematik der Abrechnungsdaten.

Bei den Aspekten, die aus dem Thema Patientensicherheit abzuleiten sind, sind die drei Ebenen der Definition zu unterscheiden (s. Kap. 2.6.2.), nämlich ob man einen Zustand beschreiben will, sich den (verborgenen) Eigenschaften der Akteure annähern möchte, oder ob man das Ziel verfolgt, deren Innovationskompetenz abzuschätzen.

Diese hier genannten Aspekte stellen die Vorüberlegungen jeder Erhebung im Bereich der Patientensicherheit dar und müssen nach Wahl der Perspektive und der Instrumente noch ergänzt werden, z.B. bei einer wissenschaftlichen Untersuchung durch die Modellbildung über das Zusammenwirken von Messung, Intervention und Kontext. Vor dem Hintergrund dieses eigentlich notwendigen Horizonts ist der Schwerpunkt der heutigen Aktivitäten sehr eng gefasst und liegt nach wie vor auf der Zustandsbeschreibung (Häufigkeit von Unerwünschten Ereignissen) auf Ebene der Organisationen (Krankenhaus) mit einfachen Methoden (v.a. klinisch-epidemiologische Erhebung mit Abrechnungsdaten, oft als „Indikatoren“ bezeichnet). Aussagen zur Innovationskompetenz oder zu den Eigenschaften der Akteure liegen kaum vor.

Aber selbst diese Situation weist blinde Flecken auf, denn diese Erhebungen werden nur in einigen wenigen Bereichen der Versorgung durchgeführt (operative Akutmedizin), und es wird dabei das lineare *end results*-Konzept (Berwick 2015) verwendet, das versucht, Patientensicherheit alleine über die quantitative Erfassung von *Outcomes* zu fassen (s. Kap. 2.2.). Gerade die Evaluation von Interventionen zu Verbesserung von Patientensicherheit ist mit diesen groben Instrumenten ein schwieriges Unterfangen. Man könnte es auch so sagen: um Patientensicherheit zu quantifizieren, wartet man erstmal ab, dass etwas passiert (UE auftreten). Diese Situation führt vor Augen, dass die Erhebung von Patientensicherheit auf dem Boden eines konzeptionellen Verständnisses stattfindet, das im Vergleich zur konzeptionellen Diskussion um Patientensicherheit, so wie in Kap. 2 dargestellt, als äußerst wenig entwickelt bezeichnet werden muss. Hierauf bezieht sich der erste von drei Punkten, die einleitend diskutiert werden müssen:

**1. Erhebungsmethodik als Spiegelbild des Konzeptes von Patientensicherheit:** Der Leser bzw. die Leserin wird in diesem Kapitel 3 ein *dejà vue* erleben. Wenn im vorangegangenen 2. Kapitel ein Konzept von Patientensicherheit entwickelt wurde, das dessen Kontextbezogenheit, dessen fach-wissenschaftliche Vielgestaltigkeit (die sechs „Schulen“) und die Innovationskompetenz der Akteure in den Mittelpunkt stellt und von einem linearen Prozessverständnis abrückt, dann sieht man sich bei der Beschäftigung mit den epidemiologischen Daten und Erhebungskonzepten wieder vollständig in die lineare Welt zurückversetzt. Die Patientensicherheitsindikatoren-Sets (PSI-Sets), um nur ein Beispiel herauszugreifen (Kap. 3.3.3.), umfassen ausschließlich Unerwünschte Ereignisse, wir sind also wieder bei einem Verständnis von Patientensicherheit angelangt, das in Kap. 2.1. diskutiert wurde und lautet: Sicherheit ist wenn nichts passiert.

Zwei schwerwiegende Fehlentwicklungen sind hier anzusprechen:

- **Dissoziation von Konzept und Erhebungsmethodik:** Die (international breit geführte) Diskussion zur Weiterentwicklung des Verständnisses von Patientensicherheit (s. Kap. 2) und die epidemiologischen Erhebungsmethoden haben sich entkoppelt. Die konzeptionelle Diskussion hat sich weit auf das Feld von Kognitionswissenschaft und Systemtheorie vorgearbeitet (das System muss lernen können und innovativ sein), die Erhebungsmethodik ist aber mit allen ihren Facetten aber im linearen „Weltbild“ steckengeblieben (Patientensicherheit ist gegeben, wenn keine postoperativen Lungenembolien auftreten). Selbst die an der Oberfläche liegende Frage der Vermeidbarkeit wird nicht einbezogen (s. Kap. 3.5.), sondern es werden die postoperativen Lungenembolien gezählt, ohne dass geprüft wird, ob diese vermeidbar wären (hierzu: Pronovost und Colantuoni 2009). Das soll gar nicht heißen, dass ein lineares Grundverständnis in einigen Situationen nicht durchaus angebracht wäre: ein zu spät geschickter Arztbrief bleibt ein zu spät geschickter Arztbrief. Trotzdem muss die Frage diskutiert werden, ob die Ereignisse, die wir z.B. mit den Patientensicherheitsindikatoren-Sets quantifizieren, wirklich Sicherheit

abbilden – denn nur in diesem Fall wird man zu einem „Lernzyklus“ auf organisatorischer oder Systemebene kommen.

• **Defizite in der Umsetzung der Erhebungsmethodik:** Unabhängig vom vorgenannten Punkt wird die Erhebungsmethodik, zusätzlich zu ihren konzeptionellen Defiziten, in vielen Fällen nicht adäquat eingesetzt. Selbst wenn man also die Beobachtungsendpunkte akzeptieren würde, werden sie nicht reliabel erfasst, und sie sind hinsichtlich ihrer Validität von zweifelhaftem Wert. Nicht reliabel heißt, dass z.B. die Ereignisse vom Untersuchungsgegenstand abhängen (*surveillance bias*) oder insensitive Datenquellen (z.B. Abrechnungsdaten) verwendet werden. Eine mangelnde Validität ist z.B. dann gegeben, wenn Ergebnisparameter verwendet werden, um die Sicherheit des aktuellen und zukünftigen Handelns zu beschreiben, in Wirklichkeit aber eher Umgebungsfaktoren oder die Kenntnisse in der Manipulation der Risikoadjustierungsmodelle wiedergegeben werden. In Kap. 3.3. wird ausführlich auf diese Aspekte eingegangen.

2. In Vorgriff auf Kap. 3.3.2. sind hier einleitend einige Bemerkungen zum Thema **Zielorientierung und Gegenstand** angezeigt. Die alten, ursprünglich auf die interne Verbesserung ausgerichteten Konzepte, die in Deutschland aus der Zeit der Einführung der Fallpauschalen und Sonderentgelte stammen und immer noch unverändert weiterverfolgt werden, stehen derzeit bzgl. ihrer Systemwirkung unter deutlichem Druck. International häufen sich Ergebnisse, die die Steuerungswirkung der Qualitätsdaten auf Systemebene in Frage stellen (Osborne et al. 2015, Etzioni et al. 2015). Das IOM gibt im Vorwort zu seinem Meilenstein-Report zu System-Indikatoren („*Core Metrics for Health and Health Care Progress*“), den es unter Leitung des langjährigen Experten David Blumenthal zusammen mit der *Blue Shield of California Foundation*, der *California Healthcare Foundation* und der *Robert Wood Foundation* erstellt hat (die nun alle in Bezug auf Qualität und Sicherheit im Gesundheitswesen wirklich nicht als ängstlich gelten), folgende Punkte zu bedenken:

- „*First, current measurement efforts are truly problematic*“ und führt den Aufwand aus, der zur Erhebung der Daten auf institutioneller Ebene anfallen,
- „*Second, as valuable as it is, **measurement is not an end in itself**. It is a tool for achieving health care goals. Readers will note **that the core metric set proposed in this report starts with goals**, proceeds through elements that embody or contribute to those goals, and then associates measures with those elements*“ (Hervorh. MS), ein Satz, der speziell in Deutschland aufmerksam gelesen werden sollte, wo man immer wieder die Ziele des ganzen Unterfangens zu diskutieren vergisst und allein von der Datenverfügbarkeit ausgeht, und

- „Third, measurement will fail if it is left to the experts. Because measures reflect goals and aspirations, **their development is fundamentally a political process in the best sense of that term**“ (IOM 2015A, Hervorh. MS).

Jede Messung, expliziter kann man es nicht sagen, braucht einen Problemaufriss und eine Zielorientierung, ohne Fragestellung sollte man jegliche Messanstrengung unterlassen (nicht zuletzt deswegen, um nicht falsche Antworten auf nicht gestellte Fragen zu erhalten) und der Messvorgang muss extern (politisch) autorisiert sein.

Wenn Unerwünschte Ereignisse aller Art (also Komplikationen operativer Maßnahmen, *adverse drug event*, technische Ereignisse etc.) über eine große Zahl von Einrichtungen erhoben werden, wie es in länderspezifischen Erhebungen oft der Fall ist, können die Verbesserung und/oder Verschlechterung einzelner Sub-Entitäten insgesamt zu einem unscharfen Bild führen, das keine Validität besitzt (Shojania und Marang-van der Mheen 2015, Baines et al. 2015). Weiterhin neigen die in den letzten Jahren durchgeführten Erhebungen dazu, die Größe des Problems zu unterschätzen, weil der Fokus lediglich auf negative Ereignisse gerichtet ist, die im Rahmen einer aktiv durchgeführten Behandlung auftreten (*errors of commission*). Wie bereits in Kap. 1 ausgeführt, kann bei einer Maßnahme, die gar nicht durchgeführt wird, auch kein Fehler beobachtet werden, und auch wenn eine Maßnahme zwar fehlerfrei durchgeführt wird, sie aber einen negativen Nutzen aufweist (Überversorgung), liegen diese Ereignisse meist außerhalb des Horizontes der verwendeten Erhebungsmethoden. Es wäre also zu fordern, dass neben den *errors of commission* (Fehler in der Durchführung) Aspekte wie das Unterlassen von Behandlungsmethoden (*errors of omission*) oder der fehlende bzw. negative Nutzen von (fehlerfrei durchgeführten, aber nicht indizierten) Behandlungen einbezogen werden (vgl. James 2013, ausführlicher in Kap. 5.2.2.2.).

In der Konsequenz ist es unabdingbar, ein auf Ziele bezogenes Vorgehen zu wählen, auch wenn man meinen könnte, beim Thema Patientensicherheit sei es einfacher als bei Qualität. Der Qualitätsbegriff ist ja ein sehr bedingter Begriff, denn er muss auf die Anforderungen und die Definition der Merkmale rekurren, um zu einem Messvorgang

**Core Measures des IOM:** (IOM 2015A)

**1. Domäne Healthy People:**

- Length of Life
- Quality of Life
- Healthy Behaviours
- Healthy Social Circumstances

**2. Domäne Care Quality:**

- Prevention
- Access to Care
- Safe Care
- Appropriate Treatment
- Patient-Centered Care

**3. Domäne Cost of Care:**

- Affordability
- Sustainability

**4. Domäne Engaged People:**

- Individual Engagement
- Community Engagement

Den Core Measures sind insgesamt 15 "Best Current Measures" zugeordnet, z.B. bei Safe Care die Rate nosokomialer Infektionen.



(zu einer expliziten Fassung von Qualität) zu kommen (zur aktuellen Diskussion s. Schrappe 2017E). Bei Patientensicherheit steht dagegen der normative, kategorische Charakter im Vordergrund: es müsste ja klar sein, um was es geht. Lasst uns die Unerwünschten Ereignisse zählen, dann kennen wir „die Sicherheit“.

Dieser Standpunkt ist im vorangehenden Kapitel (insbes. Kap. 2.3.) ausführlich diskutiert und letztlich verworfen worden. Man muss die Zahl der Unerwünschten Ereignisse (UE) zwar unbedingt kennen (s. Kap. 3.6.1.), aber Einrichtungen, bei denen keine oder nur wenige UE auftreten, sind nicht zwingend sicherer als andere (s. Kap. 2.2.). Es wurde herausgearbeitet, dass Patientensicherheit im Gegensatz zur landläufigen Annahme einen Kontext-bezogener Begriff darstellt, der nicht alleine steht, sondern sich im Kontext anderer Thematiken durchsetzen muss und im Übrigen den „aktiven“ Kontext (die Umgebungsfaktoren) für die Maßnahmen zur Verbesserung der Patientensicherheit als Resonanzboden bedarf. Patientensicherheit ist ein kontextbezogenes Konstrukt, das in einem komplexen *Throughput* herausgebildet wird, hierauf wurde auch in der Definition nach Kap. 2.6. Bezug genommen.

**3. Welche methodische Instrumente sind geeignet, um Verbesserungsprozesse und –Interventionen zu evaluieren?** Donald Berwick sieht das Ende des „*end result system*“ (Ergebnisqualität) heraufziehen und weist zum wiederholten Male darauf hin, dass für die Veränderung des Verhaltens einzelner Ärzte, Pflegenden und deren Institutionen nicht durch eine einzelne Intervention (z.B. die Information über Qualität), sondern nur durch komplexe Mehrfachinterventionen (*Complex Multicomponent Interventions*, CMCI) zu erreichen sind (Berwick 2008, 2015, Guise et al. 2014B, Schrappe 2017B, S. 410, Schrappe und Pfaff 2017A, S. 24-5). Auf diesen Interventionstyp wird in Kap. 5.7. ausführlich eingegangen, denn es ist von größter Wichtigkeit, zu Interventionen mit realistischen Erfolgchancen fortzuschreiten (s. Kap. 2.5.3.). Eine Grundbedingung hierfür ist jedoch die Entwicklung von adäquaten Messverfahren, die in der Lage sind, im komplexen Umfeld solcher Interventionen verlässliche und valide Befunde zu liefern. Immer wenn es um Messvorgänge geht, die klar umrissene Ereignisse betreffen und an einer großen Zahl von Untersuchungsgegenständen durchgeführt werden (wie z.B. bei großen epidemiologischen Erhebungen an Krankenakten), kann man es mit einer leicht operationalisierbaren Definition und somit einer transparenten und reliabel anzuwendenden Terminologie bewenden lassen. In anderen Fällen (insbesondere in komplexen Zusammenhängen und bei wissenschaftlichen Evaluationen) muss man aber von den Eigenschaften komplexer Systeme ausgehen, z.B. dass die verwendete Beobachtungsmethodik bzw. Nomenklatur vom Untersuchungsgegenstand nicht gänzlich unabhängig ist, sondern den Untersuchungsgegenstand und dessen Kontext maßgeblich beeinflusst. Weiterhin kann die klassische Trennung von Objekt (Untersuchungsgegenstand) und Subjekt (Untersucher und Methodik) nicht in jedem Fall

aufrechterhalten werden (Schrappe und Pfaff 2017, S. 7, S. 56f); auch international wird unablässig auf diese Notwendigkeit hingewiesen (Jha und Pronovost 2016). Überhaupt nicht für die Evaluation von komplexen Intervention geeignet sind Patientensicherheitsindikatoren (im Sinne einer adäquaten Definition), die zum Monitoring größerer Versorgungsbereiche primär sensitiv (nicht spezifisch) eingestellt sind (s. Kap. 3.3.3.3.), natürlich ebenso wenig wie generierende Verfahren (z.B. CIRS). Man muss zu wissenschaftlichen Settings (mit präinterventioneller Modellbildung) und klinisch-epidemiologischen Erhebungen mit Falldefinitionen greifen (s. Kap. 3.3.3.1. und 2.).

Diese genannten 3 Punkte stellen das zukünftige Arbeitsprogramm dar. Um es noch einmal zu betonen: Diese *Caveats* und Einschränkungen dürfen jedoch nicht dazu führen, dass man auf die Durchführung von entsprechenden Untersuchungen und vor allem auf die Zurkenntnisnahme epidemiologischer Befunde verzichtet, auch wenn sie nach einem linearen Modell durchgeführt wurden (ganz abgesehen davon, dass die genannten Einschränkungen in der Vergangenheit fast ausschließlich zu einer Unter- und nicht zu einer Überschätzung des Problems geführt haben). Diese Daten sind unverzichtbar, weil sie

- die Grundlage für die Priorisierung des Themas Patientensicherheit darstellen,
- als Basis für die Initiierung und Planung von sinnvollen Gegenmaßnahmen auf politischer Ebene dienen können,
- für eine adäquate Differenzierung der einzelnen Maßnahmen und
- für die Information der Öffentlichkeit über den Stand der Verbesserungsinitiativen unerlässlich sind.

So wird im vorliegenden *Weißbuch Patientensicherheit* im folgenden Kap. 3.2. zunächst nochmals die lineare Terminologie rekapituliert, um danach die Erhebungsmethodik ausführlich darzustellen. Bei der Aktualisierung der Reviews von 2007/2008 wird in Kap. 3.6. nicht das strenge Design eines Systematischen Reviews zugrunde gelegt, sondern es werden die bereits vor 10 Jahren sichtbaren Entwicklungen anhand von zahlreichen international neu erschienenen Studien diskutiert und fortgeschrieben.

### **3.2. Terminologie und Nomenklatur**

Die in der epidemiologisch orientierten Patientensicherheitsforschung verwendete Terminologie basiert auf einem linearen Verständnis von Sicherheit, zumindest soweit die Häufigkeit in großen Patientenkollektiven z.B. auf nationaler Ebene adressiert wird (Aufgaben-bezogenes Konzept, s. Kap. 2.4.8.). Trotz dieser Einschränkung auf die lineare Prozesssicht ist eine korrekte und konsistente Verwendung der Terminologie für die Messmethodik (Murff et al. 2003) und auch für ein adäquates Verständnis von Fehlern und Unerwünschten Ereignissen von entscheidender Bedeutung, denn durch falsche oder ungenaue Bezeichnungen können nicht nur missverständliche Häufigkeitsangaben, sondern vor allem auch konzeptionelle Unklarheiten impliziert werden. Als drastisches Beispiel kann die irreführende Verwendung des Begriffs „Beinahe-Fehler“ statt des adäquaten Terminus „Beinaheschaden“ gelten (oder analog „Kunstfehler“ statt Behandlungsfehler, „Fehlerkultur“ statt Sicherheitskultur etc.). Ein weiterer Aspekt ist die oft erkennbare mangelnde Abgrenzung juristischer Termini (z.B. Behandlungsfehler) von den epidemiologisch geprägten Begriffen (z.B. Vermeidbares Unerwünschtes Ereignis). Im internationalen Schrifttum zur Häufigkeit von UE und zur Evaluation von entsprechenden Interventionen zur Verbesserung der Patientensicherheit bedient man sich ausschließlich epidemiologischer Termini. Auf die maßgebliche Unterschätzung der Größenordnung des Problems, dem wir bezüglich der Patientensicherheit gegenüber stehen, durch Daten auf juristischer Grundlage wird im nachfolgenden Kap. 3.3. unter dem Begriff des *litigation gap* (Haftungslücke) nochmals ausführlich eingegangen.

Die **klinische-epidemiologische Nomenklatur** hat sich in den letzten Jahrzehnten auf internationaler Ebene weitgehend stabilisiert (s. Kasten und Abb. 3.2.-1.). Es stehen mehrere grundlegende Quellen zur Verfügung, deren Empfehlungen einen hohen Grad an Konsistenz aufweisen und es ermöglichen, Daten zur Häufigkeit auch über unterschiedliche Gesundheitssysteme und Versorgungssektoren zu übertragen und miteinander zu vergleichen – auch wenn dies nicht bedeuten darf, dass es keinen Handlungsbedarf zur Fortentwicklung der Nomenklatur gäbe (s. Kap. 3.4.). Um sich ein umfassendes und weiter vertiefendes Bild zu verschaffen, kann man auf folgende Quellen zurückgreifen (Auswahl):

- den Report *To Err Is Human* des IOM (IOM 1999),
- den erste Report der britischen Gesundheitsbehörde (Donaldson 2000),
- die Zusammenstellung des Sachverständigenrates zur Begutachtung der Entwicklung im Gesundheitswesen (SVR 2007, Nr. 602 ff),
- das kanadische *Patient Safety*-Lexikon (Davies et al. 2003),
- den *technical report* der WHO zur Klassifikation (WHO 2009),
- den Report der *National Patient Safety Foundation* in den USA (NPSF 2015) und

- das soeben neu erschienene Glossar der Arbeitsgruppe der GQMG/GMDS (Sens et al. 2018).

Der *technical report* der WHO ist insofern hilfreich, als hier auch zahlreiche abweichende Definitionen aufgeführt werden, die es ermöglichen, den terminologischen Rahmen abzuschätzen.

Kurz soll hier eine zusammenfassende Darstellung der wichtigsten Begriffe vorgenommen werden (Zusammenstellung s. Kasten). Im normativen Duktus der epidemiologischen Definitionen wird der Begriff **Patientensicherheit** (*safety*) direkt auf das Nicht-Vorliegen von Unerwünschten Ereignissen (UE) bezogen (zur grundsätzlichen Diskussion des Begriffs Patientensicherheit s. Kap. 2.4.4. und 2.4.8.). In *To Err Is Human* heißt es: „*Safety is defined as freedom from accidental injury*“, die Abwesenheit von Unerwünschten Ereignissen (IOM 1999, S. 49). Patientensicherheit wird in einem zweidimensionalen Modell als Bestandteil von Qualität eingeordnet: „*The first dimension identifies domains of quality. These include: safe care, practice that is consistent with current medical knowledge and customization. The second dimension identifies forces in the external environment that can drive quality improvement in the delivery system*“ (IOM 1999, S. 18). Der Begriff findet sich als erster und wichtigster Aspekt der *six aims for improvement* wieder (IOM 2001 S. 39) und wird analog auch von zahlreichen anderen internationalen Institutionen verwendet, so z.B. von der AHRQ (2017A) oder von der OECD in deren zweidimensionalem Qualitätskonstrukt, das sich auch in der deutschen Qualitätsdiskussion wiederfindet (Arah et al. 2006). Patientensicherheit („Patientenorientierung einschließlich Patientensicherheit“) wird ebenfalls in der QM-Richtlinie des G-BA als eines der sechs „Grundelemente“ eines (sektorenübergreifenden) Qualitätsmanagements verstanden (G-BA 2016, Teil A §3). Nach anderen Ansätzen bezieht sich Patientensicherheit mehr auf das Vorliegen präventiver Ansätze („*the reduction of risk of unnecessary harm associated with health care to an acceptable minimum*“ (WHO 2009)) und kommt dem Begriff der *patient safety practices* nahe (Shojania et al. 2002).

Ganz im Mittelpunkt der Terminologie steht der klinisch-epidemiologische Begriff des **Unerwünschten Ereignisses** (UE; *Adverse Event* (AE) oder *harm*), der das Ergebnis für den Patienten in den Mittelpunkt stellt und sich auf drei Kriterien stützt: negativ im Outcome, Behandlungs-assoziiert und nicht beabsichtigt (SVR 2007, Nr. 597). Das IOM hatte sich zunächst in *To Err Is Human* auf die ersten beiden Kriterien beschränkt (IOM 1999 S. 3), in einem der nachfolgenden Reports aber die vollständige Definition nachgeliefert (IOM 2004, S. 32, s. auch Zegers et al. 2009): „*An adverse event results in unintended harm to the patient by an act of commission or omission rather than by the underlying disease or condition of the patient.*“ Im Mittelpunkt steht der Behandlungsbezug, ein Krankheitsprogress stellt demgemäß kein UE dar. Der Begriff des

Behandlungsbezugs wird sehr weit gefasst, so dass auch ein Sturz im Treppenhaus eines Krankenhauses u.ä. als solcher gefasst wird. Für das *Medicare Patient Safety Monitoring System* (MPSMS) der AHRQ reicht eine gewisse Wahrscheinlichkeit des Zusammenhangs zwischen Ereignis und Behandlung: „*An adverse event is defined as an unintended harm, injury, or loss that is more likely associated with a patient’s interaction with the health care delivery system than from an attendant disease process*“ (Hunt et al. 2005). Das dritte Kriterium „unbeabsichtigt“ ist entscheidend bei der Abgrenzung zum Auftreten eines Schadens als Folge einer willentlichen, vorsätzlichen Nichtbeachtung bestehender Regeln (vgl. Kap. 5.4.6 zum Thema Verantwortung). Unerwünschte Ereignisse sind typischerweise auf einen Behandlungsvorgang bezogen (s. Kap. 3.1.), können z.B. aber auch auf das Unterlassen einer Behandlung zurückgehen (s.u. Abgrenzung zwischen *error of commission* und *error of omission*). Von besonderer Bedeutung sind schwerwiegende (*severe*) UE, die nach der Klassifikation der *National Coordinating Council for Medication Errors Reporting and Prevention Scale* folgendermaßen beschrieben werden (A bis D leichte UE; zit. nach James 2013, Classen et al. 2011):

E	<i>temporary harm that requires intervention</i>
F	<i>required prolonged hospital stay</i>
G	<i>permanent harm</i>
H	<i>life sustaining intervention required</i>
I	<i>contributing to death of patient</i>

In der Außenkommunikation ist die Abgrenzung des patientenbezogenen Begriffs des UE vom handlungsbezogenen Begriff des **Fehlers** (*error*) ganz entscheidend, denn umgangssprachlich wird unter Fehler oft nicht nur die Regelverletzung, sondern in Verkürzung des Terminus Behandlungsfehler ein ursächlich abzuleitender Schaden verstanden. Dem ist nicht so: ein Fehler ist ein auf das Handlungsziel abgestimmter Terminus, der entweder die Verwendung eines falschen Plans (sog. Irrtum, *mistake*) oder einen Ausführungs- bzw. Wahrnehmungsfehler (Patzler (*slip*) bzw. Versehen (*lapse*)) beinhaltet. Ein Fehler bedeutet also nicht, dass für den Patienten zwangsläufig ein negatives Ergebnis (somit ein UE) folgen *muss*, meist wird die „fehlerhafte“ Prozesskette unterbrochen und das Ergebnis nicht geschmälert (s. Begriff des Beinaheschadens). Wenn allerdings tatsächlich auf einen Fehler ein UE folgt, dann spricht man von einem vermeidbaren UE (s.u.).

Eine besondere Rolle spielt der Begriff des Fehlers bei der **Diagnostik** (Singh et al. 2017). Ein diagnostischer Fehler liegt vor, wenn eine Diagnose nicht gestellt, mit zeitlichem Verzug gestellt oder falsch gestellt wird. Als vermeidbar wird ein diagnostischer Fehler angesehen, wenn innerhalb eines diagnostischen Prozesses eine Alternative bestanden hätte, ungeachtet der Tatsache, ob der Behandler, das Behandlungsteam, das

System (z.B. Organisation) oder der Patient die Gelegenheit (*opportunity*) eines richtigen Prozederes verpasst hat (Singh et al. 2017). Auch hier gilt, dass nur ein Teil der diagnostischen Fehler zu einem UE führt (s. Kap. 5.2.2.2.).

#### **Nomenklatur zur Epidemiologie (vgl. Abb. 3.2.-1.)**

<ul style="list-style-type: none"> <li>• <b>Patientensicherheit</b> (<i>safety</i>)</li> </ul>	Abwesenheit Unerwünschter Ereignisse (IOM 1999 S. 49; SVR 2003, Nr. 360 ff; SVR 2007, Nr. 602 ff; Arah et al. 2006; NPSF 2015)
<ul style="list-style-type: none"> <li>• <b>Unerwünschtes Ereignis</b> (<i>adverse event, harm</i>)</li> </ul>	Ein unbeabsichtigtes negatives Ergebnis, das auf die Behandlung zurückgeht und nicht der bestehenden Erkrankung geschuldet ist. (IOM 1999 S. 3; IOM 2004, S. 32; SVR 2007, Nr. 597)
<ul style="list-style-type: none"> <li>• <b>Fehler</b> (<i>error</i>)</li> </ul>	Nichterreichen eines geplanten Handlungszieles oder Anwendung eines falschen Plans (IOM 1999, S. 3, 46; WHO 2009, S. 16; SVR 2003, Nr. 367))
<ul style="list-style-type: none"> <li>• <b>Irrtum</b> (<i>mistake</i>)</li> </ul>	Nichterreichen des geplanten Handlungszieles durch Verfolgen eines falschen Plans (IOM 1999, S. 46)
<ul style="list-style-type: none"> <li>• <b>Patzer</b> (<i>slip</i>)</li> </ul>	Nichterreichen eines geplanten Handlungszieles durch einen Ausführungsfehler (WHO 2009, S. 140)
<ul style="list-style-type: none"> <li>• <b>Versehen</b> (<i>lapse</i>)</li> </ul>	Nichterreichen eines geplanten Handlungszieles durch einen Wahrnehmungsfehler (Dean et al. 2002; WHO 2009, S. 122)
<ul style="list-style-type: none"> <li>• <b>Vermeidbares UE</b> (<i>preventable AE</i>)</li> </ul>	Ein auf einen Fehler zurückzuführendes unerwünschtes Ereignis (IOM 1999 S. 24)
<ul style="list-style-type: none"> <li>• <b>Beinaheschaden</b> (<i>near miss</i>)</li> </ul>	Ein Fehler ohne konsekutives Auftreten eines unerwünschten Ereignisses (IOM 2004, S. 34; SVR 2007, Nr. 599; SVR 2003 Nr. 367)
<ul style="list-style-type: none"> <li>• <b>Behandlungsfehler</b> (<i>negligent AE</i>)</li> </ul>	Ein Vermeidbares Unerwünschtes Ereignis, das die Kriterien der Sorgfaltsverletzung erfüllt (IOM 199, S. 24;
<ul style="list-style-type: none"> <li>• <b>Ereignis</b> (<i>event, incident</i>)</li> </ul>	Ein Ereignis stellt einen Zwischenfall, einen Prozess, eine Vorgehensweisen oder ein Ergebnis dar, das das Risiko für das Eintreten eines unerwünschten Ereignissen erhöht oder tatsächlich in ein unerwünschtes Ereignis mündet, und auch das Versagen von Präventionsmaßnahmen mit einschließt (ÄZQ 2005; WHO 2009, S. 16; Davies et al. 2003. S.43)
<ul style="list-style-type: none"> <li>• <b>Kritisches Ereignis</b> (<i>critical incident</i>)</li> </ul>	Ereignis, das das Risiko für das Eintreten eines schwerwiegenden unerwünschten Ereignissen erhöht oder tatsächlich in ein schwerwiegendes unerwünschtes Ereignis mündet. Ein kritisches Ereignis bedingt die sofortige Untersuchung und Reaktion (Davies et al. 2003, S. 43)

Abzugrenzen ist die willentliche (absichtliche) und missbräuchliche Verletzung von Regeln (*violation*), eine Handlung, bei der die Regeln des korrekten Verhaltens bewusst ignoriert werden („*instances in which rules of correct behaviour are consciously ignored*“ (Dean et al. 2002)). Die WHO definiert *violation* als „*a deliberate deviation from an operating procedure, standard or rules*“ (WHO 2009 S. 145). Auch der Sachverständigenrat erwähnt die *violation* in Abgrenzung zum Behandlungsfehler: „Abzugrenzen sind darüber hinaus strafrechtlich relevante Tatbestände wie vorsätzliche oder missbräuchliche Schädigung oder Gewaltanwendung“ (SVR 2003, Nr. 366).

Für die praktische Arbeit ist die Unterscheidung von Irrtum auf der einen Seite und Patzer bzw. Versehen auf der anderen Seite recht relevant. Man „befindet sich im Irrtum“, diese Redewendung deutet bereits an, dass man sich nicht bewusst ist, einen fehlerhaften Plan zu verfolgen, und man kann *cum grano salis* davon ausgehen, dass Irrtümer schwerwiegendere Folgen haben (können) als Patzer bzw. das Versehen. Der Grund liegt darin, dass die beiden letzteren bei der fehlerhaften Anwendung des richtigen Plans auftreten, man ist sich auch jederzeit bewusst, was man tut bzw. tun sollte, und die Korrektur gelingt leichter. Ein Irrtum kann dagegen ein recht stabiler Zustand sein, der lange durchgehalten wird (z.B. Fehlerkette bei der Seitenverwechslung), und bei dem selbst direkte Hinweise auf die Fehlerhaftigkeit des Vorgehens keine Beachtung mehr finden. Die Definitionen lauten folgendermaßen:

- **Irrtum** (*mistake*): „*In a mistake, the action proceeds as planned but fails to achieve its intended outcome because the planned action was wrong*“ (IOM 1999, S. 46);
- **Patzer** (*slips*): „*error which result from some failure in the execution and/or storage stage of an action sequence*“ (WHO 2009, S. 140);
- **Versehen** (*lapse*): „*failure of memory or attention*“ (Dean et al. 2002), auch dem Versagen der Gedächtnisfunktion wird große Bedeutung zugemessen (WHO 2009, S. 122).

Das IOM bezeichnet sowohl den Patzer als auch das Versehen als „*error of execution* (IOM 1999, S. 46) bzw. als „*skill-based*“ (IOM 1999, S. 140, s. auch Leape 1994), stellt jedoch als Unterschied zwischen beiden Formen heraus, dass der Patzer einer direkten Beobachtung zugänglich ist, während dies für das Versehen als einer mentalen Leistung nicht gilt (S. 46). Dieser kognitionswissenschaftliche Zugang verwendet in seiner nur wenig differierenden Systematik für den Irrtum den Begriff *rule-based* (falscher Plan) und führt ergänzend die *knowledge-based errors* ein (in einer aktiven Problemlösung werden falsche Schlüsse gezogen) (s. Kap. 2.4.5.):

Eine weitere Unterscheidung bezieht sich darauf, ob der Fehler in der aktiv durchgeführten Behandlung (*error of commission*) und als Nicht-Behandlung auftritt (*errors of omission*): „*Events can be the result of human, technological, and systems errors and can be classified as errors of commission (a direct consequence of treatment)*

or errors of omission (failure to undertake an action that should have been completed”) (IOM 2012, S. 15). Diese Unterscheidung war bereits sehr früh ein Thema, in der Australian Health Care Study (Wilson 1995) waren 50% der UE durch Unterlassungen bedingt. James (2013) geht noch weiter und unterscheidet in seinem Systematischen Review insgesamt fünf Untergruppen des Fehlers:

- *errors of commission,*
- *errors of omission,*
- *errors of communication,*
- *errors of context,*
- *diagnostic errors.*

Im Zusammenspiel zwischen Fehler und UE gibt es grundsätzlich zwei Möglichkeiten: der Fehler kann zu einem UE führen oder nicht. Im ersten Fall handelt es sich um ein **Vermeidbares Unerwünschtes Ereignis** (VUE; *preventable adverse event PAE*). Dieser Begriff steht im Mittelpunkt eines auf Prävention ausgerichteten Risiko- und Qualitätsmanagements, da hier durch Analyse der Fehlerkette und durch Vermeidung von Fehlern die betreffenden Unerwünschten Ereignisse zu vermeiden wären: „*an adverse event attributable to error is a ‘preventable adverse event’*” (IOM 1999, S. 24). Das Kriterium der „Vermeidbarkeit“ ist *in praxi* allerdings nicht leicht zu operationalisieren (s. Kap. 3.5.). Dennoch ist es für epidemiologische Fragestellungen unumgänglich, eine genaue Spezifizierung vorzunehmen, z.B. danach, ob ein Fehler im allgemeinen *health care management*, im Befolgen der *accepted practice* und/oder auf individueller oder Systemebene vorliegt. Eine andere Herangehensweise unterscheidet Fehler aufseiten der behandelnden Ärzte, des Teams und des Systems (Barnes et al. 2006).

Die WHO stellt klar, dass die Vermeidbarkeit einen starken Kontext-Bezug aufweist: „*Preventable is being accepted by the community as avoidable in the particular set of circumstances*” (WHO 2009, S. 17). Diesem Umstand kommt große Bedeutung insbesondere dann zu, wenn es sich um UE im Gefolge von unterlassenen Behandlungen (*errors of omission*) oder von Behandlungen mit fehlendem (negativem) Nutzen handelt. James (2013) definiert ein *preventable adverse event* entsprechend folgendermaßen:

*“The broadest definition encompasses all unexpected and harmful experience that a patient encounters as a result of being in the care of a medical professional or system because high quality, evidence based medical care was not delivered during hospitalization.”*

VUE, die auf unterlassene oder nicht dem Stand der Erkenntnis nach den Kriterien der *Evidence-based Medicine* Behandlungen, auf diagnostische Fehler, Kommunikationsfehler oder Ursachen in der Systemumgebung zurückgehen, sind weitaus schwerer zu erfassen als *errors of commission* (James 2013).

Im zeitlichen Verlauf können VUE unmittelbar nach dem Eintreten des Fehlers (z.B. Blutung nach Antikoagulantien-Überdosierung), nach einem mittelfristigen Intervall (z.B.



Auftreten einer nosokomial übertragenen Hepatitis C) oder langfristig auftreten (z.B. Pneumokokken-Sepsis nach unterlassener Impfung nach Splenektomie) (James 2013).

Im zweiten Fall – wenn der Fehler nicht zu einem UE führt – spricht man von einem **Beinaheschaden** (*near miss*): „A near miss is an act of commission or omission that could have harmed the patient but did not cause harm as a result of chance, prevention, or mitigation“ (IOM 2004, S. 34). Den umgangssprachlichen Begriff des Beinahefehlers sollte man vermeiden, weil es sich um einen Fehler handelt, nämlich einen „Fehler ohne konsekutives Auftreten eines unerwünschten Ereignisses“ (SVR 2007, Nr. 599). Der Begriff des Beinaheschadens ist ein wichtiger Baustein des Fehlerketten-Konzeptes und grundlegend für die Funktion von *Reporting*-Systemen (z.B. CIRS, s. IOM 1999, S. 74). Da noch keine unerwünschten Ereignisse aufgetreten sind und sie im übrigen häufiger als stattgehabte UE sind, können Beinaheschäden besonders gut zu Zwecken der Analyse und Prävention eingesetzt werden. Soweit man dem linearen Modell folgen will, weisen sie hinsichtlich des Auftretens von UE eine Indikator-ähnliche Monitoringfunktion auf (s. Kap. 3.3.3.3.). Dessen ungeachtet besteht in der Definition auch im englischsprachigen Bereich eine gewisse Varianz (Übersicht bei WHO 2009, S. 128). Der Sachverständigenrat hat bereits 2003 eine gut verständliche Definition vorgeschlagen: „Ein Behandlungsfehler bleibt ohne negative gesundheitliche Auswirkungen aufgrund glücklicher Umstände, durch rechtzeitiges, korrigierendes Eingreifen und/oder „überholende Kausalität“ (SVR 2003 Nr. 367).

Wenn ein UE nicht nur auf einen Fehler, sondern auf mangelnde Sorgfalt zurückgeht, liegt ein **negligent adverse event** vor, ein Vermeidbares Unerwünschtes Ereignis, das die Kriterien der Sorgfaltsverletzung erfüllt. Der juristische Begriff des Behandlungsfehlers muss insofern abgrenzt werden, weil dort auch eine juristische Klärung des Ursachenzusammenhangs vorgenommen worden ist (s.u.). Wie in Kap. 1.2.4. ausgeführt, wurde in den frühen Studien in den USA auf breiter Front versucht, *negligent adverse events* mit epidemiologischen Methoden zu erfassen (*California Medical Insurance Feasibility Study* (Mills 1978), *Harvard Medical Practice Study II* (Leape et al. 1991), *Harvard Medical Practice Study III* (Localio et al. 1991), *Utah-Colorado Studie* (Thomas et al. 2000)). Diese Zielrichtung warf erhebliche Probleme in der Erhebungsmethodik auf, denn es musste nicht nur das Vorliegen eines UE festgestellt werden, sondern auch eine Wertung hinsichtlich des Vorliegens eines Fehlers (Vermeidbarkeit) und darüber hinaus einer Sorgfaltsverletzung vorgenommen werden. Man versuchte sich mit einem Begriff zu helfen, der dem des Medizinischen Standards ähnelt: „We defined a negligence as care that fell **below the standard expected of physicians in their community**“ (Brennan et al. 1991, Hervorh. MS). In der Veröffentlichung der Harvard-Studie heißt es „error is not the same as negligence. In tort law, medical negligence is defined as **failure to meet the standard of practice** of an average qualified physician practicing in the specialty in

question. Negligence occurs not merely when there is error, **but when the degree of error exceeds an accepted norm**. The presence of error is a necessary but not sufficient condition for the determination of negligence” (Leape et al. 1991, Hervorh. MS). In der *Harvard Medical Practice Study III* (Localio et al. 1991) heißt es: “Negligence was considered to have occurred if the medical care that caused the adverse event was **below the expected level of performance** of the average practitioner who treated problems such as the patient’s at that time” (Hervorh. MS). Das IOM schloss sich dieser Definition später an: “Negligent adverse events represent a subset of preventable adverse events that satisfy legal criteria used in determining negligence (i.e., whether the care provided **failed to meet the standard of care reasonably expected of an average physician qualified to take care of the patient in question**)” (IOM 1999, S. 24 (Hervorh. MS), zu einer Übersicht der Definitionsansätze s. WHO 2009 S. 128). Es wird diese Thematik deswegen hier so ausführlich dargestellt, weil das *negligent adverse event* in der „Geburtsstunde“ der Patientensicherheitsdebatte in den 90er Jahren in den USA wegen der dort herrschenden Versicherungskrise eine so große Rolle spielte.

Der **Behandlungsfehler** ist kein primär epidemiologischer, sondern ein juristischer Begriff, der das Verschulden, die Verletzung der Sorgfalt und die Zurechenbarkeit zum Gegenstand hat. „Ein gerichtlich oder außergerichtlich nach Einholung eines medizinischen Gutachtens anerkannter Behandlungsfehler, der auf einem nachweisbaren Fehler (Sorgfaltspflichtverletzung im Vergleich zum medizinisch anerkannten Standard) beruht, der nachweislich (Kausalität) einen nachweisbaren Behandlungsschaden (Schadensnachweis) verursacht hat und dem bzw. den beklagten Leistungserbringern oder der beklagten Gesundheitseinrichtung zuzurechnen ist (Zurechenbarkeit)“ (SVR 2003 Nr. 367, Tab. 11). Die Sorgfaltspflichten richten sich nach dem jeweiligen bei zumutbaren Anstrengungen zugänglichen objektiven Stand der medizinischen Wissenschaft zum Zeitpunkt der Behandlung. Der Begriff des Kunstfehlers sollte nicht mehr Verwendung finden, weil er die Ausübung einer nicht regelgebundenen „Kunst“ impliziert, während die Gesundheitsversorgung demgegenüber bei aller Individualität der Beziehung zwischen Patienten und Behandlern aller Berufsgruppen durchaus von Regeln geprägt und bestimmt ist.

Abschließend soll noch auf zwei Begriffe eingegangen werden, die in epidemiologischen Untersuchungen keine sehr große Rolle spielen, jedoch im Zusammenhang mit *Reporting*-Instrumenten und anderen Analysemethoden von Bedeutung sind: das Ereignis (*event, incident*) und das kritische Ereignis (*critical incident*). Der Begriff **Ereignis** ist sehr weit gefasst und umfasst sowohl Risiken und Beinaheschäden, die noch nicht zu einem negativen unerwünschten Ereignis geführt haben, als auch stattgehabte UE. Ein Ereignis stellt einen Zwischenfall, einen Prozess, eine Vorgehensweisen oder ein Ergebnis dar, das das Risiko für das Eintreten eines unerwünschten Ereignissen erhöht oder tatsächlich

in ein unerwünschtes Ereignis mündet, und auch das Versagen von Präventionsmaßnahmen mit einschließt (modif. n. WHO 2009, S. 16; Davies et al. 2003, S. 43; ÄZQ 2005). Die WHO versteht unter einem *incident* „... a reportable circumstance, near miss, no harm incident or harmful incident (adverse event)“ (WHO 2009, S. 16). Im *Canadian Patient Safety Dictionary* wird vorgeschlagen *“That incident be defined as including events, processes, practices, or outcomes that are noteworthy by virtue of the hazards they create for, or the harms they cause, patients. Incident reporting systems are meant to capture any and all incidents that are worthy of reporting. They often fail to do so because practitioners do not know what they should report”*. Das ÄZQ verwendet den Begriff *“Zwischenfall”*: *“Ein Ereignis (Incident) im Rahmen einer Heilbehandlung, welches zu einer unbeabsichtigten und /oder unnötigen Schädigung einer Person oder zu einem Verlust hätte führen können oder geführt hat“* (ÄZQ 2005).

Das **Kritische Ereignis** (*critical Incident*) grenzt sich nur quantitativ vom Ereignis (*incident*) ab, indem es schwerwiegende UE betrifft (zur Schweregradeinteilung s.o.). Wie der Begriff Ereignis bezieht sich das kritische Ereignis sowohl auf Risiken als auch auf die konkrete Verwirklichung von schwerwiegenden UE. Das *Canadian Patient Safety Dictionary* empfiehlt daher als Definition: *„That critical incident be defined as an incident resulting in serious harm (loss of life, limb, or vital organ) to the patient, or the significant risk thereof. Incidents are considered critical when there is an evident need for immediate investigation and response. The investigation is designed to identify contributing factors and the response includes actions to reduce the likelihood of recurrence”* (Davies et al. 2003, S. 43). Der Begriff des *critical event* differiert kaum vom Begriff des *sentinel event*, das jedoch zum Begriffsumfeld des Indikatorenkonzepts gehört. Beiden Termini ist der hohe Schweregrad des (potentiell eintretenden) Ereignisses und die hohe Dringlichkeit sofortiger Präventions- und Gegenmaßnahmen gemeinsam. In der praktischen Arbeit ist zu beachten, dass im Gegensatz zum Wortlaut der Bezeichnung *Critical Incident Reporting System*, das schwerwiegende Ereignisse impliziert, in einem CIRS in erster Linie Beinaheschäden und weniger schwerwiegende UE registriert werden, da diese häufiger, weniger sanktionsgefährdet und eher einer Prävention zugänglich sind.

Im Zusammenhang mit der Arzneimitteltherapiesicherheit (AMTS) wird der Begriff des Unerwünschten Ereignisses auf die Arzneimittelanwendung bezogen: Ein **Unerwünschtes Arzneimittelereignis** (UAE) „ist ein schädliches Ereignis, das in einem zeitlichen Zusammenhang mit einer Arzneimittelanwendung auftritt“ (Möller und Aly 2012, Aly 2015). Eine **Unerwünschte Arzneimittelreaktion (UAR)** entspricht dem älteren Begriff der „Nebenwirkung“: „Nebenwirkungen sind bei Arzneimitteln, die zur Anwendung beim Menschen bestimmt sind, schädliche und unbeabsichtigte Reaktionen auf das Arzneimittel. Dabei ist zwischen Nebenwirkungen

> bei bestimmungsgemäßigem Gebrauch,

- > in Folge eines Medikationsfehlers und
- > in Folge von Missbrauch oder beruflicher Exposition

zu unterscheiden“ (Aly 2015). Analog zum Begriff des Fehlers gilt als **Medikationsfehler** „ein Abweichen von dem für den Patienten optimalen Medikationsprozess, das zu einer grundsätzlich vermeidbaren Schädigung des Patienten führt oder führen könnte“ (Aly 2015). Es besteht folglich eine Kongruenz mit den allgemeinen Begrifflichkeiten.

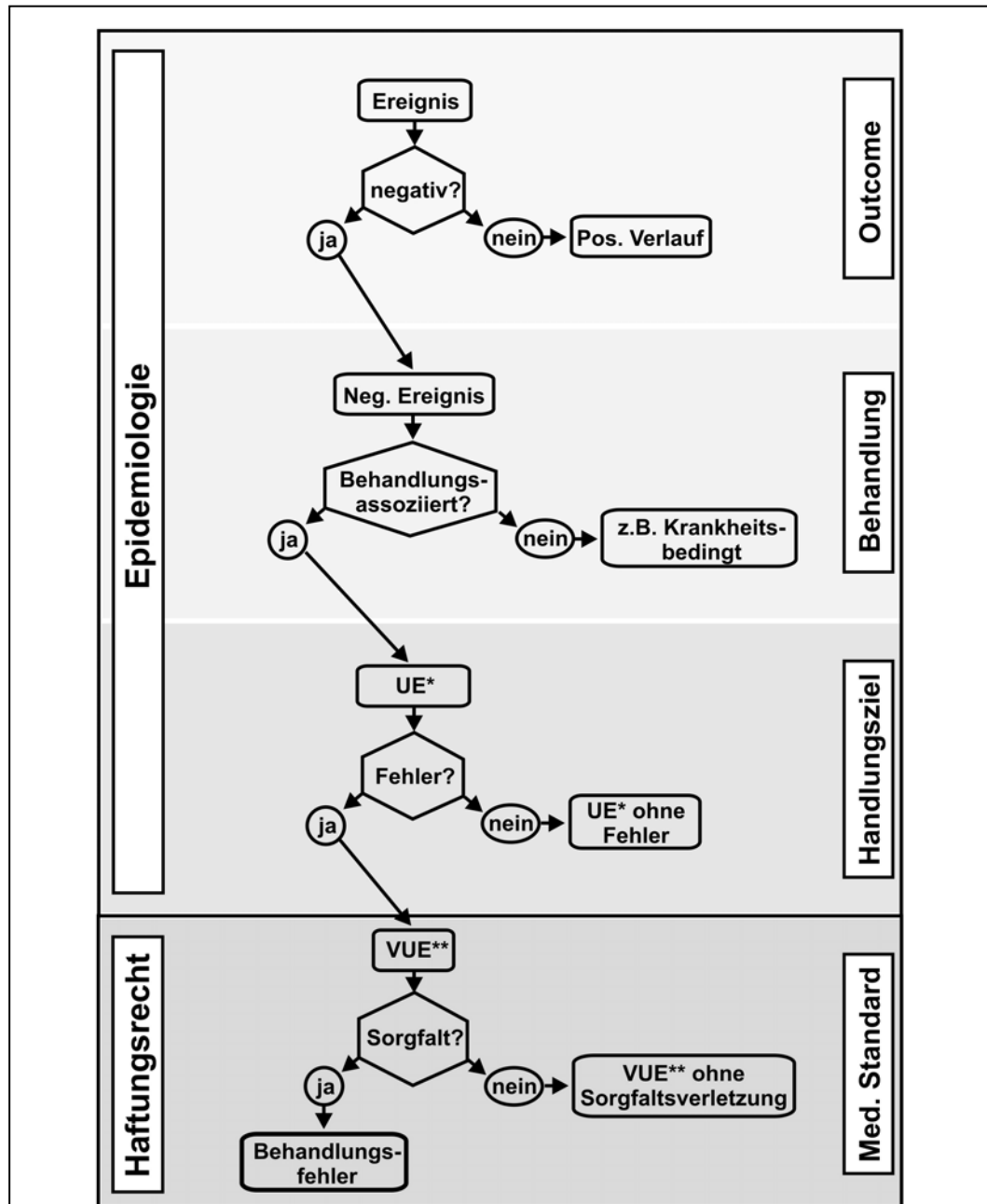


Abb. 3.2.-1: Nomenklatur zur Epidemiologie von Patientensicherheits-relevanten Ereignissen. Schematische Darstellung auf der Basis eines linearen bzw. normativen Prozess-Modells (s. Kap. 2.4.4.). Juristische Begriffe wie Behandlungsfehler sind abzugrenzen. \*UE = Unerwünschtes Ereignis, \*\*VUE = Vermeidbares Unerwünschtes Ereignis

### **3.3. Die kritische Rolle der Erhebungsmethodik**

#### **3.3.1. Einführung: vier Schritte**

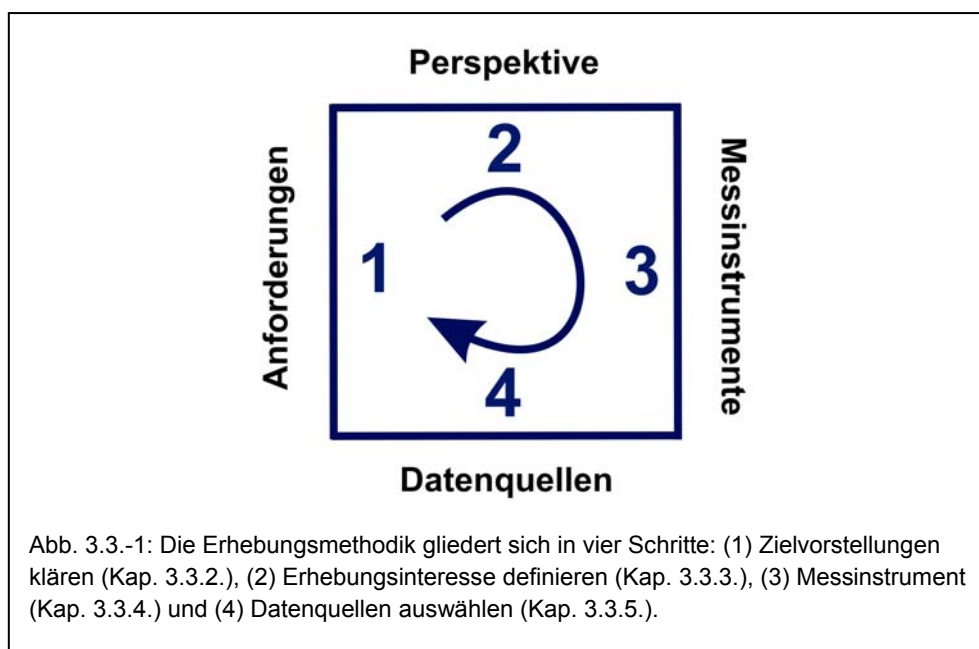
Immer wenn man sich anschickt, ein Konstrukt quantitativ zu fassen, kommen dessen potentielle Schwachpunkte zum Vorschein. Normalerweise wird angenommen, dass bei Messvorgängen im Bereich Patientensicherheit die Bestimmung und Abgrenzung von Zielen und eine Klärung des Erkenntnisinteresses nicht vorgenommen werden müssen, denn es geht um die Erfassung von Sicherheit und damit ist die Richtung klar. Allerdings wurde in den vorangehenden Kapiteln ausführlich auf die Kontextbedingtheit des Begriffs Patientensicherheit eingegangen (Kap. 2.3., s. auch Kap. 3.1.) und daraus die Konsequenz abgeleitet, dass hier sehr wohl eine Zielbestimmung am Anfang stehen muss: Patientensicherheit ist kein absoluter, kategorialer Begriff, obwohl es den Anschein hat. Es spielt zudem eine erhebliche Rolle, ob die Erfassung z.B. als wissenschaftliches Projekt, als Maßnahme des Risikomanagements oder als breit angelegtes Monitoring-Projekt initiiert wird.

Die **Zielorientierung**, die auch einschließt, dass man zur Kommunikation der Ziele eine vermittelbare Systematik vorhält, wird bei Messvorgängen im Bereich der Patientensicherheit meist unterschlagen. Der Erhebungsgegenstand, z.B. die Zahl Unerwünschter Ereignisse, erscheint als zu selbsterklärend, so dass man gar nicht reflektiert, ob diese Zahl eigentlich den Grad der Sicherheit wiedergeben kann. Eine Systematik, die es erlaubt, Ziele auszuwählen, darzustellen und gegeneinander abzuwägen, wurde in Kap. 2.3.3. als dreidimensionales Konstrukt vorgeschlagen, das sich an dem entsprechenden Orientierungsrahmen aus „Qualität 2030“ orientiert (Schrappe 2015, s. 68ff). Die drei Dimensionen umfassen die Perspektivdimension, die Strukturdimension und die Bedarfsdimension (vgl. Kap. 3.3.2.). Die **Perspektivdimension** erscheint besonders aktuell, da man bis heute meist die Anbieter-bezogene Perspektive einnimmt (z.B. nosokomiale Infektionen in einer stationären Einrichtung) und es immer noch große Defizite in der Darstellung von Sicherheit aus Patientenperspektive, regionaler oder Populationsperspektive gibt. Aber auch die **Strukturdimension** hat eine große Bedeutung, denn Messwerte richten sich einerseits an strukturellen Gegebenheiten aus (z.B. an einer sektoral orientierten Versorgung) und implizieren allein schon durch die Erhebung strukturell differenzierte Effekte (z.B. die Optimierung des sektoralen Sektors bei Eingriffs-bezogenen Ergebnisindikatoren). Die Morbiditäts- oder **Bedarfsdimension** ist demgegenüber fast selbsterklärend, denn die meist auf operative Akuterkrankungen ausgerichtete Auswahl der Unerwünschten Ereignisse, die erfasst werden sollen, spiegelt das reale Krankheitsspektrum nur unvollständig wieder (z.B. fehlen die chronischen,

konservativ zu behandelnden Erkrankungen und Mehrfacherkrankungen). Im Übrigen fehlt völlig der Präventionsbezug, der ebenso zu dieser Dimension gehört.

Die Zielorientierung stellt jedoch nur den ersten, allerdings ausgesprochen wichtigen Schritt der Planung einer Erhebung zur Patientensicherheit dar. Insgesamt besteht dieser Prozess aus **vier Schritten** (s. Abb. 3.3.-1), die in einer festgelegten Reihenfolge abgearbeitet werden müssen (s. ausführlicher in Kap. 5.3.):

- Zielorientierung,
- Erkenntnisinteresse,
- Messinstrumente und
- Datenquelle.



Die Klärung des **Erkenntnisinteresses** (s. Kap. 3.3.3.) ist vielleicht der wichtigste Vorbereitungsschritt. Vier Herangehensweisen stehen zur Auswahl:

- die **wissenschaftliche Evaluation** mit ihren umfassenden Modellannahmen, Pilotierungen und einer sorgfältigen Interpretation der Ergebnisse (unter Rückgriff auf das Modell),
- die **klinisch-epidemiologische Untersuchung**, die Ereignisse unter Verwendung von Falldefinitionen bei möglichst hohen, im Alltag aber realisierbaren Werten bezüglich Sensitivität und Spezifität,
- das **Monitoring durch Indikatoren**, die unter Verzicht auf hohe eine Spezifität sehr sensitiv eingestellt sind (damit man kein Ereignis übersieht), und die zur Beobachtung größeren Versorgungsbereiche geeignet sind, und

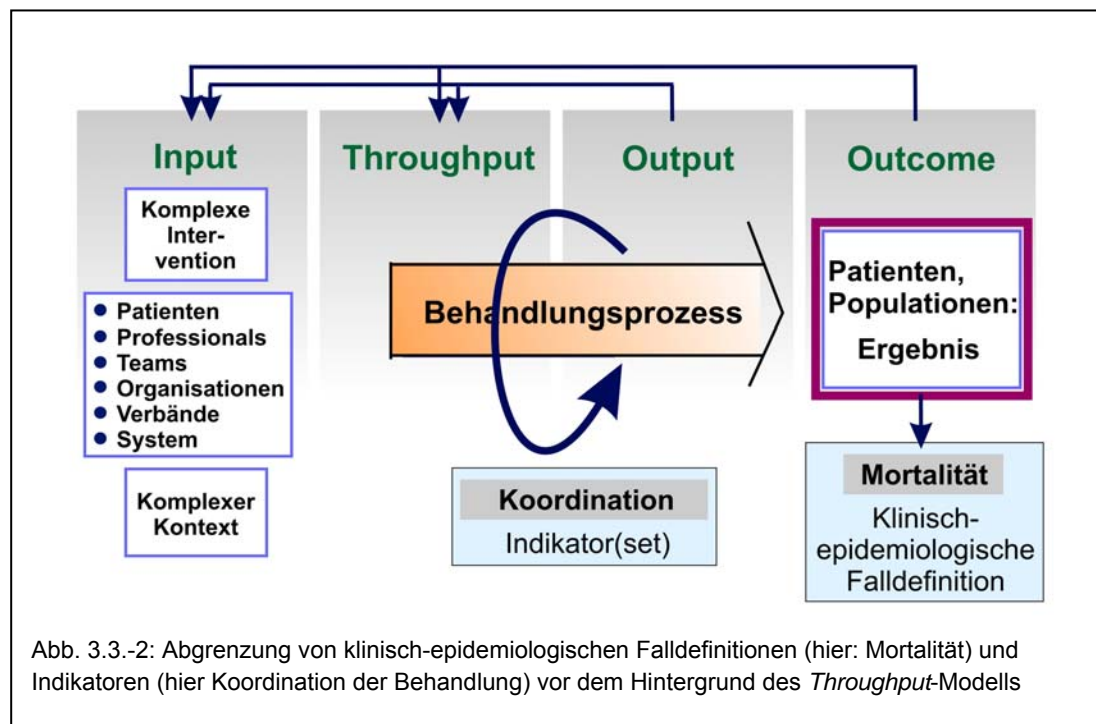
- die **generierenden Verfahren**, die im Sinne des *detection loop* (s. Abb. 3.3.-3) vorher unbekannte Informationen erfassen, deren Fehlen ebenso wenig bekannt war (sog. doppelt blinde Flecken, *Unknown Unknowns*). Diese Daten können damit zum Gegenstand nachfolgender Untersuchungen gemacht werden (z.B. CIRS, Unterlagen aus dem Beschwerdemanagement) oder als Aufgreifkriterien für Folgeuntersuchungen dienen (z.B. Registerstudien mit nicht vollständigem Nenner, nicht sensitive Abrechnungsdaten etc.).

Wichtig ist es sich zu vergegenwärtigen, dass jeder dieser Zugangswege deutlich differierende statistische Eigenschaften der Erfassungsmethodik bedingt. Eine wissenschaftliche Evaluation kann alle erdenklichen Anstrengungen unternehmen, um eine möglichst hohe Sensitivität und gleichzeitig Spezifität zu verwirklichen, während die generierenden Verfahren, sozusagen als anderes Extrem, nur das „Aufgreifen“ als Ziel haben und auf messbare Sensitivitäts- bzw. Spezifitätswerte verzichten.

Ein sehr wichtiger Punkt beim Thema Patientensicherheit besteht in der großen praktischen Bedeutung der **klinisch-epidemiologischen Untersuchungen**. So basieren die großen Erfassungen zu (Vermeidbaren) Unerwünschten Ereignissen im Grunde auf klinisch-epidemiologischen Falldefinitionen, sei es im institutionellen Rahmen, sei es z.B. in den großen Untersuchungen nach dem *Harvard Medical Practice Study-Design* auf nationaler Ebene (s. Kap. 3.3.4.4.). Das große Beispiel und zugleich Vorbild ist die Erfassung nosokomialer Infektionen, die mit ihrer Methodik bis in die 70er Jahre des letzten Jahrhunderts zurückreicht. In der Konsequenz aus jahrzehntelanger Optimierung dieser Methodik verfügt man heute über Falldefinitionen, die nicht nur über Institutionen, sondern auch über Landesgrenzen hinweg vergleichbare Daten erbringen. Es bleibt eines der ungelösten Probleme und gleichzeitig eine der großen Perspektiven der weiteren Entwicklung (s. Kap. 5.3.3.), dass über die Erstellung von klinischen Falldefinitionen auch in anderen Bereichen der Gesundheitsversorgung (z.B. postoperative Thrombosen, Unerwünschte Arzneimittelereignisse etc.) ähnliche klinisch und zugleich epidemiologisch relevante Daten erhoben werden (s. z.B. MPSMS, Kap. 3.3.3.3.).

Die Abgrenzung zu den **Patientensicherheitsindikatoren** ist besonders wichtig, denn diese haben (entgegen der oft geäußerten Meinung) nicht das Ziel, Unerwünschte Ereignisse quantitativ zu erfassen, sondern sie gehen der Aufgabe nach, ohne einen allzu großen Aufwand (aber mit hoher Sensitivität) in größeren Versorgungsbereichen analog einer Warneinrichtung Situationen zu erkennen, in denen das Auftreten von Unerwünschten Ereignissen über einer definierten Wahrscheinlichkeitsschwelle liegt. Sprechen sie an, sollten die wirklich vorliegenden Unerwünschten Ereignisse vollständig erfasst sein, auch wenn es hierdurch zur falsch-positiven Erfassung von nicht-auffälligen Ereignissen kommt (die geringe Spezifität wird durch Nachuntersuchungen ausgeglichen). Bei den meisten sog. Patientensicherheitsindikatoren-Sets (PSI-Sets) handelt es sich im

Grunde um Zusammenstellungen aus klinisch-epidemiologischen Falldefinitionen und nicht um validierte Indikatoren, die Sicherheitsprobleme vorhersagen (z.B. PSI90-Set der AHRQ, s. Kap. 3.3.3.3.3.).



In Abb. 3.3.-2. wird das Verhältnis von klinisch-epidemiologischen Falldefinitionen und Indikatoren nochmals vor dem Hintergrund des *Throughput*-Modells dargestellt: klinisch-epidemiologische Falldefinitionen stellen den Ist-Zustand dar (meist anhand von *Outcome*-Parametern, hier: Mortalität), während Indikatoren in ihrer Ampelfunktion ungünstige Verläufe vorhersagen (z.B. mittelsangaben zur Koordination). Der Behandlungsprozess wird also eher durch Indikatoren beschrieben, die im optimalen Fall die Unerwünschten Ereignisse vorherzusagen in der Lage sind.

Nach Klärung des Erkenntnisinteresses erfolgen zwei weniger kritische Schritte, zum einen die Auswahl der **Messinstrumente**, die von der Frage der Datenquelle abzugrenzen ist (s. Kap. 3.3.4.). Zu unterscheiden sind direkte Beobachtung, Befragungen, Erhebungen zur Patientensicherheitskultur, Trigger-gestützte Erhebungen und *Big Data*-Analysen. Jeweils sind hier Prävalenz, Periodenprävalenz und Inzidenz zu unterscheiden, die eine unterschiedliche Aussagekraft aufweisen, und außerdem der institutionelle und überinstitutionelle Ansatz.



Zuletzt wird die (oder mehrere) am besten geeignete **Datenquelle** und die Art der Daten ausgewählt, die die angestrebten Zwecke am besten erfüllen (s. Kap. 3.3.5.). Das Spektrum reicht vom externen *Chart Review* über Trigger-gestützte Aktenuntersuchungen, über den intern generierten Report bis zu Versicherungsdaten, Haftpflichtversicherungsdaten und Abrechnungs- sowie Sozialdaten, außerdem sind je nach Aufgabenstellung Meldungen aus Fehlermeldesystemen und anderen generierenden Verfahren (z.B. Patientenbeschwerden) heranzuziehen. Die Analyse von unstrukturierten Daten wie z.B. bei *Big Data* stellt eine weitere Option dar. Hinsichtlich der Datenart ist die Entscheidung zu treffen, ob man eher Struktur-, Prozess- oder Ergebnisindikatoren verwenden sollte und ob eine Risikoadjustierung notwendig ist.

Im weiteren sind natürlich einige allgemeine Aspekte hinzuzufügen, die jedoch aus thematischen Gründen Publikationen zur Qualitätsthematik vorbehalten bleiben müssen. Spezifisch für Patientensicherheit sind die Aspekte geringe Häufigkeit, geringe Effektmaße bei Interventionen etc., wie sie in Kap. 2.5.2. dargestellt sind. Gerade die Seltenheit stellt eine Herausforderung dar, denn ein Risiko von 1:10.000 heißt nichts anderes als: 9999mal geht alles gut, und trotzdem muss man die Prävention fortführen und weiter evaluieren, eine enorme Anforderung an die Aufmerksamkeit (Hollnagel 2014 S. 47).

### **3.3.2. Zielorientierung**

Die Zielorientierung stellt den ersten, sehr wichtigen Schritt der Planung einer Erhebung zur Patientensicherheit dar. Die nachfolgenden Schritte (Zielorientierung, Erkenntnisinteresse, Messinstrumente und Datenquelle) sollen nicht geringgeschätzt werden, doch ein fehlender Zielbezug führt zwangsläufig zu großen „blinden Flecken“ in dem Wissen über Patientensicherheit (Derose und Petitti 2003, Smith und Busse 2010). Die Beispiele sind Legion: man misst den Stand der Patientensicherheit auf dem Gebiet der operativen Akutmedizin, so gut es die verwendete Methodik auch hergibt, und trotzdem kann man, wie in den vorangegangenen Kapiteln schon mehrfach angemerkt, mit diesem Ansatz die eigentlichen Probleme vollständig übersehen. Zum Beispiel kann es möglich sein, dass

- diese Probleme gar nicht im Bereich der operativen Akutmedizin liegen (sondern z.B. in der Behandlung von chronisch Kranken),
- diese nicht berücksichtigen, ob indizierte Behandlungen unterlassen worden sind (*errors of omission*), so dass sie gar nicht in den Zahlen auftauchen (eine

unterlassene Behandlung kann keine behandlungsbedingten Unerwünschten Ereignisse zur Folge haben),

- die akutmedizinische Maßnahmen, die zwar ohne Unerwünschtes Ereignis durchgeführt werden, auf eine erfolglose oder unterlassene Prävention (z.B. Impfung) zurückgehen, und dass

- die Behandlungen zwar ohne Unerwünschte Ereignisse abgelaufen sind, es aber nicht indizierte Behandlungen waren (Übersorgung), so dass die (wenngleich schadensfreien) Behandlungen selbst ein Unerwünschtes Ereignis darstellen.

Selbst wenn diese Fälle ausgeschlossen werden können, könnte es fünftens immer noch möglich sein, dass

- die Behandlungen in Einrichtungen mit einem privilegierten Patientenaufkommen mit günstigem Risikoprofil durchgeführt werden, während (wir nehmen ein Beispiel aus den USA) im benachbarten *safety net*-Krankenhaus (Redlener und Grant 2009), das es sich zur Aufgabe gemacht hat, auch den sozial benachteiligten Patienten zur Verfügung zu stehen, wegen des ungünstigeren Risikoprofils der dortigen Patienten eine höhere Zahl von Unerwünschten Ereignissen auftreten (Risikoselektion).

Es wäre in diesem geschilderten Fall also dringend geboten gewesen, sich vorher Gedanken zu machen, ob die akutmedizinisch-operative Ausrichtung eigentlich der Problemstellung entspricht, die man als vordringlich ansieht. Dies hätte auch die Möglichkeit eröffnet, sich über die Problematik des Morbiditätsspektrums, unterlassener Maßnahmen und Prävention, der Übersorgung oder Risikoselektion zu orientieren.

In Kap. 2.3. wurde weiterhin darauf hingewiesen, dass die geforderte Zielorientierung für das Thema Patientensicherheit eine wichtige Funktion in der Kommunikation und Allianzbildung hat, und zwar sowohl im institutionellen Bereich als auch auf Systemebene. Auch Ziele, die Sicherheit adressieren, brauchen Unterstützung und müssen in das „Gesamtverhalten“ der jeweiligen Ebene integriert werden. So rekuriert das Thema Sicherheitskultur, dies hier als (sehr wichtiges) Beispiel, immer wieder auf die Gesamtheit der Werte und Normen, die Sicherheit angehen.

Die Formulierung und Kommunikation von Zielen ist jedoch einfacher gesagt als getan. Es bedarf dafür ein von den Adressaten geteiltes Verständnis, auf das man sich gemeinsam beziehen kann, und das eine Priorisierung auch konkurrierender Ziele ermöglicht. In dem Fall, dass ein solches „Koordinatensystem“ nicht vorliegt, muss man in der Lage sein, ein solches System aktiv in die Diskussion zu bringen. Dies ist im Bereich Patientensicherheit (und Qualität) nicht so einfach, denn dieses Konstrukt ist nur in einem interdependent-mehrdimensionalen Zusammenhang zu verstehen. Patientensicherheit kann man nicht, das wurde in Kap. 2. ja klar, „über einen Kamm scheren“, sondern es gibt verschiedene Aspekte, die auf ganz unterschiedlichen Voraussetzungen aufbauen. Eine dieser Dimensionen ist der Bedarf, was bedeutet, dass sich die Beschäftigung mit

Patientensicherheit auf einen Bereich der Gesundheitsversorgung beziehen sollte, in dem ein Schwerpunkt des medizinisch-pflegerischen oder demographischen Handlungsbedarfs zu finden ist. Würde man Patientensicherheit dagegen in einem Zusammenhang thematisieren, der in der Versorgung keine aktuelle Rolle spielt, setzt man sich dem Vorwurf der Irrelevanz aus.

Um die Priorisierung von Anforderungen, die Zielformulierung und der Evaluation der Zielerreichung zu erleichtern, ist daher ein mehrdimensionales Koordinatensystem notwendig. Das hier vorgeschlagene System (s. Kap. 2.3.3.) wurde ursprünglich als Qualitätsmodell entwickelt, kommt hier aber in einer modifizierten Form für das Thema Patientensicherheit zum Einsatz (s. Abb. 2.3.-2, vgl. Schrappe 2015, S. 68ff) und beinhaltet drei Dimensionen:

1. Perspektivdimension
2. Strukturdimension
3. Bedarfsdimension

Die **Perspektivdimension** beschreibt ähnlich wie in der Gesundheitsökonomie den Horizont der Betrachtung. Es werden sieben Perspektiven unterschieden (s. Abb. 2.3.-2):

- Wissenschaft,
- Gesellschaft,
- Region/Population,
- Nutzen,
- Patienten,
- Institution bzw. Anbieter und
- Professionen bzw. Berufsgruppen.

Je nach Perspektive ergibt die Analyse völlig andere Ergebnisse: was aus institutioneller Sicht wichtig erscheint (z.B. Senkung der Komplikationen), ist aus Patientensicht zwar auf den ersten Blick auch interessant, aber für es ist vielleicht noch wichtiger, den langfristigen Therapieerfolg zu kennen oder zu verbessern (z.B. die Fähigkeit, sich schmerzfrei fortzubewegen). Man muss genau hinsehen: die wissenschaftliche Sicht muss nicht mit der gesellschaftlichen Sicht übereinstimmen, gleiches gilt für die gesellschaftliche Sicht in Abgleich mit der regionalen oder Populations-Perspektive. Gleichzeitig ist es nicht sinnvoll, bestimmte Perspektiven grundsätzlich unbeachtet zu lassen, z.B. muss man sich um die Akzeptanz der Berufsgruppen bemühen.

Die **Strukturdimension** bezieht sich auf die Weiterentwicklung (oder die Hemmung) der wichtigsten strukturellen Systemprobleme. Das klassische Beispiel ist die Behinderung des Integrationsgedankens durch sektorale oder anbieter-bezogene Initiativen, die im Grunde einer sektoralen Optimierung dienen und den notwendigen Integrationsprozess nicht nur nicht fördern, sondern ihn sogar hemmen können.

Letztlich ist die **Bedarfsdimension** eine *conditio sine qua non*, wenngleich sie doch immer wieder übergangen wird. Wie in Kap. 5.2.4. noch ausführlich zu schildern sein wird, beschäftigt man sich in Deutschland recht intensiv mit Komplikationen im operativen Akutbereich, jedoch nur wenig mit der Medikationssicherheit bei chronisch mehrfacherkrankten, älteren Patienten; weitere Bereiche, die noch nicht in den Blick genommen worden sind, wurden anfangs dieses Kapitels genannt. Ganz wichtig: Hierzu gehören auch Aspekte der Prävention.

Diese drei Dimensionen erlauben es, ein Projekt oder einen Ansatz zur Verbesserung der Patientensicherheit einzuordnen und zu kommunizieren, aus welcher Perspektive er an das Problem herangeht, welche strukturellen Konsequenzen daraus abzuleiten sein werden, und welche Bedeutung er hinsichtlich des Morbiditätsspektrums aufweist (oder gar der Prävention dient). Es gibt allerdings einen immer wiederkehrenden Ausspruch, den man vielerorts und oft hört, der die Thematik gut beschreibt: statt eine Diskussion darüber zu führen, wo Patientensicherheit eigentlich wirklich prioritär erhoben und verbessert werden sollte: „*fangen wir doch schon mal damit an, mit den Daten, die wir haben*“. Dieses ist der häufig gehörte Einstieg in eine Maßnahme, die ohne eine kritische Zielorientierung viele Ressourcen verbraucht und Hoffnungen nährt, ohne das man eigentlich geklärt hat, welches Problem man eigentlich lösen möchte.

Es gibt allerdings prominente Ausnahmen. So wird in dem methodisch sehr fortgeschrittenen *Medicare Patient Safety Monitoring System* (MPSMS) in den USA, mit dem Daten zu Unerwünschten Ereignissen im Krankenhausbereich erhoben wird, an ganz prominenter Stelle die Frage gestellt: „*For what purpose will the information be used?*“. Sehr wichtig, hier geht es rein um eine quantitative Erhebung (s. Kap. 3.3.3.2.): „*... the role of the MPSMS is only the measurement of adverse events. The project has no capacity to initiate interventions or affect change in the topics observed* (Hunt et al. 2005). Das ist eine deutliche Aussage, die auch schon das Erkenntnisinteresse thematisiert.

### **3.3.3. Das Erkenntnisinteresse der Erhebung**

Abzugrenzen von der Perspektivdimension (s.o.) ist das Erkenntnisinteresse einer Erhebung: was will man eigentlich wissen? Will man wissenschaftlich evaluieren, will man zählen, will man beobachten (Monitoring), will man „blinde Flecken“ beheben – all das sind unterschiedliche Ansatzpunkte. Diese Ansatzpunkte gilt es, bei der Planung und bei der Auswertung fein säuberlich auseinanderzuhalten, vor allem da die statistischen Anforderungen sehr unterschiedlich sind.

Das enorme Interesse an der Thematik Patientensicherheit hat zu einer sehr großen Anzahl von Studien zum Thema der Größenordnung des Problems geführt (s. Kap. 3.6.). In der Folge hat das Interesse an der verwendeten Methodik stark zugenommen, und zwar zusätzlich zu der im vorangehenden Kapitel diskutierten Frage der verwendeten Perspektive. In diesem Zusammenhang ist die erhebliche Innovationskraft hervorzuheben, die dabei von der Patientensicherheitsforschung und –praxis ausgeht, denn sie hat zwei methodische Ansätze wieder verstärkt in den Mittelpunkt der Aufmerksamkeit gerückt, die gerade auch im benachbarten Bereich der Qualitätsforschung in Vergessenheit geraten waren: die Verwendung von Erhebungen mit klinisch-epidemiologischen Falldefinitionen (und ihren Weiterentwicklungen wie Trigger-gestützte Verfahren) und die zentrale Rolle der sog. generierenden Verfahren (Schrappe 2015, S. 80). Unter den letztgenannten Verfahren sind die *Critical Incident Reporting Systems* (CIRS) am prominentesten, die in der Lage sind, „zweifach blinde Flecken“ bzw. „*Unknown Unknowns*“ erkennbar und in der Folge auch messbar zu machen. Neben den analytisch ausgerichteten Methoden zur Fehlerketten- und Ursachenanalyse darf dieser Umstand als eines der Hauptverdienste der Diskussion zur Patientensicherheit gelten.

In den nachfolgenden Abschnitten werden die vier Perspektiven des Erkenntnisinteresses gefolgt von den wichtigsten Messinstrumenten und Datenquellen kurz dargestellt (s. Abb. 3.3.-3, Tab. 3.3.-1), in Einzelnen

- komplexe Messvorgänge zur wissenschaftlichen Evaluation („*scientific pathway*“).
- Messung mittels klinisch-epidemiologischer Falldefinitionen („*epidemiological pathway*“),
- Monitoring durch Indikatoren („*monitoring pathway*“) und
- generierende Verfahren (sog. „*detection loop*“).

Die genannten vier Erkenntnisperspektiven verfolgen grundsätzlich andere Ziele. Am deutlichsten sind die generierenden Verfahren abzugrenzen (*detection loop*), die über Ereignisse informieren, die vorher unbekannt waren, und deren Nicht-Existenz ebenfalls nicht bekannt waren (*unknown unknowns*). Sie sind nicht zur Quantifizierung geeignet (Sensitivität und Spezifität nicht über 1%) und umfassen auch Verfahren, die mit sog. Aufgreifkriterien arbeiten (z.B. Obduktion, Abrechnungsdaten bei unvollständiger Dokumentation etc.) (s. Kap. 3.3.3.4.). Wissenschaftliche Verfahren (*scientific pathway*) erreichen ein Maximum an (gleichzeitiger) Sensitivität und Spezifität, sind aber wegen ihres Aufwandes nur für Untersuchungen im kleineren Rahmen geeignet. Dies ist z.B. der Fall bei der Evaluation von komplexen Mehrfachinterventionen (*Complex Multicomponent Interventions*, CMCI, s. Kap. 5.7.), bei denen man über die Wahl der wissenschaftlichen Methodik besonders gründlich Rechenschaft ablegen muss (s. Kap. 3.3.3.1.). Klinisch-epidemiologische Falldefinitionen (*epidemiological pathway*) stehen im Zusammenhang mit Patientensicherheit ganz im Mittelpunkt, der Großteil der epidemiologischen

Untersuchungen zur Häufigkeit von UE und Unterformen sowie zahlreiche im großen Maßstab durchgeführte Interventionsstudien sind hier einzuordnen (Kap. 3.3.3.2.). Hierzu gehören auch die Trigger-gestützten Verfahren, zu deren Grundtyp das *Harvard Medical Practice Study* (HMPS) – Design zu rechnen ist. Diese Methoden stehen zwar einem umfangreicheren Einsatz offen, erfordern aber dennoch ein spezialisiertes Vorgehen und entsprechende Ressourcen. Dagegen sind Patientensicherheits-Indikatoren (PSI) für einen breiten Einsatz geeignet (s. Kap. 3.3.3.3.), denn sie beschränken sich auf das Monitoring (*monitoring pathway*). Sie müssen eine hohe Sensitivität aufweisen, denn man will keine Ereignisse übersehen, jedoch muss man Kompromisse bei der Spezifität machen (sog. Intrinsic Ungerechtigkeit, s. Schrappe 2015, S. 89f).

	<b>Erkenntnisinteresse</b>	<b>Zielereignis</b>	<b>Statistik</b>	<b>Anmerkung</b>
<i>Scientific Pathway</i>	Wissenschaftliche Evaluation unter Modellannahmen	Wirkung komplexer Interventionen	Sensitivität und Spezifität maximal	Aufwendig, aber für „doppelte Komplexität“ geeignet
<i>Epidemiological Pathway</i>	Quantifizierung definierter Ereignisse	<i>Safety</i> : Unerwünschtes Ereignis (UE)	Sensitivität und Spezifität hoch	Für Routine geeignet, aber nicht ohne Aufwand
<i>Monitoring Pathway</i>	Monitoring und Risikobewertung	Vorhersage von UE	Sensitivität hoch, Spezifität sekundär	Monitoring großer Versorgungsbereiche, Ampelfunktion
<i>Detection loop</i>	Identifikation von <i>Unknown Unknowns</i>	Unbekannt	Sensitivität und Spezifität <1%	Exploration, nicht zur Quantifizierung geeignet

Tab. 3.3.-1: Abgrenzung der unterschiedlichen Perspektiven des Erkenntnisinteresses (modif. n. Schrappe 2015 S. 81), vgl. Text und Abb. 3.3.-3.

Für den Stand der deutschen Diskussion darf als charakteristisch gelten, dass klinisch-epidemiologische Methoden und Indikatoren nicht sauber voneinander getrennt werden. Es werden zahlreiche, sehr wichtige Daten auf der Basis von Falldefinitionen dokumentiert (z.B. Komplikationsraten), aber es ist oft nicht nachvollziehbar, ob diese Daten nun wirklich etwas über Sicherheit aussagen (s. Ausführungen in Kap. 2.2.) oder Sicherheit für einen Versorgungsbereich vorhersagen (die zentrale Aufgabe von Indikatoren). Hierauf muss ausführlich eingegangen werden, um ein Verständnis vom

Konzept der Patientensicherheitsindikatoren (PSI) zu gewinnen, das mit einem weiterentwickelten Konzept von Patientensicherheit in Deckung zu bringen ist (s. Kap. 3.3.3.3.). Gleichmaßen ist zu diskutieren, ob zum Monitoring bestimmte Indikatoren zur Evaluation wissenschaftlicher Fragestellungen verwendet werden können, was wegen der geringen Spezifität regelmäßig zu Problemen führt. Wenn in der Literatur immer wieder auf die Schwierigkeit hingewiesen wird, die Erfolge von Interventionen zu Verbesserung der Patientensicherheit nachzuweisen, dann kann eine der möglichen Antworten in den hier kurz dargestellten methodischen Problemen liegen.

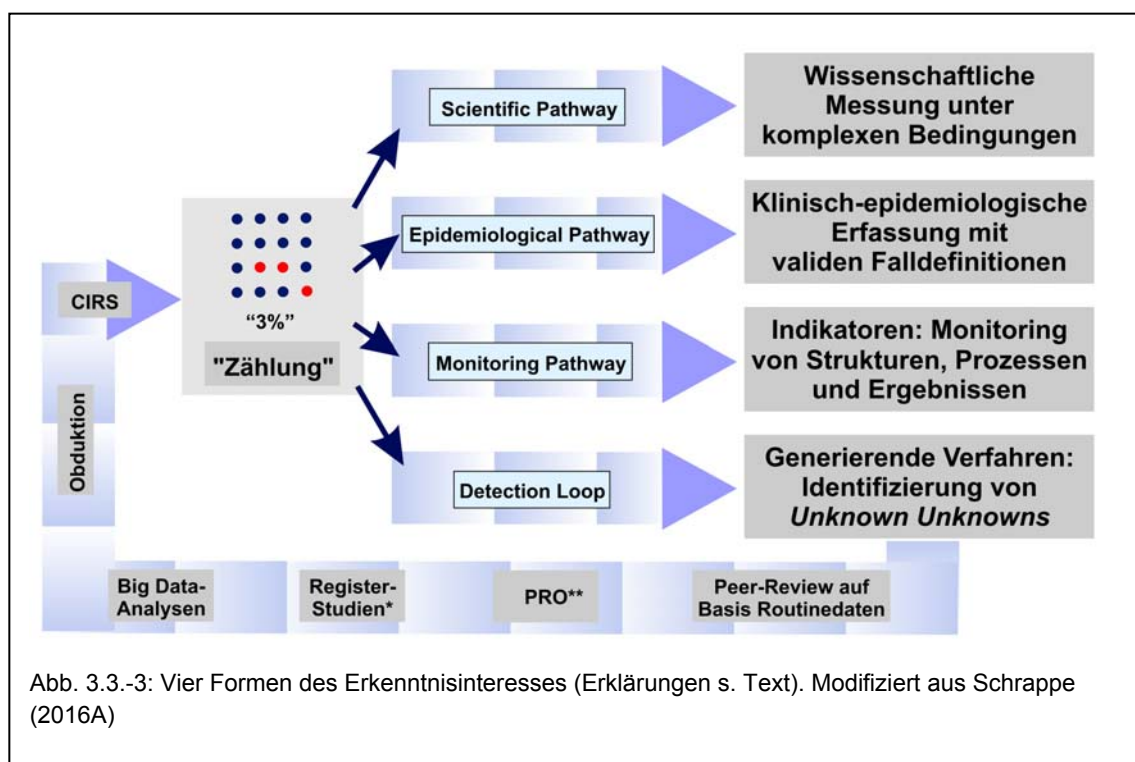


Abb. 3.3.-3: Vier Formen des Erkenntnisinteresses (Erklärungen s. Text). Modifiziert aus Schrappe (2016A)

Ein weiterer Punkt betrifft die Frage, ob Erhebungen im Bereich der Patientensicherheit als lineare Messvorgänge betrachtet werden können. Dieser Punkt reicht noch weiter in das konzeptionelle Verständnis von Patientensicherheit hinein, denn wie bereits ausführlich in Kap. 2.3. dargestellt, legt es der dort diskutierte normative Charakter der Definition von Patientensicherheit nahe, diese „monochromatisch“ über das Vorhandensein bzw. die Abwesenheit von Unerwünschten Ereignissen zu beschreiben. Dieser Ansatz ist verständlich, denn auf den ersten Blick erscheint der Aushandlungsprozess über die „inhärenten Merkmale“ und die „Anforderungen“, die regelhaft der Quantifizierung von Qualität vorausgehen müssen, beim Thema Patientensicherheit nicht notwendig zu sein (Pronovost et al. 2007). Zu Beginn einer Patientensicherheitsbewegung wie Anfang des letzten Jahrzehnts in Deutschland brachte ein solches lineares Vorgehen auch große Vorteile hinsichtlich der Durchsetzung

und Akzeptanz des Patientensicherheitsgedankens mit sich, denn man war rasch mit klaren Zahlen sprechfähig und konnte die Priorität des Themas unter Beweis stellen. Auf lange Sicht stößt dieses Verständnis jedoch aus mehreren Gründen an seine Grenzen, denn

- eine lineare Quantifizierung lässt die Komplexität des Messvorganges selbst außer Acht und kann positive oder negative Einflüsse der Messung selbst auf den Untersuchungsgegenstand übersehen;
- Patientensicherheit ist als ein soziales Konstrukt aufzufassen ist, das sich nur in Annäherungen linear abbilden lässt und deshalb einer präexperimentellen Theoriebildung bedarf;
- die angenommene Linearität führt zu der Erwartung, dass Maßnahmen zur Verbesserung der Patientensicherheit rasch und direkt (also linear) zu nachweisbaren Verbesserungen führen, was fast nie der Fall ist.

Der Grund für den letzten Punkt liegt bei seltenen Ereignissen wie Patientenverwechslungen auf statistischem Gebiet (kleine Fallzahlen), bei häufigeren Ereignissen bei der multifaktoriellen Entstehung und zahlreichen gegenläufigen Entwicklungen (höhere Vigilanz, bessere Dokumentation mit Erhöhung der Raten, paradoxe Effekte etc.).

Dies gibt genau die Situation derzeit in Deutschland wieder. Die Konsequenz kann jedoch nicht im ausschließlichen Einsatz qualitativer Methoden liegen (so wichtig diese sind), so wie es z.B. von Vincent et al. (2014) es mit ihren fünf Fragen implizit vorgeschlagen wird (s. Kasten). Stattdessen sollten durchaus quantitativen Methoden eingesetzt werden, die zu verwertbaren quantitativen Maßen führen, mit denen man Patientensicherheit darstellen, vergleichen und bei Interventionen eine Verbesserung nachweisen kann. Es ist dabei jedoch zu beachten, ob man sich mit einem Monitoring-Ansatz mit Indikatoren oder einer klinisch-epidemiologischen Erhebung begnügen kann oder man zu einer wissenschaftlichen Untersuchung greifen muss. Letztere die Messung bzw. Erfassung von Merkmalen, die die Patientensicherheit beschreiben, in ein Untersuchungskonzept oder Modell fassen, das vor der Untersuchung erstellt wird und z.B. Angaben darüber enthält, wie die Tatsache, dass der Erfassungsvorgang selbst als komplexe Interventionen aufzufassen ist. Auf dieses Thema wird im nachfolgenden Kapitel genauer eingegangen.



- ▶ Has patient care been safe in the past?  
We need to assess rates of past harm to patients, both physical and psychological.
  - ▶ Are our clinical systems and processes reliable?  
This is the reliability of safety critical processes and systems but also the capacity of the staff to follow safety critical procedures.
  - ▶ Is care safe today?  
This is the information and capacity to monitor safety on an hourly or daily basis. We refer to this as 'sensitivity to operations'.
  - ▶ Will care be safe in the future?  
This refers to the ability to anticipate, and be prepared for, problems and threats to safety
  - ▶ Are we responding and improving?  
The capacity of an organisation to detect, analyse, integrate, respond and improve from, safety information
- (Vincent et al. 2014)

### 3.3.3.1. Wissenschaftliche Perspektive (*scientific pathway*)

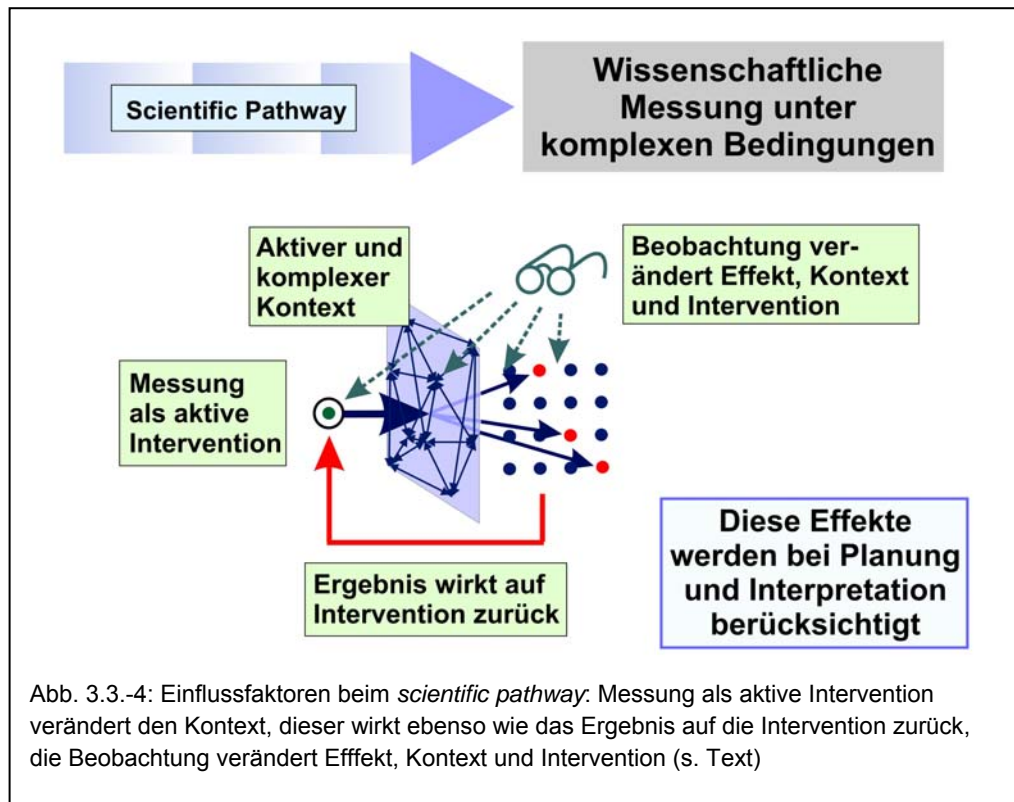
Die wissenschaftliche Perspektive (*scientific pathway*) wird in der Qualitäts- und Patientensicherheitsforschung eingenommen, bedient sich der Konzepte der Versorgungsforschung und ist vor Kurzem im entsprechenden Memorandum des Deutschen Netzwerk Versorgungsforschung ausführlich dargestellt worden (Geraedts et al. 2017). Ähnlich wie die Versorgungsforschung stellt die Patientensicherheitsforschung die beiden zentralen Postulate der biomedizinisch-klinischen Forschung in Frage, die darin bestehen, dass zwischen Untersucher (Subjekt) und Untersuchungsgegenstand (Objekt) eine scharfe Trennung hergestellt werden kann und alle Umgebungsbedingungen zu kontrollieren sind (s. Schrappe und Pfaff 2017, S. 7ff). Diese Annahmen, die in bestimmten Bereichen der klinischen Forschung von großer Wichtigkeit sind, weil nur so auf der Ebene der absoluten Wirksamkeit (*efficacy*) Behandlungsmethoden auf ihren Nutzen überprüft werden können (randomisierter Versuch), um anschließend unter Alltagsbedingungen ihre relative Wirksamkeit (*effectiveness*) in der Routinebehandlung unter Beweis stellen zu können, sind im Bereich der Patientensicherheitsforschung nicht in jedem Fall zu realisieren.

Diese Einschränkungen haben mehrere Gründe. Zum Einen befindet sich gerade die Thematik Patientensicherheit in einem hochgradig durch soziale Konstrukte beeinflussten Kontext, ja stellt nach der Ansicht vieler Autoren selbst ein soziales Konstrukt dar (Pawson und Tilley 2004). Zum Anderen löst allein die Tatsache der Messung von Ereignissen eine deutliche Veränderung des Verhaltens und der Bedeutung dieser

Ereignisse aus, die jeder Patientensicherheitspraktiker aus eigener Anschauung kennt – allein die Ankündigung einer Untersuchung verändert alles. Weiterhin kann man nicht davon ausgehen, dass während der Dauer der Untersuchung der Kontext, das Verhalten und die äußeren Umstände konstant bleiben, denn die Ergebnisse und die im Laufe der Untersuchung sich ändernde Bedeutung der Untersuchung wirken auf den Kontext und die Messvorgänge zurück. Dies gilt schon für die Erhebung von *prima vista* rein quantitativ anmutenden Indikatoren wie Komplikationsraten oder Prozessparametern, aber erst recht für Evaluationen höhergradiger Konstrukte wie z.B. der Sicherheitskultur. Linearen Ansätzen, die Patientensicherheitskultur sozusagen „direkt“ messen und z.B. longitudinal vergleichen, ist daher mit Vorsicht zu begegnen, insbesondere der Rückwirkung auf den Kontext der Messung und auf das Sicherheitsverhalten ist größte Aufmerksamkeit zu schenken.

Aus diesen Überlegungen folgt die Einsicht, dass jeder Messvorgang selbst als komplexe Intervention zu betrachten ist, auch ohne dass es sich hier um die Evaluation einer komplexen Intervention (z.B. Einführung einer Checkliste) handeln muss. Unter den vier hier geschilderten Formen des Erkenntnisinteresses, die mit diesem Umstand unterschiedlich umgehen, stellt der wissenschaftliche Zugang insofern eine Besonderheit dar, als hier versucht wird, im Rahmen eines durchaus aufwendigen Untersuchungsaufbaus die genannten Einflussfaktoren in einer Theoriebildung in der Vorphase der Untersuchung zu antizipieren. Es darf hier nicht der Gesichtspunkt unterschlagen werden, dass es in der Wissenschaft eine große Kontroverse zu der Frage gibt, ob diese Herangehensweise sinnvoll und überhaupt möglich ist (s. Cohn et al. 2013, Leviton 2011). Folgende vier Aspekte müssen in der Planung, der Durchführung und natürlich auch Auswertung sowie Ergebnisinterpretation berücksichtigt werden (s. Abb. 3.3.-4):

- die Tatsache der Durchführung der Untersuchung verändert die Umgebungsfaktoren bzw. den Kontext, wobei berücksichtigt werden muss,
- dass der Kontext im Gesundheitswesen bereits ohne Durchführung einer Untersuchung als höchst komplexer Zustand begriffen werden muss;
- die Messung bzw. Beobachtung wird als aktive Intervention verstanden und hat einen Einfluss auf den Untersuchungsgegenstand (Greenhalgh und Swinglehurst 2011), der wegen der doppelten Komplexität (Schrappe und Pfaff 2017, S. 24f) von Intervention und Kontext nur schwer vorhersagbar ist, und
- das Ergebnis – soweit im Verlauf bekannt oder erahnt – wirkt auf den Untersuchungsgegenstand und den Kontext zurück, so dass während der Untersuchungsdauer das Setting nicht konstant bleibt.



Diese Zusammenhänge sind Gegenstand der Versorgungsforschung und der *implementation research*. Folgende Aspekte sind stichpunktartig hervorzuheben:

➔ **Intervention und Kontext sind gleichberechtigte Einflussfaktoren:** lineare Input-Output-Modelle sind nicht hinreichend. Um das Zusammenspiel von aktivem Kontext und komplexer Intervention adäquat zu beobachten, sind sog. „OXO-Modelle“ notwendig (*Observe – Perturbation – Observation*), bei denen in Abgrenzung zu „CMO-Modellen“ (*Context + Mechanism = Outcome*) nicht davon ausgegangen wird, dass die Intervention (z.B. Einführung einer Leitlinie) rein additiv zu einem stabilen Kontext hinzugefügt wird (Berwick 2008, Pawson und Tilley 2004). Zusammenfassend ergeben sich folglich fünf Untersuchungsdomänen (Damschroder et al. 2009):

- Intervention mit ihren *core elements* (die „peripheren Eigenschaften“ der Intervention können evtl. angepasst werden),
- äußerer Kontext (z.B. soziales oder politisches Umfeld),
- interner Kontext (z.B. organisatorische Struktur),
- beteiligte Individuen, insbesondere ihre Einstellungen und die geübte Praxis, und letztlich
- der Prozess der Implementierung (z.B. Schulungen).

Bei der Planung der Evaluation müssen also Kontext, individuelle Einstellungen und Implementierung mit einfließen, was durchaus erhebliche Auswirkungen auf Ressourcen und Zeitbedarf haben kann. Ganz entscheidend ist natürlich die Frage, auf welcher Ebene die Evaluation durchgeführt wird (Kontextebene), in erster Linie unterscheidet man

- Ebene des Gesundheitssystems
- institutionelle Ebene
- individuelle Ebene (Verhaltensänderung).

→ **Untersucher als *embedded researcher*** (Vindrola-Padros et al. 2017): Die Untersucher sind als zwar gewohnt, mit qualitativen, explorativen und generierenden Verfahren zu arbeiten, die auch keine klare Trennung von Untersuchungsobjekt und – subjekt aufweisen, trotzdem stehen Evaluationen von Interventionen zur Verbesserung von Patientensicherheit unter großem Druck, denn sie müssen gleichzeitig mit der Komplexität der Interventionen (es handelt sich meist um zusammengesetzte Mehrfachinterventionen) und der Komplexität eines über die Zeit nicht stabilen Kontextes umgehen können, der nicht nur durch die Intervention, sondern auch zusätzlich noch durch die Beobachtung selbst verändert wird. Diese Situation wird unter dem Begriff der Doppelten Komplexität zusammengefasst (s. Kap. 2.5.4.).

→ **Methodik und Untersuchungsgegenstand sind interdependent:** Die Wahl der adäquaten Methodik hängt also nicht nur von der untersuchten Intervention ab, sondern die Methodik ist selbst ein aktiver Bestandteil der Intervention, die Auswahl der untersuchten Intervention hängt sogar erheblich von der eingesetzten Methodik ab. So könnte z.B. ein ausschließlicher Einsatz der randomisierten Kontrollgruppenbildung (Methodik) dazu führen, dass im Endeffekt nur Wissen zu isolierten, biomedizinisch dominierten Interventionen zur Verfügung steht, weil komplexe Interventionen schlechter im randomisierten Design zu untersuchen sind bzw. dort geringere Effektmaße ergeben: „... *for the simple reason that these are the advances that have been studied. That is where the evidence is.*“ (Leape et al. 2002).

→ **Adäquate Theoriebildung ist unbedingte Voraussetzung:** In den Empfehlungen zur Methodik in der Versorgungsforschung wird grundsätzlich auf die große Bedeutung einer *a priori* stattfindenden Theoriebildung verwiesen, die den Untersuchungsgegenstand strukturiert, Interventionen und Kontext-Bedingungen darstellt sowie Aussagen über sinnvolle Endpunkte macht (MRC 2008). Dieses Vorgehen erscheint vor dem Hintergrund der genannten Wechselwirkungen von Kontext, Intervention und Methodik verständlich, allerdings fließt natürlich ein Vorwissen des Untersuchers in die Evaluation mit ein.

→ **Interventionen sind im Regelfall komplex:** Wie generell in der Versorgungsforschung ist auch hier davon auszugehen, dass die verwendeten Interventionen Kontext-sensibel sind und in wechselseitigem Verhältnis zur eingesetzten

Methodik stehen. Beide, Kontext und Intervention, sind komplex (sog. „doppelte Komplexität“, s Kap. 2.5.4.), es hat sich der Begriff der *complex multicomponent interventions* (CMCI's) eingebürgert (Berwick 2008, Guise et al. 2014B, Schrappe und Pfaff 2017A, S. 24-5). Die Frage, ob sich die CMCI wieder in ihre Einzelmaßnahmen zerlegen lässt, wird weiter in Kap. 5.7. diskutiert.

### 3.3.3.2. Klinisch-epidemiologische Erhebung

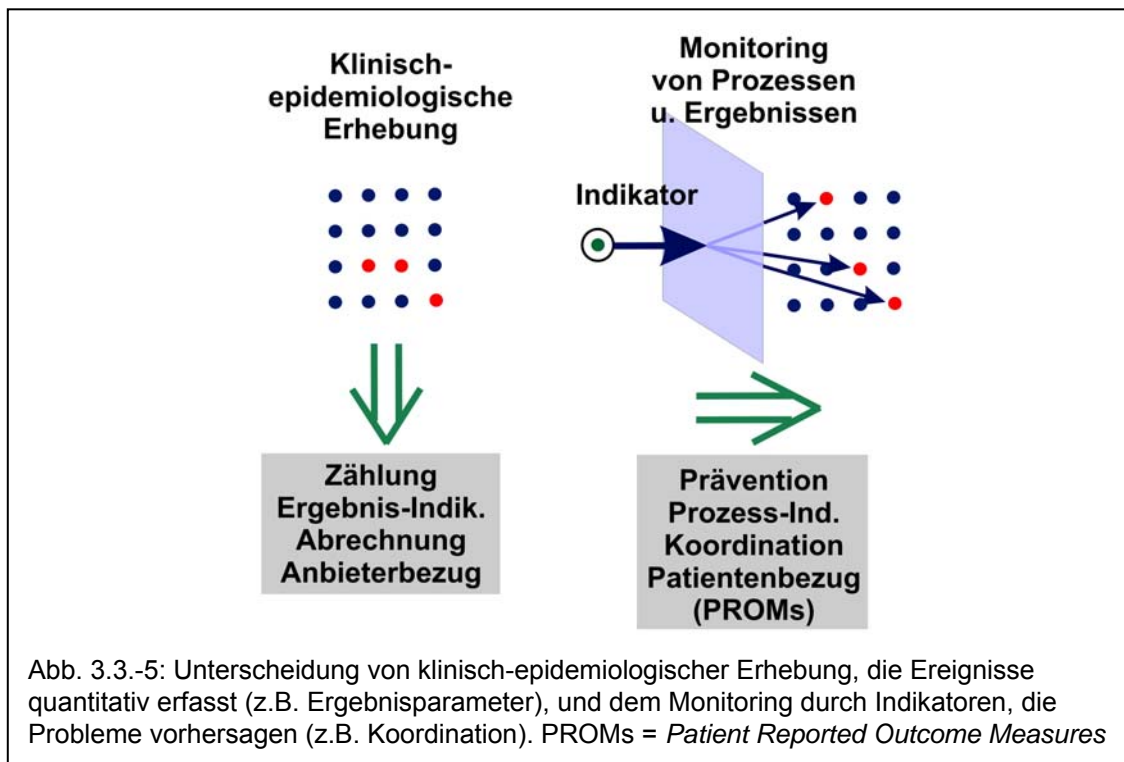
Das Thema Patientensicherheit hebt sich durch die große Bedeutung der klinisch-epidemiologischen Untersuchungsmethoden und der generierenden Verfahren (z.B. CIRS) von Erhebungen im Qualitätsbereich ab. Klinisch-epidemiologische Falldefinitionen wurden naturgemäß zunächst zur Erfassung infektiologischer Komplikationen (nosokomiale Infektionen) eingesetzt. In der Infektiologie sind seit den 80er Jahren des letzten Jahrhunderts epidemiologisch intendierte Definitionen entwickelt worden, die komplexe Sachverhalte stabil, Institutionen- und sogar Länder-übergreifend wiedergeben können (Gaynes und Solomon 1996). Wegen der spezifischen Charakteristika von nosokomialen Infektionen (s. Kasten) haben sich international eigene Berufe herausgebildet (*Infection Control*-Spezialisten, Krankenhaushygiene, Hygienefachkräfte etc.). Die Thematik (Erfassung, Entwicklung von darauf aufbauenden Steuerungsmechanismen) wurde so erfolgreich handgehabt, dass es teilweise sogar zu Abgrenzungen zu Maßnahmen im Qualitäts- und Patientensicherheitsbereich kam (Talbot et al. 2013). Klinisch-epidemiologische Falldefinitionen müssen nicht nur von Indikatoren, sondern auch von klinischen Diagnosen sauber abgegrenzt werden, denn die Falldefinitionen implizieren weder eine ärztliche Diagnose noch eine Therapieentscheidung (obwohl klinisch-

#### **Steckbrief: Nosokomiale Infektionen**

- 1** Nosokomiale Infektionen (NI) sind das einzige Qualitäts- bzw. Sicherheitsproblem, das übertragbar ist.
- 2** NI treten im Krankenhaus (alter Begriff: Krankenhausinfektionen), aber auch in der ambulanten Versorgung auf.
- 3** NI können drei epidemiologische Formen annehmen: sporadisch (punktuell), epidemisch (Indexperson) und endemisch (unbekanntes Erregerreservoir – die gefährlichste Form).
- 4** In Deutschland liegt die Prävalenz bei Krankenhauspatienten stabil zwischen 3 und 4% der Aufnahmen (ohne bei Aufnahme bereits bestehende NI).
- 5** NI sind eng verbunden mit dem Problem der Antibiotikaresistenz.
- 6** Die wichtigsten Manifestationen sind Harnwegsinfektionen, Pneumonie und postoperative Wundinfektionen, nosokomiale Sepsis und gastrointestinale Infektionen.
- 7** Ein wirksames *infection control*-Konzept umfasst die Übertragung (z.B. Barriere-methoden) und gleichzeitig die Standardisierung der Antibiotikatherapie (Ziel: Resistenzbildung).

epidemiologische Falldefinitionen gelegentlich Therapieentscheidungen als Fallkriterium verwenden, z.B. Antibiotikagabe als Kriterium für nosokomiale Infektionen).

In der Folge sind klinisch-epidemiologische Falldefinitionen auch für nicht-infektiologische Entitäten entwickelt worden (z.B. postoperative Thrombosen, Stürze etc.). Für die Erfassung von ganzen Gruppen von Unerwünschten Ereignissen wie z.B. im *Harvard Medical Practice Study Design* (s. Kap. 3.3.4.4.) werden Trigger-gestützte Verfahren entwickelt, die mehrstufig arbeiten. Allen diesen Definitionen ist zu eigen, dass sie auf bestimmte Messinstrumenten (z.B. Prävalenzuntersuchung) und Datenquellen (z.B. Krankenakten) anwendbare Kriterien vorhalten. Sie erfordern zwar nicht einen so hohen Messaufwand wie die wissenschaftlich intendierten Erfassungen (s. 3.3.3.1.), sind jedoch trotzdem auf speziell trainierte Personen angewiesen und bringen insofern einen Aufwand mit sich, der deutlich über der Verwendung von Indikatoren oder auch CIRS-Meldungen liegt. Auch hinsichtlich ihrer statistischen Anforderungen heben sie sich deutlich von den anderen Verfahren ab, denn bei klinisch-epidemiologischen Falldefinitionen muss auf ein ausgeglichenes Verhältnis einer relativ hohen Sensitivität und Spezifität geachtet werden. Wissenschaftliche Untersuchungen können wegen ihrer Modellbildung und dem besonderen Aufwand, der hier zugrunde gelegt werden kann, mit noch höheren Werten aufwarten, sind aber nicht für den Routinebetrieb geeignet. Indikatoren grenzen sich von Falldefinitionen durch ihre hohe Sensitivität und niedrige Spezifität ab (s. Kap. 3.3.3.3.).



Die mangelnde Differenzierung von klinisch-epidemiologischen Falldefinitionen und Indikatoren (seien es nun Qualitäts- oder Patientensicherheitsindikatoren, zur Unterscheidung s. auch Kap. 3.3.3.3.2.) stellt in Deutschland eines der größten Hindernisse in der Weiterentwicklung der Externen Qualitätssicherung nach §136 SGB V dar, denn die meisten dort erhobenen Parameter sind zwar brauchbare klinisch-epidemiologische Fallspezifikationen, die man auch kennen muss (z.B. Diagnose-bezogene Mortalität oder Komplikationen), weisen aber keine nachgewiesene Validität hinsichtlich der Vorhersage von Unerwünschten Ereignissen und Defiziten im Bereich Patientensicherheit auf (s. Schrappe 2015, S. 86ff). Zu dieser mangelnden Validität kommt noch eine mangelnde Reliabilität hinzu, zumindest soweit es sich um Abrechnungsdaten handelt (Jha und Pronovost 2016).

Im Einzelfall gibt es natürlich Sicherheits-relevante Parameter, die einerseits aus klinisch-epidemiologischer Intention erhoben werden (z.B. Katheterinfektionen auf Intensivstationen), die aber gleichzeitig auch gute Indikatoren z.B. für die Sicherheit der Versorgung in der Intensivmedizin darstellen. Für die Mortalität, dies als anderes Beispiel, gilt dies dagegen nicht: die Mortalität ist ein wichtiger Parameter, der erhoben und in einer Institution bekannt sein muss, er ist aber kein guter Patientensicherheits- bzw. Qualitätsindikator, weil er von sehr vielen anderen Faktoren beeinflusst wird (Smith et al. 2010, S. 9, Hogan et al. 2015) und in seiner Risiko-adjustierten Form Gegenstand von *gaming* bezüglich der adjustierungsrelevanten Nebendiagnosen ist (z.B. Gupta et al. 2017A).

Studien zur klinisch-epidemiologischen Erfassung von UE und Vermeidbaren UE sind zahlreich publiziert worden (Hanskamp-Sebregts et al. 2016). In einem der größten je durchgeführten Systematischen Reviews zu diesem Thema (Lessing et al. 2010) konnte kein Einfluss des Landes, in dem die Erhebung durchgeführt worden war, auf das Ergebnis festgestellt werden. Die Streuung, die die Ergebnisse der 241 eingeschlossenen Studien erwartungsgemäß zeigten, war in erster Linie durch die Zahl der eingeschlossenen Patienten erklärbar (weitere Informationen s. Kap. 3.6.1.).

Zum differenzierten Einsatz von klinisch-epidemiologischen Falldefinitionen und zum Weiterentwicklungsbedarf (z.B. Nutzung für Prozessparameter aus dem *Throughput* der Verwirklichung von Patientensicherheit) wird in Kapitel 5.3.3. Stellung genommen. Falls man sich z.B. zu einer verpflichtenden Meldung von bestimmten *sentinel events* entschließen sollte, muss man hier auch Falldefinitionen entwickeln.

### 3.3.3.3. Monitoring durch Indikatoren

#### 3.3.3.3.1. Einführung

Indikatoren sind hoch-sensitiv eingestellte Parameter, die unerwünschte Ereignisse vorhersagen (JCAHO 1991), sie sind nicht identisch mit Sicherheit oder Qualität (Sens et al. 2018). Die klassische Definition des Qualitätsindikators durch die *Joint Commission* sei hier wiederholt, da sie auch für das Verständnis von Patientensicherheitsindikatoren (PSI) eine große Bedeutung hat:

*„An indicator is a quantitative measure that can be used to monitor and evaluate the quality of important governance, management, clinical, and support functions that affect patient outcomes. An indicator is not a direct measure of quality. **Rather, it is a tool that can be used to assess performance that can direct attention to potential performance issues that may require more intense review within an organisation**“* (JCAHO 1991, s. auch Sens et al. 2018, Hervorh. MS).

Um im routinemäßigen Einsatz in großen Versorgungsbereichen eine aussagekräftige Sensitivität erreichen zu können (möglichst alle Unerwünschten Ereignisse sollen angezeigt werden), muss man Kompromisse in der Spezifität machen (falsch-positive Ergebnisse werden in der nachgehenden Analyse aufgedeckt). Die Validität von Indikatoren besteht in der Sensitivität bezüglich dieser Vorhersagefunktion und ist nicht zu verwechseln mit der Zuverlässigkeit des Messvorganges (Reliabilität) z.B. in der Erfassung einer Komplikationsrate (s. Abb. 3.3.-5). Diese Unterscheidung ist in ihrer zentralen Bedeutung wirklich nicht zu überschätzen, denn sie berührt den Kern der Arbeit mit Indikatoren. Will man mit Indikatoren „zählen“ (was eigentlich Aufgabe klinisch-epidemiologischer Falldefinitionen wäre), dann wäre (neben der Sensitivität) die Spezifität (bzw. der Positive Prädiktive Wert) sehr wichtig, denn man will falsch-positive Ergebnisse vermeiden, von all positiven Ergebnissen sollen möglichst viele richtig positiv sein. Will man aber eine Monitoring-Funktion einrichten (wozu Indikatoren in der Tat bestimmt sind), dann sollte man von allen tatsächlich vorliegenden Ereignissen keine übersehen, man darf also möglichst keine falsch-negativen Befunde tolerieren (hohe Sensitivität bzw. hoher Negativer Prädiktiver Wert).

Zwei häufige Missverständnisse:

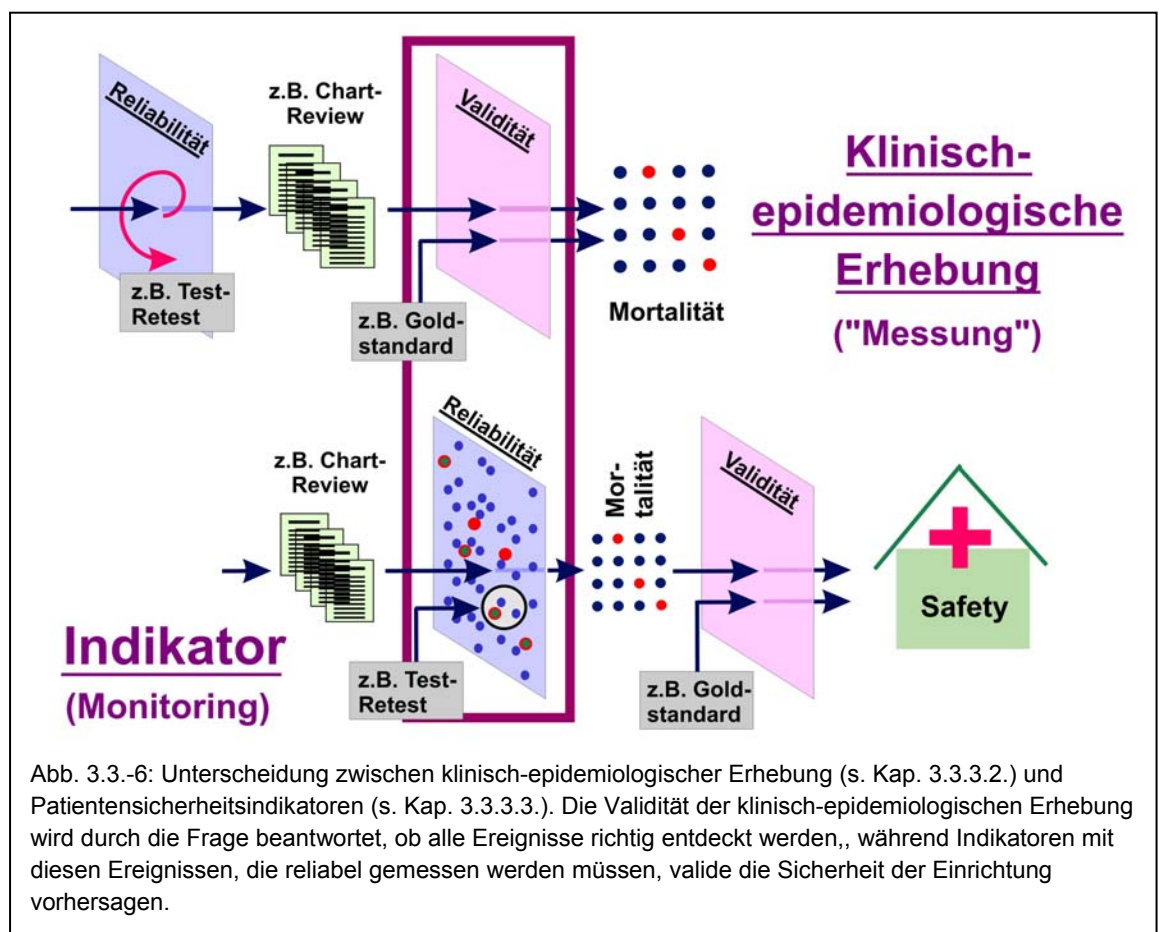
**Patientensicherheit und Klinische Diagnostik**

**1** Klinisch-epidemiologische Falldefinitionen (z.B. für nosokomiale Infektionen) sind nicht mit klinischen Diagnosen gleichzusetzen.

**2** Die klinische Diagnostik arbeitet mit Methoden hoher Spezifität und hohem Positivem Prädiktivem Wert, Indikatoren dagegen mit hoher Sensitivität und hohem Negativem Prädiktivem Wert.



Hier liegt auch eine der Gründe (s. Kasten), warum Indikatoren gerade im ärztlichen Bereich so schwer verständlich zu machen sind, denn Ärzte arbeiten in ihrer Diagnostik mit hoch-spezifischen Verfahren (der festgestellt Befund muss tatsächlich vorliegen (keine falsch-positiven Ergebnisse), weil sonst umsonst operiert wird) und haben Schwierigkeiten zu akzeptieren, dass bei der Arbeit mit Indikatoren das Vorliegen falsch-positiver Ergebnisse toleriert wird. Wie man allerdings in manchen Veröffentlichungen z.B. aus den USA erkennt, ist dieses Missverständnis weit verbreitet, so kritisieren Winters et al. (2016) in ihrem Systematischen Review den AHRQ-PSI-Set (s. Kap. 3.3.3.3.) aufgrund seines niedrigen Positiven Prädiktiven Wertes.



In der Konsequenz ergibt sich aus der beschriebenen Problematik auch ein unterschiedliches Verständnis der Validität (s. Abb. 3.3.-6.). Ein Indikator ist valide, wenn er alle Ereignisse erkennt oder vorhersagt. Die Reliabilität des Indikators beschreibt, ob das Indikatorereignis (das das Maß darstellt, das zum Monitoring eingesetzt wird) zuverlässig erkannt wird. Dies ist wichtig, denn ein nicht reliabler Indikator (z.B. nosokomiale Infektionen erfasst durch Abrechnungsdaten, s. Kap. 3.3.5.) kann nie valide

sein. Eine quantitative, klinische-epidemiologische Erfassung ist demgegenüber dann valide, wenn sie erfasst, was sie erfassen soll. Die Rate von Katheterinfektionen kann also einmal ein klinisch-epidemiologisch gemessenes Datum zum Status Quo sein, ein andermal ein Indikatorereignis, das Sicherheitsprobleme vorhersagt (z.B. in einer Intensivstation, s.o.).

Die *Joint Commission* (JCAHO) definiert die Validität von Indikatoren als „*the degree to which the indicator identifies events that merit further review*“ (JCAHO 1991, S., 12). Dies spielt gegenwärtig auch in der deutschen Diskussion eine große Rolle, da das Institut für Qualität und Transparenz im Gesundheitswesen (IQTIG) in seinem Methodenpapier (IQTIG 2017) daran festhält, den Begriff der Validität im Sinne der Reliabilität (Zuverlässigkeit) zu interpretieren (ausführliche Darstellung Schrappe 2017F, s. außerdem Kap. 5.3.1.):

„Die Datenvalidität beschreibt dabei, ob die zur Berechnung des Indikatorwerts verwendeten Daten den Vorgaben des Indikators hinsichtlich Vollzähligkeit, Vollständigkeit und Dateninhalten entsprechen oder ob es z. B. durch Erfassungs- oder Übertragungsfehler zu Unterschieden zwischen erhobenem Wert und dokumentiertem Wert gekommen ist“ (IQTIG 2017, S. 119, s. auch S. 120f).

Die Einführung in die Thematik der Indikatoren muss hier exemplarisch bleiben und wird in Kap. 5.3.4. wieder aufgenommen, die Systematik, statistische Einteilung und der Fokus können nur am Rande gestreift werden (Schrappe 2015, S. 86ff). Nachgeordnet den statistischen Grundgrößen (hohe Sensitivität, hoher NPW, niedrigere Spezifität) sollten Indikatoren mehrere Anforderungen erfüllen, zu nennen ist besonders die

- **Selektion:** ein Parameter wird ausgewählt aus mehreren Parametern nach seiner Sensitivität, Unerwünschte Ereignisse in einem ganzen Versorgungsbereich zu beobachten,
- **Abstraktion:** ein Indikator ist nicht mit Sicherheit bzw. Qualität identisch, sondern sagt Risiken voraus,
- **Zielgerichtetheit:** Indikatoren werden zielorientiert entwickelt und eingesetzt (s. Kap. 2.3.3.),
- **Machbarkeit:** technisch und messtechnisch umsetzbar, und natürlich die
- **Validität:** der Indikator sagt die Probleme mit hoher Sensitivität voraus, für die er entwickelt wurde (Schrappe 2010A).

Neben der sog. Rumba-Regel (**R**elevant, **U**nderstandable, **M**easureable, **B**ehaviourable, **A**chievable) und den Anforderungen der AHRQ bei der Entwicklung des *Healthcare Cost and Utilization Projects* (HCUP) (Davies et al. 2001) spielt in Deutschland das OECD-Konzept nach Arah et al. (2006) eine Rolle, das die Anforderungen *effectiveness, safety, access* und *patient-centeredness* unterscheidet.

Als Vorbereitung zu den nachfolgenden Kapiteln wird hier kurz auf die drei Aspekte eingegangen, den *Patient Reported Outcomes*, den sog. *area*-Indikatoren und der Problematik der Indikatoren-Sets. **Patient Reported Outcomes** (PRO) stellen einen wichtigen Einstieg in das Thema unterschiedlicher Qualitätsperspektiven dar (s. Kap. 2.4.2., Kap. 5.2.2.1.) und sind international z.B. bereits im Zusammenhang mit P4P-Programmen in Gebrauch (Ryan und Blustein 2012). Patientenorientierung und –zentrierung gehören zu den zentralen Trends in Praxis und Wissenschaft der letzten Jahre. Bei der Entwicklung von Präventionsempfehlungen zur Vermeidung von Unerwünschten Ereignissen spielt der „aktive Patient“ eine wichtige Rolle: man versucht, die Beobachtungs- und Auffassungsgabe der Patienten aktiv zu Zwecken der Verbesserung der Patientensicherheit zu nutzen (Schrappe 2010, S. 385). Im Mittelpunkt stehen dabei die *Patient-Reported Outcome Measures* (PROM's), also die Dinge, die der Patient selbst und nur er selbst berichten kann (z.B. postoperative Schmerztherapie (Varagunam et al. 2014)). Sie betreffen vornehmlich den Bereich der Prozessindikatoren (Koordination, Information, Kommunikation). Für die Externe Qualitätssicherung nach §136 SGB V in Deutschland wurde von Döbler und Geraedts (2017) in einer kürzlichen Analyse festgestellt, dass von den 239 Indikatoren des Erfassungsjahres 2016 nur ein einziger Indikator die Patientenperspektive betrifft. Entsprechende Entwicklungen durch das IQTiG nach den Bestimmungen des §137a SGB V sind daher von großer Wichtigkeit.

Die **area-Indikatoren**, die Auskunft über die Sicherheit (und Qualität) in regionaler oder Populationsperspektive geben, sind in der internationalen Literatur in großem Umfang beschrieben und wurden bereits im Sachverständigenratsgutachten von 2007 in die Diskussion eingebracht (SVR 2008, Nr. 501). Der AHRQ-PSI-Set enthält sieben *area*-Indikatoren (s. Tab. 3.3.-3), die jedoch inhaltlich wenig begründet erscheinen (Kap. 3.3.3.3.3.). Diese Indikatoren sind von Leistungsanbieter-bezogenen Indikatoren, von Indikatoren zu Einzelprozeduren und von (erkrankungsbezogenen) sektorübergreifenden Indikatoren abzugrenzen und greifen primär Versorgungsprobleme auf, die in der Zusammenarbeit der Sektoren und Leistungsanbieter eine Rolle spielen (ausführlicher Kap. 5.2.2.3.). Die regionale Perspektive muss in der Begrifflichkeit und im Konzept durch den Populationsbezug ergänzt werden (Nolte et al. 2010, Hildebrandt 2015), denn neben die geographischen Dimension tritt z.B. der Versicherungs- oder der funktionale Aspekt (wer koordiniert die Behandlung, s. von Stillfried und Czihal 2014). Wichtig ist an dieser Stelle die Erkenntnis, dass Regionen unter den Bedingungen der Sektorierung keine „homogenen Qualitätsräume“ darstellen, dass also die Qualität der einzelnen Sektoren innerhalb einer Region nicht miteinander korreliert (Herrin et al. 2016). Wahrscheinlich ist es unumgänglich, im Zusammenhang mit Patientensicherheit in Zukunft auch die Einführung einer populationsbezogenen Vergütung zu diskutieren. Unter diesen Bedingungen stellen z.B. die ungeplanten Wiederaufnahmeraten (Herrin et al. 2015) oder

die ambulant-sensitiven Krankenhausaufnahmen (Fischbach 2016, Sundmacher und Schüttig 2016) wichtige Indikatoren dar.

Der dritte Punkt betrifft die **Indikatoren-Sets** (s. Kap. 3.3.3.3.3.). Die Gesundheitsversorgung ist ein multidimensionales, mitunter komplexes Geschehen. Gut ausgewählte und validierte Indikatoren können zwar ihre „Ampelfunktion“ für größere Bereiche der Gesundheitsversorgung wahrnehmen, trotzdem bleiben sie immer selektiv. In einigen Fällen ist dies hinreichend, wenn z.B. nur ein bestimmter Bereich der Gesundheitsversorgung beobachtet werden soll. In anderen Fällen werden Indikatorensets herangezogen, um auf einen Blick größere Bereiche der Versorgung abbilden zu können. Indikatoren-Sets sind auch außerhalb des Gesundheitswesens in Gebrauch (Goddard und Jacobs 2010) und können als *composite measures* auf einen einzigen Wert kondensiert werden, dies ist aber nicht zwingend notwendig. Es handelt sich immer um mehr als eine reine Zusammenstellungen von Einzelindikatoren, denn sie können die Validität der Einzelindikatoren auch verschlechtern, wenn man die Validität des Gesamt-Sets nicht kritisch untersucht. Kondensierte Indikatoren-Sets werden definiert als

“a composite measure is a combination of two or more individual measures in a single measure that results in a single score” (NQF 2009 S. 2, ähnlich Peterson et al. 2010, Metge et al. 2009).

Das IOM charakterisiert Sets in ihrem *Vital Core Measures Set* durch ihre systemische Bedeutung, ihre Outcome-Orientierung, ihre Bedeutung für Patienten, ihre Datensparsamkeit, ihre Repräsentativität und ihren Nutzen auf unterschiedlichen Ebenen des Systems (IOM 2015A, Box S-2). Zusätzlich wird oft die bessere Vergleichbarkeit von Einrichtungen durch die Kondensierung auf einen einzelnen Wert genannt (Peterson et al. 2010, Shwartz et al. 2015). Als Adressaten werden individuelle Patienten und Erkrankungen, Institutionen mit ihren unterschiedlichen Hierarchie-Ebenen, Versicherungen (z.B. Impfquote), die Bevölkerung, das System und die Wissenschaft genannt (Hyder et al. 2012, IOM 2006 S. 91ff, Shwartz et al. 2015).

Eines der ältesten und am besten erprobten Indikatorensets wurde von der *Health Care Financing Administration* (HCFA) bzw. später dem *Centers for Medicare & Medicaid Services* (CMS) entwickelt und für den Vergleich der Versorgungsqualität der Bundesstaaten der USA verwendet (Jencks et al. 2000, Jencks et al. 2003). Es basierte auf 6 Erkrankungen mit insgesamt 24 Einzelindikatoren (akuter Myokardinfarkt, chronisches Herzversagen, Schlaganfall, ambulant erworbene Pneumonie, Brustkrebs und Diabetes). Das Instrument ergibt über die Zeit sehr hochgradig korrelierende Ergebnisse.

Mit Indikatoren-Sets ist immer wieder die Hoffnung verknüpft worden, dass wegen der höheren Fallzahlen und Ereignishäufigkeiten Unterschiede besser zu erkennen sind als bei der Nutzung von Einzelindikatoren, wo gerade unter den Bedingungen der Risikoadjustierung kleine Gruppengrößen eine sehr große Streuung aufweisen und Unterschiede nicht mehr nachweisbar sind. Genauso ist es allerdings möglich, dass kleine Unterschiede verwischt werden, weil die Vielzahl der Indikatoren ein „Rauschen“ verursachen. In jedem Fall geht die Hoffnung nur unter der Bedingung in Erfüllung, dass die Einzelindikatoren transparent ausgewählt sind und voneinander unabhängige Ereignisse darstellen (Smith 2010, S. 14). Zur Auswahl der Einzelindikatoren gibt es verschiedene Vorschläge, die alle nicht maßgeblich von Auswahlkriterien bei Einzelindikatoren abweichen (Jencks et al. 2003, NQF 2009, Peterson et al. 2010, Schwartz et al. 2015), im Mittelpunkt stehen Validität und Reliabilität, Machbarkeit und Relevanz.

Die Angabe „der Indikatorenset ergibt einen Wert von 90%“ kann heißen, dass alle Indikatoren bei 90% liegen oder alle 10 Indikatoren bei 100% und ein Indikator bei 0%. Man bezeichnet dieses Phänomen als Feuerlöscher-Problem, denn wenn man sich vorstellt, einer von 10 Indikatoren sei extrem wichtig (Vorhandensein eines Feuerlöschers), dann kann man sich mit der alleinigen Kenntnis der Angabe „der Set steht bei 90%“ nicht zufrieden geben. Hier hilft ein schrittweises Vorgehen weiter: zuerst wird der Indikator Feuerlöscher abgefragt und nur dann die anderen Indikatoren, wenn hier ein bestimmter Grenzwert erreicht wird (oder bei binären Indikatoren ein „ja“). Ein ähnlicher Effekt tritt ein, wenn einer oder mehrere Einzelindikatoren ein *ceiling* aufweisen und eine Verbesserung eines anderen Indikators, der noch Verbesserungspotential besitzt, dadurch verborgen bleibt (IOM 2006, S. 92). Gleiches gilt, wenn ein (oder mehrere) Indikator(en) völlig irrelevant sind: eine evtl. Verbesserung oder Verschlechterung einer anderen Indikators, der vielleicht im Set gut abgebildet hätte werden können, wird durch den irrelevanten Indikator verwischt, so dass die Validität des Indikatoren-Sets unter der der Einzelindikatoren liegt (Peterson et al. 2010).

Dieses Beispiel führt zu Problematik der Gewichtung der Einzelindikatoren innerhalb des Sets, die bzgl. der US-amerikanischen VBP-Programme derzeit anhand des PSI90 Indikatorsets kritisch diskutiert wird (Rajaram et al. 2015; s. Kap. 3.3.3.3.3.). In der Literatur sind mehrere Methoden beschrieben, mit denen man die Gewichtung erstellen und absichern kann (Übersicht z.B. bei Goddard und Jacobs 2010). Ein weiterer Punkt besteht in der Form der Aggregation. Man kann die einzelnen Indikatoren als einzelne Kriterien aufsummieren (8/10 Indikatoren erfüllen die Anforderungen), kontinuierliche Skalen anwenden („die Indikatoren stehen bei 7,5“), einen Durchschnitt bilden oder man kann über Entscheidungsbäume vorgehen (NQF 2009, S. 4), wenn die Gewichtungen sehr unterschiedlich sind (Feuerlöscher-Beispiel). Unter dem Strich wird die Verwendung

von Indikatoren-Sets in der wissenschaftlichen Literatur nicht von allen Autoren geteilt, insbesondere wenn es um einzelne Versorgungsbereiche geht: „*Despite these multiple purposes, composite measures are not useful for targeting specific areas of improvement*“ (Shwartz et al. 2015).

### 3.3.3.3.2. Patientensicherheitsindikatoren

Patientensicherheitsindikatoren (PSI) stellen im Zusammenhang mit der Diskussion der quantitativen Bestimmung von Patientensicherheit eines der wichtigsten und zugleich widersprüchlichsten Themen dar (s. Kap. 5.3.4.). Gemeinhin wird angenommen, dass man mit PSI Patientensicherheit direkt quantitativ erfassen könne. Aus historischem Blickwinkel ist dies auch leicht nachvollziehbar, so wurde das Thema in den USA und in Deutschland mit Angaben über die Sterblichkeit an Unerwünschten Ereignissen im Krankenhaus in der öffentlichen und fachlichen Wahrnehmung auf die Tagesordnung gesetzt (s. Kap. 1.2.1.). Dies war nicht falsch, denn die Daten waren sachkundig erhoben und valide – und es wird auch in Zukunft notwendig sein, Daten dieser Art zu erheben. Allerdings ist es von größter Bedeutung zu erkennen, dass es dabei nicht im eigentlichen Sinne um Patientensicherheitsindikatoren, sondern um die Ergebnisse von quantitativen Erhebungen mit klinisch-epidemiologischen Falldefinitionen handelte, die Unerwünschte Ereignisse (UE) erfassten (s. Kap. 3.3.3.2., s. Kasten). Es ist also ein sehr differenziertes Umgehen mit den Begrifflichkeiten notwendig.

#### **Steckbrief Patientensicherheits-indikatoren (PSI)** (s. Kap. 5.3.4.)

- 1** Definition: Qualitätsindikatoren mit Schwerpunkt Patientensicherheit.
- 2** Faktisch: PSI stellen heutzutage in den allermeisten Fällen Unerwünschte Ereignisse (UE) dar, die mit klinisch-epidemiologischem Erkenntnisinteresse erhoben werden.
- 3** PSI-Sets entsprechen heute weitgehend „UE-Sets“.
- 4** Die Kenntnis der UE-Raten ist wichtig, sie stellen ein entscheidendes *Outcome* dar ...
- 5** ... aber es ist unklar, ob sie tatsächlich die Sicherheit der stattgehabten oder zukünftigen Versorgung beschreiben.
- 6** Die heute verwendeten PSI entsprechen eher einem überkommenen linearen Verständnis von Patientensicherheit.
- 7** Es besteht enormer Handlungsbedarf für die (Weiter-)Entwicklung von PSI (s. Kap. 5.3.4.).

Soweit man sich auf der Basis einer adäquaten Definition des Indikators bewegt (s. Kap. 3.3.3.3.1.), ist im Grunde die **Definition** von Patientensicherheitsindikatoren (PSI) in ihrer Abgrenzung vom allgemeinen Begriff der Qualitätsindikatoren wenig problematisch, denn sie beruht lediglich auf einer inhaltlichen Gewichtung: PSI sind Qualitätsindikatoren mit

dem Schwerpunkt Patientensicherheit (Davies 2002, SVR 2008, Nr. 616). PSI sind zu bezeichnen als

*„specific quality indicators which also reflect the quality of care inside hospitals, but focus on aspects of patient safety. Specifically, PSIs screen for problems that patients experience as a result of exposure to the healthcare system, and that are likely amenable to prevention by changes at the system or provider level“* (McDonald et al. 2002).

Der Sachverständigenrat betont in seiner schwerpunktmäßigen Beschäftigung mit diesem Thema besonders die Übereinstimmung des Indikatorenkonzeptes mit dem Konzept des Beinaheschadens; letztere können als Parameter angesehen werden, die im Sinne ihrer prozessualen Vorhersagefunktion als gute Kandidaten für PSI gelten (ebd., Nr. 615, s. auch Kap. 5.3.4.). Es werden aber auch einige andere Besonderheiten von PSI herausgearbeitet:

- PSI sagen besonders schwere und relevante Ereignisse voraus, deren Nichtkenntnis nicht zu vertreten wäre (**hohe Sensitivität**);
- die Dringlichkeit ist hoch, Wiederholungen müssen rasch ausgeschlossen werden, der Lernzyklus ist kurz, daher bestehen hohe Anforderungen an Beeinflussbarkeit, Konsequenz, **Machbarkeit** und praktischen Relevanz;
- die **Prozessanalyse** nach unerwünschten Ereignissen muss **normativer** ablaufen als dies im Qualitätsmanagement gängig ist, daher ist zu fordern, dass PSI in starkem Maße regelbasiert sind, um den normativen Zugang zur Analyse von Schadensfällen zu erleichtern.

International und in Deutschland sind zahlreiche PSI beschrieben (ausführliche Diskussion s. Kuske et al. 2013). Der Sachverständigenrat hat in seinem Gutachten 2007 eine umfassende Literaturübersicht über die international gebräuchlichen Patientensicherheitsindikatoren angefertigt (SVR 2008, Nr. 649ff) und daraus einen Set von 30 PSI abgeleitet (s. Tab. 3.3.-2). Diese PSI sind in Anlehnung an die Systematik der JCAHO (*Joint Commission for Accreditation of Health Care Organizations*) eingeteilt in

- globale Indikatoren,
- übergreifende Indikatoren,
- diagnosebezogene Indikatoren,
- fallbezogene Indikatoren und
- organisatorische Indikatoren.

Der Systematische Review wurde später deutlich ausgeweitet (Kuske et al. 2013).

<b><u>Bereiche</u></b>	<b><u>Indikatoren</u></b>
<b>1. Globale Indikatoren</b>	1.-a. Sterblichkeit bei DRGs mit niedriger Mortalitätsrate
	1.-b. Dekubitus
	1.-c. In-Hospital Hüftfraktur
<b>2. Übergreifende Indikatoren</b>	2.-a. Perioperative Mortalität
	2.-b. Nosokomialer Myokardinfarkt
2.1. Wiederaufnahme	2.1.-c. Ungeplante stationäre Wiederaufnahme innerhalb von 30 Tagen
	2.1.-b. Ungeplante Aufnahme bzw. Rückkehr auf die Intensivstation
2.2. Intraoperativ	2.2.-a. Anästhesiekomplikation
	2.2.-b. Eingriffs- und Seitenverwechslung
	2.2.-c. Belassen eines Fremdkörpers während des Eingriffs
2.3. Postoperativ	2.3.-a. Ungeplante Re-Operation
	2.3.-b. Postoperative pulmonale Embolie oder tiefe Venenthrombose
	2.3.-c. Postoperative Sepsis
	2.3.-d. Postoperative Hämorrhagie oder Hämatom
2.4. Ausgewählte nosokomiale Infektionen (AHRQ)	2.4.-a. Wundinfektion
	2.4.-b. Beatmungsbedingte Pneumonie
	2.4.-c. Infektionen von intravasalen und Harnwegskathetern und Drainagen
2.5. Technical Devices	2.5.-a. Unerwünschte Ereignisse im Zusammenhang mit Medizinprodukten (AMDE: adverse medical device events)
2.6. Medikamentenbedingt	2.6.-a. Medikationsfehler
	2.6.-b. Mortalität infolge einer vermeidbaren unerwünschten Medikamentenwirkung
2.7. Sentinel events	2.7.-a. Iatrogener Pneumothorax
	2.7.-b. Transfusionsreaktion
	2.8.-c. Erfolgreiche Wiederbelebung
<b>3. Diagnose-bezogene Indikatoren</b>	3.-a. Schlaganfall nach einem herzchirurgischen Eingriff 3.-b. Amputation bei Diabetes-Patienten 3.-c. Amputation nach einem gefäßchirurgischen Eingriff
<b>4. Fachspezifische Indikatoren:</b> Beispiel Geburtshilfe	4.-a. Geburtshilfliches Trauma – Vaginale Entbindung mit Instrument
	4.-b. Geburtshilfliches Trauma – Vaginale Entbindung ohne Instrument
<b>5. Organisatorische Indikatoren:</b> 5.1. Vorfälle	5.1.-a. Körperliche Zwangsmaßnahmen (Dauer, Zeit, Grund, Verletzungen)
5.2. Personal- und Einrichtungsressourcen	5.2.-a. Arbeitszeit

Tab. 3.3.-2: PSI-Set des Sachverständigenrates im Gutachten 2007 (SVR 2008 Nr. 649)



Wenn man sich beispielhaft die PSI aus Tab. 3.3.-2 anschaut (weitere Beispiele sind in Kap. 3.3.3.3. dargestellt), dann ergibt sich jedoch der oben bereits erwähnte, in seiner Deutlichkeit überraschende Befund: bei (fast) allen dieser Indikatoren handelt es sich eigentlich nicht um Indikatoren, sondern um typische klinisch-epidemiologische Falldefinitionen, die behandlungsbedingte Unerwünschte Ereignisse (UE) wiedergeben. Auch in der Externen Qualitätssicherung nach §136 SGB V sind die meisten verwendeten Parameter (153/239, 64%) als Patientensicherheits-relevant anzusehen (Döbler und Geraedts 2017), aber auch hier handelt es sich hier nicht um Indikatoren im oben beschriebenen Sinne, sondern um Ergebnisse klinisch-epidemiologischer Falldefinitionen, eine Vorhersage von Unerwünschten Ereignissen im Sinne des genuinen Indikatorkonzeptes ist in vielen Fällen nicht validiert. Die Aussage macht den großen Weiterentwicklungsbedarf auf dem Gebiet der Indikatoren bzw. PSI deutlich (Kap. 5.3.4.).

Zunächst ist man vielleicht geneigt, diese Beobachtung als minder relevant einzuschätzen – Komplikationen sind eben Komplikationen – aber es ergeben sich Schwierigkeiten spätestens dann, wenn man sich an das in Kap. 2 hergeleitete **multimodale Konzept von Patientensicherheit** erinnert (Kap. 2.6.2.). Dieses Konzept sagt aus, dass Sicherheit bzw. Patientensicherheit nicht abschließend mit einem günstigen *Outcome* zu beschreiben ist, sondern das Sicherheitsverhalten und das Umgehen mit Risiken umfasst, weiterhin die Problemlösungskompetenzen der Experten und Teams vor Ort, das Lernen der Organisationen, die Innovationskompetenz, natürlich die Komplexität des Systems und die Fähigkeit der Systeme, mit emergenten Ereignissen umzugehen – mit den Worten von Sidney Dekker und Nancy Levenson (2014): *“The question it [the system approach, MS] pursues is not why bad operators make mistakes, but why good ones do.”* *The good ones*: damit sind diejenigen gemeint, die vielleicht geringe Raten postoperativer Komplikationen erreichen konnten, während in ihren Teams, organisatorischen Einheiten und Systemen Sicherheit jedoch keine Priorität darstellt.

Es liegt also, wie ja bereits einleitend in Kap. 3.1. festgestellt, im jetzigen Verständnis von Patientensicherheitsindikatoren eine **„Dissoziation von Konzept und Erhebungsmethodik“** vor - die gegenwärtig genutzten PSI entsprechen dem im ersten Teil von Kap. 2 diskutierten Verständnis von Patientensicherheit als „Abwesenheit von Unerwünschten Ereignissen“. Die hinter den derzeitigen verwendeten PSI stehende Konzeption entspricht dem linearen Prozess-Verständnis (s. Kap. 2.4.4.) und basiert auf der Annahme, dass das *Outcome* als Endpunkt der Fehlerkette die Sicherheit des Prozesses widerspiegelt. Wie in Kapitel 2 (2.4.5-2.4.7.) ausführlich diskutiert, kann dieses Prozess-orientierte Verständnis zwar wichtige Hilfestellungen geben, ist aber nicht in der Lage, die Prozesse im Bereich der Teams, der Organisation und innerhalb komplexer Systeme abschließend zu beschreiben. Schon allein die Frage der

Vermeidbarkeit kann so nicht beantwortet werden (s. Kap. 3.5., Pronovost und Colantuoni 2009).

Dies führt zu der Problematik, dass die gegenwärtig verwendeten „PSI“ hinsichtlich der **prospektiven Validität von PSI** für die Sicherheit der Versorgung weitgehend versagen. Die Frage lautet ja: können diese PSI die Sicherheit der kommenden Versorgung vorhersagen, richten sie also im Sinne der JCAHO-Definition (s. Kap. 3.3.3.3.1.) das Interesse wirklich auf eventuell vorhandene Problembereiche? Man muss hier sehr vorsichtig sein, denn die Annahme, dass um so stärker und häufiger die *Outcome*-bezogenen PSI der gegenwärtigen Form ansprechen, die Schadensereignisse um so häufiger oder schwerwiegender seien, ist nicht nur theoretisch fragwürdig (s.o.), sondern auch empirisch wenig unterlegt (Brown et al. 2008C). Einzelne Studien zeigen zwar, dass für Krankenhäuser und Operateure, die heute eine höhere Mortalität aufweisen, dies auch im darauffolgenden Jahr der Fall ist – dies wäre eine wirklich Validierung eines PSI (Jha und Epstein 2006 mit den Daten des *New York State Cardiac Surgery Report Systems (CSRS)*). Im Normalfall kann von diesem Zusammenhang jedoch nicht ausgegangen werden. Hinzu kommt die **mangelnde Reliabilität der heute verwendeten PSI**. Man sollte meinen, diese sei bei Komplikationen oder gar bei einem Parameter wie der Mortalität kein Problem, aber dies ist nicht der Fall. Hierauf wird in Kap. 3.3.5. näher eingegangen. Es muss also neu nachgedacht werden (s. Kap. 5.3.4.), es ist dringend erforderlich, PSI in ihrem Fokus und ihrer Messmethodik weiterzuentwickeln.

#### **PSI: vier Fragestellungen**

##### **1 Kongruenz mit dem Konzept**

**Patientensicherheit:** geben die gegenwärtig genutzten PSI Auskunft über die tatsächlich vorgehaltene Patientensicherheit?

##### **2 Prospektive Validität der PSI für die Sicherheit der Versorgung:**

können die gegenwärtig genutzten PSI die Sicherheit der kommenden Versorgung vorhersagen, dirigieren sie also wirklich das Interesse auf die vorhandenen Problembereiche?

**3 Reliabilität und Validität:** welche Anforderungen sind an die Reliabilität von PSI zu stellen, und welche Folgen hat dies für die Validität?

**4 Weiterentwicklung:** wie sieht die Entwicklungsperspektive für die Thematik PSI aus (s. Kap. 5.3.).

### 3.3.3.3.3. PSI-Sets

Patientensicherheit ist ein multidimensionales Geschehen und stellt daher eine sinnvolle Einsatzmöglichkeit für Indikatoren-Sets dar (s. Kap. 3.3.3.1.). Es gibt im internationalen Kontext zahlreiche PSI-Sets, die Übertragbarkeit auf Deutschland wurde teilweise untersucht und ist für den stationären Sektor recht gut gegeben (Weingärtner et al. 2013). Auch hier muss die Einschränkung vorausgeschickt werden, dass es sich eigentlich um „UE-Sets“ handelt, also Zusammenstellungen von mit klinisch-epidemiologischen Falldefinitionen erfassten Komplikationen (s. vorangehendes Kapitel). Trotzdem sollen diese Sets hier unter dem Titel „PSI-Sets“ dargestellt werden.

Europa wurde das Indikatoren-Set des Projektes ‘*Safety Improvement for Patients in Europe*’ (SimPatiE) entwickelt, das 28 Indikatoren umfasst (Kristensen et al. 2009). Die OECD entwickelte das *OECD Health Care Quality Indicator (HCQI) Project* mit einer Auswertung hinsichtlich *Patient Safety* auf Ebene der Gesundheitssysteme mit 21 Indikatoren (Kelley und Hurst 2006, McLoughlin et al. 2006, Drösler et al. 2009). Das PSI-Set des *National Quality Forum* (NQF 2009) umfasst 8 Indikatoren (Dekubitus, Pneumothorax, Nosokomiale Infektionen, postoperative Hüftfraktur, thromboembolisches Ereignis, Sepsis, postoperative Wunddehiszenz, intraoperative Verletzung (*accidental puncture or laceration*)), zusätzlich gibt es ein pädiatrisches Set.

Auf Routinedaten (s. Kap. 3.3.5.) basiert das bekannte **AHRQ-PSI Set**, das 20 Indikatoren umfasst (McDonald et al. 2002, Romano et al. 2003) und aus dem *Complications Screening Program* (CSP) hervorging (Weingart et al. 2000, Weingart und Iezzoni 2003). In Tab. 3.3.-3 ist die aktuelle Aufstellung für das Jahr 2015 wiedergegeben, einschließlich der beobachteten Raten für das Jahr 2012 (AHRQ 2017B). Die Häufigkeitsangaben werden weitergehend in Kap. 3.3.5. (Nutzung von Abrechnungsdaten) und Kap. 3.6.2. diskutiert (zur Epidemiologie). Die AHRQ-PSI werden von allen US-amerikanischen Krankenhäusern erhoben und beinhalten sowohl 7 *area*-Indikatoren (s. Kap. 3.3.3.3.1.) als auch das PSI90-Indikatorenset, das im *Hospital Quality Star Rating* des CMS (Bilimoria und Barnard 2016), im *Value Based Purchasing*, dem amerikanischen P4P-Äquivalent, und im ebenfalls als P4P-Programm konzipierten *Hospital-Acquired Condition Reduction Program* (HACP) eine große Rolle spielt (Kahn et al. 2015).

PSI	Patient Safety Indicator	PSI90	Area	Rate/1000*
02	Death rate in low-mortality diagnosis-related groups (DRGs)			0,32
03	Pressure ulcer rate	X		0,50
04	Death rate among surgical inpatients with serious treatable complications			118,62
05	Retained surgical item or unretrieved device fragment count		X	k.A.
06	Iatrogenic pneumothorax rate	X	X	0,34
07	Central venous catheter-related bloodstream infection rate	X	X	0,21
08	Postoperativ hip fracture rate	X		0,04
09	Perioperative hemorrhage of hematoma rate		X	5,11
10	Postoperative physiologic and metabolic derangement rate			0,69
11	Postoperative respiratory failure rate			10,05
12	Perioperative pulmonary embolism of deep vein thrombosis rate	X		4,99
13	Postoperative sepsis rate	X		9,61
14	Postoperative wound dehiscence rate	X	X	1,86
15	Accidental puncture or laceration rate	X	X	1,89
16	Transfusion reaction count		X	k.A.
17	Birth trauma rate – injury to neonate			1,89
18	Obstetric trauma rate – vaginal delilvery without instrument			133,19
19	Obstetric trauma rate – vaginal delivery with instrument			20,97
90	Patient safety to selected indicators (composite measure)			-

Tab. 3.3.-3: AHRQ-PSI Set 2015 mit Ergebnissen aus 2012 (AHRQ 2017B). \*Die Rate bezieht sich auf exponierte Patienten (*at risk*), die den Ein- und Ausschlusskriterien genügen (Spezifikationen unter [www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V50/TechSpecs/PSI\\_50\\_updates\\_techsspecs.zip](http://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V50/TechSpecs/PSI_50_updates_techsspecs.zip)). Beispiel: PSI 13 (Postoperative Sepsis) schließt alle operierten Elektiv-Patienten älter als 17 Jahre ohne Sepsis bei Aufnahme ein, die z.B. nicht an einem Dekubitus, Immunkompromittierung, Tumorerkrankung oder primären Infektion leiden. Diese Kriterien sind bei dem Vergleich der Raten streng zu beachten, trotzdem sind die Daten zur Diskussion der Größenordnung und der methodischen Voraussetzungen (z.B. Datenquelle: Abrechnungsdaten) wichtig.

Der PSI90-Set, der aus 8 Einzelindikatoren zusammengesetzt ist (s. Tab. 3.3.-3), wird derzeit sehr kontrovers diskutiert (Bilimoria und Barnard 2016, JHA und Pronovost 2016). Der Indikatoren-Set basiert auf administrativen Daten, die sehr insensitiv sind, nur in einem Bereich (nosokomiale Infektionen) werden klinisch-epidemiologische Daten verwendet (Jha und Pronovost 2016, s. auch Kap. 3.3.5.). Das Set ist inhomogen und teilweise einem *detection-* bzw. *surveillance bias* unterworfen (PSI 2 Dekubitus und PSI 12 thromboembolisches Ereignis), die beobachtete Rate bildet also in erster Linie den diagnostischen Aufwand ab, der betrieben wird, um die Komplikation festzustellen (Bilimoria et al. 2013, Schwartz et al. 2015). Die Kritik an der mangelnden Spezifität bzw. dem niedrigen Positiven Prädiktive Wert von PSI15 (intraoperative Verletzung) kann man noch relativieren (s. Ausführungen zu den statistischen Anforderungen, denen Indikatoren unterliegen – Spezifität steht nicht im Vordergrund (s. Kap. 3.3.3.3.1.)), jedoch ist die Kritik an der mangelnden Risikoadjustierung dieser sehr *Outcome*-orientierten Indikatoren nicht von der Hand zu weisen, genauso wichtig erscheint die Kritik an der sehr inhomogenen Gewichtung (PSI15 Faktor 0,4917, PSI 12 0,2579, ..., PSI08, 0,0011) (Austin et al. 2016, Rajaram et al. 2015, Winters et al. 2016). Natürlich ergibt der AHRQ-PSI-Set sehr interessante und differenzierte Einsichten über die Häufigkeiten der (eigentlich als UE-Raten zu bezeichnenden) „Indikatoren“, gerade auch im Hinblick auf die genutzten Datenquellen (Abrechnungsdaten). Hierauf wird in Kapitel 3.3.5. ausführlich eingegangen.

Ein methodisch deutlich umfassenderes PSI-Set ist unter der Bezeichnung **Medicare Patient Safety Monitoring System (MPSMS)** seit dem Jahr 2001 von den *Centers of Medicare and Medicaid Services* (CMS) gemeinsam der AHRQ, den CDC, der FDA und anderen Einrichtungen entwickelt worden (Hunt et al. 2005). Das PSI-Set besteht aktuell aus 21 Indikatoren aus 4 Domänen und umfasst neben den typischen nosokomialen Infektionen und postoperativen Komplikationen (einschließlich der Kontrastmittel-assoziierten Nephropathie) auch Unerwünschte Arzneimittelereignisse (Digoxin, Antikoagulation, Hypoglykämie) sowie explizit die *Clostridium difficile*-, MRSA- und VRE-Infektionsraten. Es ist also weitaus breiter aufgestellt als z.B. das AHRQ-Set und schließt Teile des konservativen Bereichs mit ein (s. Tab. 3.3.-4). Es steht hinsichtlich der Datenquellen auf „zwei Beinen“, es beruht nämlich auf einer Kombination von administrativen Daten (zur Erhöhung der Reliabilität) und einer Algorithmus-gestützten Analyse von Krankenakten (s. Kap. 3.3.5.), entspricht also in etwa einer Kombination des PSI-Sets der AHRQ mit dem *Harvard Medical Practice Study* (HMPS) *Design* (s. Kap. 3.3.3.4.). In partiellem Unterschied zum HMPS-Design wird die Analyse der Krankenakten nicht von den Behandelnden oder in der Institution angefertigt (was in den HMPS-Studien nicht immer ganz klar ist), sondern explizit an einer externen Stelle, und die Reviewer werden aufgefordert, nur nach dem dokumentierten Status zu gehen und nicht ihr eigene klinisches Urteil einfließen zu lassen (Differenz von klinisch-epidemiologischer Falldefinition und klinischer Diagnose, s. Kap. 3.3.3.2.).

Das Instrument ist kosteneffektiv und hinsichtlich seiner Reliabilität sowie Konsistenz gut untersucht (Classen et al. 2016, Hunt et al. 2005). Es verfolgt ein klar umrissenes Ziel, und – das ist die absolute Ausnahme – konstatiert im Gegensatz zu allen anderen „PSI-Sets“ ganz deutlich, dass es um die Erfassung von Unerwünschten Ereignissen geht, und nicht um die Feststellung der Vermeidbarkeit der UE (Classen et al. 2016) bzw. die Erfassung von Indikatoren. Weiterhin wird auch nicht die Absicht verfolgt, das Augenmerk auf Verbesserungsmaßnahmen zu richten:

*“Within that larger mission, the role of the MPSMS is only the measurement of adverse events. **The project has no capacity to initiate interventions or affect change in the topics observed.** That capacity is the work of other quality improvement projects within QIG [Quality Improvement Group des CMS, MS]. Specialization of MPSMS functionality to focus only on measurement significantly reduces the operational complexity“ (Hunt et al. 2005, Hervorh. MS).*

Vom Jahr 2005 bis 2011 wurden in den USA mehr als 60.000 Patienten mit Herzversagen, Myokardinfarkt, Pneumonie und größeren operativen Eingriffen mit dem MPSMS erfasst und ein partieller Abfall der Häufigkeit von UE gefunden (Wang et al. 2014). Im Jahr 2014 wurden in 1109 Krankenhäusern 20.000 zufällig ausgewählte Krankenakten von Patienten mit den gleichen Diagnosen wie bei Wang et al. (2014) analysiert (Classen et al. 2016). Die Relevanz und Machbarkeit des Instrumentes darf also als gegeben angesehen werden. Die Limitationen des Verfahrens (Classen et al. 2016) bestehen in der retrospektiven Perspektive, Fehler in der expliziten Analyse, in lokalen Unterschieden in der Dokumentation und Dokumentationslücken sowie in der fehlenden Kompetenz in der Feststellung der Vermeidbarkeit. Gerade der letzte Punkt stellt einen Nachteil gegenüber dem impliziteren HMPS-Design dar. Das MPSMS wird derzeit in das *Quality and Safety Review System (CSRS)* weiterentwickelt und soll um automatisierte Auswertungsmodule der elektronischen Krankenakte ergänzt werden. Eine genauere vergleichende Analyse wird in den beiden kommenden Kapitel vorgenommen.

<b>MPSMS-Indicator: Measure name</b>	<b>Domain</b>	<b>Data available</b>	<b>Rate %</b>
Adverse events associated with digoxin	ADE	2004 – 2014	0,9
Adverse events associated with hypoglycemic agents	ADE	2004 – 2014	8,8
Adverse events associated with heparin	ADE	2004 – 2014	11,1
Adverse events associated with low molecular weight heparin and factor Xa inhibitors	ADE	2004 – 2014	3,5
Adverse events associated with warfarin	ADE	2004 – 2014	4,8
Hospital-acquired pressure ulcers	Gen	2004 – 2014	3,7
Inpatient falls	Gen	2005 – 2014	0,9
Central line – associated bloodstream infections	HAI	2002 – 2014	0,3
Postoperative pneumonia	HAI	2002 – 2014	1,5
Hospital-acquired antibiotic-associated C. difficile	HAI	2004 – 2014	0,5
Catheter-associated urinary tract infections	HAI	2005 – 2014	2,6
Hospital-acquired MRSA	HAI	2005 – 2014	0,06
Hospital-acquired vancomycin-resistant enterococcus	HAI	2005 – 2014	0,05
Ventilator-associated pneumonia	HAI	2005 – 2014	11
Adverse events associated with hip joint replacement	PP	2002 – 2014	4,5
Adverse events associated with knee joint replacement	PP	2002 – 2014	2,6
Mechanical complications associated with central lines	PP	2002 – 2014	3,3
Postoperative venous thromboembolic events	PP	2002 – 2014	0,5
Postoperative cardiac events (cardiac and noncardiac surgeries)	PP	2004 – 2014	0,8
Adverse events associated with femoral artery puncture for catheter angiographic procedures	PP	2005 – 2014	2,1
Contrast nephropathy associated with catheter angiography	PP	2005 – 2014	11,9

Tab. 3.3.-4: 21 Indikatoren-Set des *Medicare Patient Safety Monitoring System* (MPSMS) (Chassin et al. 2016). ADE *Adverse Drug Event*, Gen *General*, HAI *Hospital-Acquired Infection*, PP *Postprocedure*. Rate: % der exponierten Patienten (ohne CI) für 2014.

Ein aus deutscher Sicht in die Zukunft weisendes Indikatoren-Set wurde von der **Leapfrog-Gruppe** in den USA entwickelt. Die Leapfrog-Gruppe hat sich vor mehr als 10 Jahren als Organisation von Unternehmen gegründet und kooperiert eng mit den *Centers of Medicare and Medicaid Services* (CMS) (s. [www.leapfrog.org](http://www.leapfrog.org), Moran und Scanlon 2013). 2500 amerikanische Krankenhäusern werden anhand eines Scores (A (sehr gut) bis F) verglichen, der sich aus insgesamt 27 Items zusammensetzt (Leapfrog-Group 2017). Das Indikatoren-Set setzt sich aus folgenden Gruppen zusammen:

- 12 Prozess- und Struktur-Indikatoren
  - 2 *American Hospital Association* (AHA)-Indikatoren
    - CPOE: *Computerized Physician Order Entry*
    - IPS: *ICU Physician Staffing*
  - 5 NQF *Safe Practices Scores*
    - SP 1: *Culture of Safety Leadership Structures and Systems*
    - SP 2: *Culture Measurement, Feedback, and Intervention*
    - SP 4: *Risks and Hazards*
    - SP 9: *Nursing Workforce*
    - SP 19: *Hand Hygiene*
  - 5 *Hospital Consumer Assessment*-Indikatoren (HCAHPS, entwickelt von CMS und AHRQ)
    - H-Comp-1 *Nurse Communication*
    - H-Comp-2 *Doctor Communication*
    - H-Comp-3 *Staff Responsiveness*
    - H-Comp-5 *Communications about Medicines*
    - H-Comp-6 *Discharge information*
- 15 Outcome-Indikatoren
  - 3 *Hospital Acquired Conditions* (HAC-Indikatoren)
    - Foreign Object Retained*
    - Air Embolism*
    - Falls and Trauma*
  - 5 CMS *Hospital Compare*-Indikatoren
    - Central Line Associated Bloodstream Infection (CLABSI)*
    - Catheter-Associated Urinary Tract Infection (CAUTI)*
    - Surgical Site Infection* Kolorektale Chirurgie (SSI Colon)
    - Methicillin-resistente Staph. aureus (MRSA)
    - Clostridium difficile* Enterocolitis
  - 7 Indikatoren aus dem ARHQ-PSI-Set
    - PSI 3: *Pressure Ulcer*



- PSI 4: *Death among Surgical Inpatients*
- PSI 6: *Iatrogenic Pneumothorax*
- PSI 11: *Postoperative Respiratory Failure*
- PSI 12: *Postoperative Pulmonary Embolism/Deep Vein Thrombosis*
- PSI 14: *Postoperative Wound Dehiscence*
- PSI 15: *Accidental Puncture or Laceration*

Die auf die *End-Results* fixierten *Outcome*-Indikatoren werden also durch Indikatoren ergänzt, die einem adäquaten Verständnis der Verwirklichung von Sicherheit entsprechen, z.B. IT-bezogene Indikatoren, Personalausstattung, Führung, Sicherheitskultur, Hände-Hygiene und vor allem Patienten-bezogene Indikatoren, die die Kommunikation, Information und Koordination betreffen (vgl. Austin und Derks 2016). Ein PSI-Set dieser Art sollte in Deutschland in den Mittelpunkt der Diskussion gerückt werden (s. Kap. 5.3.8.).

### 3.3.3.4. Generierende Verfahren: CIRS und Co.

Eines der beiden großen Verdienste der Diskussion um Patientensicherheit besteht in der Entwicklung von Verfahren, die es Organisationen erlauben, Ereignisse wahrzunehmen, die sie bisher „nicht auf dem Radar hatten“, wobei ihnen dieses Nicht-Erkennen nicht als Problem bewusst war (sog. „*Unknown Unknowns*“). Es handelt sich nicht einfach um „Blinde Flecken“, denn hier kann ein Außenstehender ja den Finger heben (s. Abb. 3.3.-7), sondern niemand – weder im Innen- noch im Außenraum – weiß von der Existenz und Nichtexistenz dieser Ereignisse (man könnte sagen: es handelt sich um „doppelt-blinde Flecken“). In gewisser Form kann man solche Instrumente wie CIRS als eine Maßnahme ansehen, die es mit der Emergenz komplexer Systeme aufnimmt (s. Kap. 2.4.7.5.).

Der Begriff „Generierende Verfahren“ ist aus einer längerdauernden Diskussion entstanden, die die Suche nach einem Oberbegriff für „CIRS und Co.“ mit zahlreichen Experten zum Gegenstand hatte (Schrappe 2015, S. 80ff). Diese Bezeichnung soll zum Ausdruck bringen, dass diese

	Dem Selbst bekannt	Dem Selbst nicht bekannt
Den Anderen bekannt	Bekanntes Wissen	Unbekanntes Wissen
Den Anderen nicht bekannt	Bekanntes Unwissen	Unbekanntes Unwissen

Abb. 3.3.-7: Abgewandelte Form des Johari-Fensters zur Illustration der *Unknown Unknowns* rechts unten (n. Schreyögg 1999, S. 497)

Verfahren Ereignisse entdecken („generieren“) und in die weitere Analyse einbringen, die vorher wegen ihrer Unbekanntheit den klinisch-epidemiologischen Instrumenten (einfache Erfassung mit Falldefinitionen, Risikofaktoranalyse, Modellbildung etc.) nicht zugänglich waren. Ein historisches Beispiel kann dies illustrieren: wäre der Londoner Arzt John Snow während der Londoner Cholera Epidemie 1854 nicht unter dem Mikroskop auf ihm unbekannte kommaförmige Korpuskel gestoßen und dann auf die Idee verfallen, dass einer der Brunnen den Krankheitsherd darstellt, hätte er seine epidemiologische Untersuchungen und somit die Entdeckung der Pumpe in der *Broad Street* nicht durchführen können, denn zu dieser Zeit waren die Kontagionisten (*Vibrio Cholerae* war noch nicht entdeckt) gegenüber den Miasmatikern, die die Erkrankung nicht als Infektion, sondern als eine Art „Ausdünstung schlechter Luft“ verstanden, noch in der absoluten Minderzahl (die Idee einer Infektionskrankheit war unbekannt).

Die *Critical Incident Reporting Systeme* stellen nicht die einzige Form eines generierenden Verfahrens dar. Die gleiche Funktion wird z.B. von *Morbidity-Mortality Conferences*, Interviews von Patientenbeauftragten, Analyse von Freitexten in Patientenzufriedenheitsbefragungen und Auswertungen von Beschwerden wahrgenommen, immer wenn die Ereignisse vorher unbekannt waren und daher nicht Gegenstand gezielter prospektiver Erhebungen sein können (Thomas und Petersen 2003). Die verschiedenen Zugangsmöglichkeiten ergeben charakteristischerweise Ergebnisse, die sich nicht überlappen, d.h. jede Methode entdeckt „eigene“ Ereignisse (Hogan et al. 2008).

Die Meldung von den betreffenden Ereignissen findet freiwillig und spontan statt („*spontaneous active event reporting*“, Farley et al. 2008). Im Zusammenhang mit der Erfassung von Unerwünschten Ereignissen müssen die generierenden Verfahren zwar genannt werden, obwohl sie eigentlich zur quantitativen Erfassung nicht geeignet sind, denn sie sind weder sensitiv noch spezifisch und erfassen nur ca. 1% aller UE (Classen et al. 2011, Cullen et al. 1995, Farley et al. 2010, Howell et al. 2015, Murff et al. 2003). Dieser Umstand wurde schon früh als Problem erkannt (IOM 2004, S. 201) und die Diskussion um Erfassungssysteme mit höherer Sensitivität angestoßen. Gelegentlich wurde sogar direkt kritisiert, dass statt einer validen Erhebung nur noch *incident reporting* betrieben würde (Vincent et al. 2014),

Man kann die Diskussion anstoßen, ob nicht auch die Gesamtheit der Erfassungsmethoden, die im Nenner keinen 100%-Ansatz beinhalten, als generierende Verfahren bezeichnet werden sollten (Classen et al. 2016). In Abb. 3.3.-3 sind daher auch Registerstudien (z.B. Schiedsstellen-Statistiken) und *Peer Review* Verfahren auf der Basis nicht-vollständiger Datenquellen mit einbezogen. Die erfassten Ereignisse treten als Aufgreifkriterien auf, lösen also eine Analyse aus (was sehr wichtig ist), aber können nicht

beanspruchen, valide Daten zur Häufigkeit in der Gesamtpopulation zu geben. Schiedsstellendaten können (s. Kap. 3.6.) sehr interessante Informationen z.B. über Ereignisse liefern, an die man nie gedacht hätte (Kennzeichnung des zu operierenden Knies auf dem Thrombosestrumpf), aber eine absolute, auf eine Population bezogene Häufigkeitsangabe erhält man so nicht (Thomas und Petersen 2003).

Eine besondere Rolle spielen in den letzten Jahren *big data*-Analysen, die auf unstrukturierte und sehr umfangreiche Datenmengen zugreifen (s. Kap. 3.3.4.5.). Krankenunterlagen aller Art können auf Signale untersucht werden, die z.B. in pharmakoepidemiologischer Hinsicht Auskunft über Unerwünschte Arzneimittelereignisse aufmerksam machen können (Kush und Goldman 2014, Garbe und Pigeot 2015). Auch können Patienten die Möglichkeit erhalten, über ihre Computer oder Smartphones mit Hilfe von Digitalen Gesundheitsanwendungen (MedApps) Ereignisse mitzuteilen, die als Unerwünschte Ereignisse analysiert werden können (Bahk et al. 2015, Wasson et al. 2007). Auch hier besteht kein Populationsbezug, da ähnlich wie bei CIRS-Meldungen für Häufigkeitsabschätzungen der Bezug fehlt (die Meldefrequenz kann z.B. primär von der Technik-Affinität abhängen), aber zur Detektion von Einzelereignissen, die bei Studien oder in der tagtäglichen Behandlung nicht weiter auffallen, sind diese Instrumente gut geeignet.

Um so wichtiger sind solche Daten aus generierenden Verfahren für die lernende Organisation und das *learning health system* (s. Kap. 2.4.6.3.-5.). Wenn Reason (1998) seine schon zitierte Frage „*Can safer culture be engineered?*“ mit „ja“ beantwortet, dann legt er vier „Subkulturen“ zugrunde: *reporting culture*, *just culture*, *flexible culture* und *learning culture*, allesamt nur als Managementleistung vorstellbar. Inhaltlich geht es um *beliefs* und *values*, Begriffe, die aus dem Zusammenhang der Organisationskultur stammen und hier in Bezug auf Sicherheit verwendet werden (Ashcroft et al. 2005, Gaal et al. 2011) und als Leistung der Organisation zu verstehen sind (Pettigrew 1979). Auch Analysen aus dem anthropologischen Blickwinkel kommen zum gleichen Schluss (Weick 1998). Implizit wird angenommen, dass es durch die Nutzung von *Critical Incident Reporting* Systemen zur Analyse und Korrektur von Prozessen kommt, die sich vor allem auf solche Prozessschritte beziehen, über deren Existenz oder Fehleranfälligkeit vorher keine Informationen vorlagen. Die sog. *Reporting Culture* ist daher auch ein wichtiger Bestandteil der Sicherheitskultur (Reason 1998, Dixon-Woods 2010).

In Kap. 5.3.5. wird konkret auf die Anforderungen an generierende Verfahren eingegangen.

### 3.3.4. Messinstrumente

#### 3.3.4.1. Allgemeines

Die Auswahl der Messinstrumente ist der Zielfestlegung und der Festlegung des Erkenntnisinteresses nachgeordnet (s. Abb. 3.3.-1), es muss also vorweg festgelegt werden, welches Problem zu welchem Zweck und aus welchem Erkenntnisinteresse (wissenschaftlich, Monitoring etc.) untersucht werden soll. Zusätzlich muss die Messinstrumente von der Frage der Datenquelle abgegrenzt werden. Diese Vorgehensweise wird im vorliegenden Weißbuch über das multimodale Konzept von Patientensicherheit abgeleitet (Kap. 2.6.2.), das Patientensicherheit als kontextbezogenes Konstrukt versteht, das durch oft verborgene Einflussfaktoren auf der Ebene der Personen, Teams, Organisationen und Systeme gestaltet wird, und nicht als absolutes *Outcome* gesehen wird, das als Ergebnis eines linearen Prozesses nur „abgezählt“ werden muss (s. Kap. 2.5.3.). Außerdem wird die Messung von Sicherheit selbst als komplexe Intervention verstanden.

Wenn man sich in diesem Licht mit der Messinstrumente zu Patientensicherheit beschäftigt, sind zweierlei Probleme nicht zu übersehen. Zum Einen weisen die unterschiedlichen Methoden in ihren Ergebnissen keine große Überlappung auf, offensichtlich messen sie also nicht den gleichen Gegenstand (Hogan et al. 2008, Weismann et al. 2008). Zum Anderen ist der Eindruck nicht zu vermeiden, dass sich zwischen dem Konzept von Patientensicherheit und der derzeit gängigen Methodik eine große Kluft aufgetan hat („Dissoziation von Konzept und Erhebungsmethodik“, s. Kap. 3.1.), so einfach und voraussetzungsfrei sind die Konstrukte, die dem Messvorgang bislang hinterlegt sind. Beide Punkte wiegen um so schwerer, als dass die mangelnde Reliabilität bzw. Validität der Erhebungsmethoden bzw. Messinstrumente (Shojania und Thomas 2013, Wong et al. 2015) immer wieder als einer der denkbaren Gründe dafür angeführt wird, dass die Wirksamkeit der Verbesserungsbemühungen nicht nachgewiesen werden kann. Da somit eine positive Rückmeldung ausbleibt, lassen sich diese Verbesserungsinitiativen auch nicht nachhaltig implementieren (Baines et al. 2013, Classen et al. 2011, Landrigan et al. 2010, Wachter 2010).

Dabei gibt es schon bei der technischen Umsetzung der Erhebung zahlreiche Argumente, die Vorannahmen beinhalten. Die Darstellung muss hier aus Platzgründen stichwortartig bleiben, deshalb sind hier nur wenige Beispiele aufgeführt:

- **Grundsätzliche Fragen:** Bei der Planung der Erhebung sind einige basale Fragen vorab zu klären:

- **Prävalenz- vs. Inzidenz:** die Prävalenzuntersuchung überschätzt die Häufigkeit. Der Unterschied hängt z.B. von der Verweildauer ab und kann nach der Formel von Rhome und Sudderth (1981) korrigiert werden (Gastmeier et al. 2001).

- **Punkt- vs. Perioden-Prävalenz:** Die Punktprävalenz (zu einem Zeitpunkt erhoben) unterschätzt die Häufigkeit gegenüber der Periodenprävalenz, die das Auftreten von UE während des stationären Aufenthaltes zum Gegenstand hat (Zingg et al. 2014).

- **retrospektiver vs. prospektiver Ansatz:** dieser Punkt ist viel wichtiger als oft angenommen. Zunächst gibt es Anhaltspunkte dafür, dass beide Methoden jeweils einen hohen Prozentsatz unterschiedlicher Ereignisse erfassen (Michel et al. 2004). Natürlich kann die retrospektive Untersuchung mit der Bestimmung der *Odds Ratio* nur in den entsprechenden Datenquellen (z.B. Krankenunterlagen, Abrechnungsdaten) dokumentierte Daten einschließen, während man in der prospektiven Untersuchung frei ist, bekannte Daten aller Art zu erfassen. Die Erfahrung zeigt allerdings, dass hier aus Unerfahrenheit oft zu viele Daten erfasst werden, die dann die Auswertung erschweren.

- Beschreibung als **Proportion oder Rate:** die Rate grenzt sich von der Proportion dadurch ab, dass im Nenner eine andere Einheit als im Zähler erhoben wird als im Zähler (z.B. UE/100 Patienten). Es ist regelmäßig zu klären, wie man mit mehreren Ereignissen pro Bezugsgröße umgeht (z.B. mehrere UE bei einem Patienten - indem man nur das erste Ereignis aufnimmt).

- **Berücksichtigung der Exposition:** Die einfachste Exposition bei UE ist die Behandlungsdauer (z.B. Krankenhausaufenthalt), daher korrigiert man im Krankenhaus über die Verweildauer: UE pro 100 Patiententage. Meist muss man zusätzlich auf das spezifische Risiko korrigieren: Katheterinfektionen auf 1000 Kathetertage.

• **Bedeutung der Datenquelle:** die Wahl der Datenquelle präjudiziert maßgeblich das zu erwartende Ergebnis (s. Kap. 3.3.5.).

• **kurzfristige Verfügbarkeit der Messergebnisse:** für die Auswirkung der Messung ist der Umgang mit den Daten auch in zeitlicher Perspektive entscheidend, denn für den Erfolg von Maßnahmen zur Verminderung von Unerwünschten Ereignissen und zur Verbesserung der Patientensicherheit ist eine valide Quantifizierung mit kurzer Rückmeldelatenz unerlässlich, weil sich sonst kein Lerneffekt einstellen kann (vgl. James 2013, Pronovost et al. 2016).

• **Annahmen über die Reliabilität der Daten:** Abgesehen von den klassischen Zuverlässigkeitskriterien ist der Zugang zu dem Messergebnis (*detection bias*) und der

Einfluss der aufgewandten diagnostischen Maßnahmen zur Identifikation dieser Ereignisse (*surveillance bias*) von größter Wichtigkeit (Bilimoria et al. 2013).

• **Überschätzung der Risikoadjustierung:** Im Bereich der Patientensicherheit stehen Ergebnis-relevante Parameter naturgemäß oft im Mittelpunkt. So weit diese Daten für Vergleiche oder gar für Steuerungszwecke verwendet werden (z.B. P4P), liegt eine Adjustierung über die bekannten Risikofaktoren nahe (zur Methodik s. Agoritsas et al. 2017). Dies kann auch Prozessparameter betreffen (z.B. Impfraten). Vier Punkte, denen im Rahmen der Messung und der Ausgestaltung deren Auswirkungen eine erhebliche Relevanz zukommt, sind in den letzten Jahren zunehmend in den Mittelpunkt des Interesses gerückt:

- **Unvollständig:** eine Risikoadjustierung kann nie alle Ausgangsunterschiede ausgleichen, es verbleibt immer ein Anreiz zur Risikoselektion.

- das **Fallzahl-Prävalenz-Problem:** um so mehr Parameter in die Risikoadjustierungsmodelle einbezogen werden, um so kleiner wird die Gruppengröße und um so höher die Streuung – ein Vergleich verschiedener Einrichtungen wird immer schwieriger (König et al. 2014).

- die **Manipulation der Risikoadjustierungsmodelle:** in den letzten Jahren sind zahlreiche Berichte erschienen, die die Gefahr eines *Upcoding* der Risikofaktoren für risikoadjustierte Ergebnisindikatoren beschreiben (Barringhaus et al. 2011, Gupta et al. 2017A): „... *up-coding of high-risk variables in PCI may occur in public reporting states, which inflates predicted risk and improves risk-adjusted outcomes in the absence of actual improvements in care*“ (Wadhera und Bhatt 2017).

- die **Akzeptanz** der Risikoadjustierung: im Rahmen des *Hospital Readmission Reduction Program* (HRRP) in den USA, das zum Komplex der Krankenhaus-bezogenen P4P-Programme gehört, wird eine Risikoadjustierung nach sozioökonomischen Kriterien durchgeführt, die in Kritik geraten ist, weil damit die tatsächlich schlechtere Versorgung in Regionen mit schlechterer sozioökonomischer Ausgangslage verschleiert und damit die dortige Unterversorgung zementiert wird (Merkow et al. 2015, Fiscella et al. 2014, Sheingold et al. 2016).

Es ist im Rahmen dieses Weißbuch nicht möglich, alle Messinstrumente umfassend darzustellen (Michel et al. 2004, Thomas und Petersen 2003). Eine vergleichende Annäherung an die Vielfalt dieser Methoden ist vom Sachverständigenrat vorgenommen worden (SVR 2008, Nr. 611f). Es sollen daher vier Instrumente herausgegriffen werden:

- (1) direkte Beobachtung
- (2) Befragungen
- (3) Trigger-gestützte Instrumente (einschl. HMPS-Design)
- (4) *Big-Data* Analysen

#### 3.3.4.2. Direkte Beobachtung

Die direkte Beobachtung durch trainiertes Personal erlaubt es, Methoden aus der qualitativen Forschung mit zu integrieren (Leslie et al. 2014), die Sichtweise der Teams und Mitarbeiter am *sharp end* zu erfassen und Aussagen über das Vorliegen von Fehlern und Beinaheschäden und somit zur Vermeidbarkeit von UE zu machen (Flynn et al. 2002). Der Fokus der Erhebung verschiebt sich von der Erfassung von „*work as imagined*“ zu „*work as done*“, man kann also erfassen, was wirklich geschieht und vermeidet eine Präformierung der Untersuchungssituation durch das Beobachtungssetting (Catchpole et al. 2017). Die Häufigkeit der UE ist in diesen auch prospektiv durchzuführenden Studien deutlich höher als z.B. in *Chart Reviews*, so fanden in der ersten derartigen Studie Andrews et al. (1997) *serious adverse events* in 17,7% der Patienten.

Gegenwärtig scheint es zu einem Revival der Studien zu kommen, die die direkte Beobachtung nutzen. Dies kann durchaus mit der größeren Bedeutung der Versorgungsforschung und qualitativen Methoden zu tun haben, geht vor allem aber auf die mangelnde Sensitivität derjenigen Instrumente zurück (s.u.), die mit Abrechnungsdaten oder retrospektiven Krankenaktenanalysen (*Chart Reviews*) arbeiten (Thomas 2015). In einer prospektiven Studie mit einer Kombination von einer zeitnaher Analyse der Krankenunterlagen mit gleichzeitigen Interviews der an der Betreuung der Patienten beteiligten Personen wurde bei 141 Patienten in 10% mindestens ein vermeidbares UE (plus in 23% ein potentielles UE) festgestellt (Wong et al. 2015). Diese Methoden sind natürlich zeitaufwendig, ergeben jedoch ein sehr genaues Bild, und sollten vor allem dann eingesetzt werden, wenn genau dies beabsichtigt ist (z.B. zur Analyse von Risikoschwerpunkten).

Typische Anwendungsfelder sind z.B. Studien zu Medikationsfehlern (Flynn et al. 2002) oder der *Compliance* von Standards bzw. Leitlinien (z.B. Erasmus et al. 2010, Systematischer Review zur Händedesinfektion). Nachteile der direkten Beobachtung liegen im Aufwand und in der Tatsache, dass auch hier Effekte der sozialen Erwünschtheit oder des Hawthorne-Effekts auftreten (Wirkung der Beobachtung selbst).

### 3.3.4.3. Befragungen

#### 3.3.4.3.1. *Patient-Reported Outcome Measures*

In erster Linie richten sich Befragungen an Patienten und an Mitarbeiter, aber in Einzelfällen können auch die Träger der Einrichtungen, die Aufsichtsbehörden und Versicherungen befragt werden. Im Zusammenhang mit der Epidemiologie von UE ist die **Befragung von Patienten** besonders wichtig, vor allem vor dem Hintergrund eines vermuteten Informationsdefizits auf Seiten der Patienten. Letzteres kann als klar widerlegt gelten: Patienten haben ein ziemlich adäquates Bild über die Häufigkeit von Unerwünschten Ereignissen (sie sagen: „Fehler“) in der Gesundheitsversorgung und sind sich dabei mit ebenfalls befragten Ärzten einig (Blendon et al. 2002). Die Daten werden in Kap. 3.6.2.6. eingehend dargestellt.

Diese Befragungsergebnisse entsprechen zunächst der klinisch-epidemiologischen Erkenntnisperspektive. Das *Patient Measure of Safety* (PMOS)-Instrument stellt ein standardisiertes Befragungsinstrument dar, das Patientenangaben über Sicherheit erhebt (McEachan et al. 2014). Das international zusammengesetzte *Patient-Reported Outcomes Safety Event Reporting (PROSPER) Consortium* schlägt den Begriff der **Patient Reported Outcomes of Adverse Events (PRO-AE)** vor, definiert als solche UE, die von Patienten ohne zusätzliche Interpretation durch Angehörige der Gesundheitsberufe berichtet werden:

*“A PRO-AE is any untoward medical occurrence, whether or not considered treatment/intervention related, that is reported or transmitted directly by the patient without interpretation by a clinician or anyone else. PRO-AEs may be collected by both structured and unstructured reports”* (Banerjee et al. 2013).

Die Befragung von Patienten kann eine erhebliche Zahl von zusätzlichen UE erbringen, die mit den klassischen Instrumenten (z.B. retrospektiver Chart Review nach dem HMPS-Design) nicht erkannt werden können (Weismann et al. 2008). Zu den *Patient Reported Outcomes Measures* (PROM) im Sinne von Patientensicherheitsindikatoren wurde bereits in Kap. 3.3.3.3.1. Stellung genommen. Die Ergebnisse von Patienten-Befragungen sind statistisch recht gut mit den Ergebnissen von Mitarbeiter-Befragungen zu sicherheitsrelevanten Ereignissen korreliert (Modak et al. 2007, Raleigh et al. 2009, Sexton et al. 2006).

In der zweiten Linie können PROMs auch als regelrechte Indikatoren eingesetzt werden, die Sicherheit und das Auftreten von UE vorhersagen (s. Vorschlag zum Indikatoren-Set „Patientensicherheit aus Patientenperspektive“ in Kap. 5.3.4.).



### 3.3.4.3.2. Erhebungen der Patientensicherheitskultur

Der zweite wichtige Komplex zum Thema Befragungen sind die **Erhebungen zur Patientensicherheitskultur** (vgl. Schrappe 2016B). Es gibt sehr implizite (Vincent et al. 2014, Pronovost et al. 2006B) und eher explizite Instrumente zur Erhebung der Patientensicherheitskultur. Es lassen sich bei den Instrumenten zur Erfassung der Patientensicherheitskultur folgende Elemente unterscheiden (Literatur in Auswahl):

- **Führung (*leadership*)** wird von fast allen Autoren genannt (z.B. Aarons et al. 2014), häufig konkretisiert als "*supervisory leadership for safety*" (Ginsburg et al. 2014), "*supervisor/manager expectations and actions promoting safety*" (Sorra et al. 2016 (AHRQ)), und "*perceptions of management*" (Sexton et al. 2006), auch interpretiert in Bezug auf "*non-punitive approach*" (Sorra et al. 2016 (AHRQ)) und die Personalausstattung (*staffing*) (Colla et al. 2005). In einer Delphi-Befragung im europäischen LINNEAUS-Projekt zu Patientensicherheitsindikatoren in der Primärversorgung wurde *leadership* an erster Stelle genannt (Frigola-Capell et al. 2012);
- **Organisationslernen** (Halligan und Zecevic 2011) oft in Zusammenhang mit „*continuous improvement*“ (Sorra et al. 2016 (AHRQ));
- **Reporting** für die CIRS-Systeme (Colla et al. 2005, Halligan und Zecevic 2011), dessen Häufigkeit (Sorra et al. 2016 (AHRQ)) und die Nachverfolgung der Meldungen (Ginsburg et al. 2014);
- **Kommunikation** (Sacks et al. 2015) einschließlich der Risikokommunikation im Sinne von „*communication openness*“ (Sorra et al. 2016 (AHRQ)), „*open communication founded on trust*“, „*judgement-free environment*“ und „*job repercussions of error*“ (Halligan und Zecevic 2011, Van Noord et al. 2010) und „*Feedback/communication about errors and organizational learning/continuous improvements*“ (Sorra et al. 2016 (AHRQ), Van Noord et al. 2010).
- **Teamwork**: reflektiert die kleinste organisatorische Einheit der *frontline provider* (Halligan und Zecevic 2011, Hutchinson et al. 2006, Sacks et al. 2015), gelegentlich wird jedoch Teamwork im Sinne der Kooperation unterschieden zwischen „*across hospital units*“ und „*within hospital units*“ (Sorra et al. 2016 (AHRQ), Van Noord et al. 2010) oder unter dem Begriff des „*teamwork climate*“ (Modak et al. 2007, Sexton et al. 2006) zusammengefasst (vgl. 2.4.5. und Kap. 5);
- **Strukturen und Prozesse** im Sinne von „*policies/procedures*“ (Colla et al. 2005) und „*hospital handoffs and transitions*“ (Sorra et al. 2016 (AHRQ), Van Noord et al. 2010);
- **Mitarbeiterperspektive** spielt eine geringe Rolle („*Job Satisfaction, Working Conditions, and Stress Recognition*“ (Modak et al. 2007, Sexton et al. 2006)).

- **Verankerung** als “*shared belief in the importance of safety*” (Halligan und Zecevic 2011) und “*overall perceptions of safety*” (Sorra et al. 2016 (AHRQ), Van Noord et al. 2010);
- **Safety climate**: einen Sonderfall stellen die aus dem SAQ (s.u.) entwickelten Instrumente dar, die das Sicherheitsklima soweit vom Kulturbegriff differenzieren, dass sie das Klima als eigene Dimension der Kultur führen (Hutchinson et al. 2006, Modak et al. 2007, Sexton et al. 2006, Sacks et al. 2015, s. hierzu auch Kap. 2.4.6.6.).

Als **Evaluationsinstrumente** spielen vier Instrumente die wichtigste Rolle (Flin et al. 2006, Halligan et al. 2011, Wischet und Ethinger 2009):

- *Hospital Survey on Patient Safety Culture* (HSOPSC) der AHRQ
- *Safety Attitudes Questionnaire* (SAQ)
- *Patient Safety Culture in Healthcare Organizations Survey* (PSCHO)
- *Modified Stanford Patient Safety Culture Survey Instrument* (MSI) [später weiterentwickelt zum *Canadian Patient Safety Climate Survey* (Can PSCS)]

Der ***Hospital Survey on Patient Safety Culture* (HSOPSC)** der AHRQ ist neben dem SAQ (s.u.) das am meisten verbreitete Instrument (Sorra et al. 2016). Es umfasst 12 Dimensionen (s. Kasten) mit 42 Items. Der HSOPSC ist gut validiert (Chen et al. 2012, Palmieri 2010, Sorra und Dyer 2010) und hat vielfach Anwendung gefunden (Burström et al. 2014, Blegen et al. 2010, Jones et al. 2013, Lee et al. 2016, Thomas und Galla 2013, Verbeek-van Noord et al. 2015). Der HSOPSC wird international häufig eingesetzt (z.B. Deutschland (Pfeiffer und Manser 2010), Niederlande, Belgien, Frankreich, Spanien, Schottland, Libanon, Palästina, Jemen, Äthiopien, China, Taiwan, Japan (Literatur bei Schrappe 2016B), wobei die Übertragbarkeit nicht immer problemlos ist (Perneger et al. 2014, Frankreich, Waterson et al. 2010, Großbritannien). Es gibt das Instrument auch in einer Ausführung für die ambulante Versorgung (Astier-Pena et al. 2015, Parker et al. 2015), für die Pflege (Cappelen et al. 2016, Thomas et al. 2012) und für Krankenhausmanager (Hammer et al. 2011). In den Niederlanden wurde er für die ambulante Versorgung zum Instrument SCOPE (*systematic culture inquiry on patient safety in primary care*) weiterentwickelt (Verbakel et al. 2014 und 2015).

## **HSOPSC: Patient Safety Culture Composites and Definitions**

(Sorra et al. 2016, aus Schrappe 2016B)

### **Patient Safety Culture Composite Definition**

*The extent to which...*

#### **Communication Openness**

Staff freely speak up if they see something that may negatively affect a patient and feel free to question those with more authority.

#### **Feedback and Communication About Error**

Staff are informed about errors that happen, are given feedback about changes implemented, and discuss ways to prevent errors.

#### **Frequency of Events Reported**

Mistakes of the following types are reported: (1) mistakes caught and corrected before affecting the patient, (2) mistakes with no potential to harm the patient, and (3) mistakes that could harm the patient but do not.

#### **Handoffs and Transitions**

Important patient care information is transferred across hospital units and during shift changes.

#### **Management Support for Patient Safety**

Hospital management provides a work climate that promotes patient safety and shows that patient safety is a top priority.

#### **Nonpunitive Response to Error**

Staff feel that their mistakes and event reports are not held against them and that mistakes are not kept in their personnel file.

#### **Organizational Learning—Continuous Improvement**

Mistakes have led to positive changes and changes are evaluated for effectiveness.

#### **Overall Perceptions of Patient Safety**

Procedures and systems are good at preventing errors and there is a lack of patient safety problems.

#### **Staffing**

There are enough staff to handle the workload and work hours are appropriate to provide the best care for patients.

#### **Supervisor/Manager Expectations and Actions Promoting Patient Safety**

Supervisors/managers consider staff suggestions for improving patient safety, praise staff for following patient safety procedures, and do not overlook patient safety problems.

#### **Teamwork Across Units**

Hospital units cooperate and coordinate with one another to provide the best care for patients.

#### **Teamwork Within Units**

Staff support each other, treat each other with respect, and work together as a team.

Der **Safety Attitudes Questionnaire (SAQ)** ist das zweite wichtige Instrument (Sexton et al. 2006), entwickelt aus dem *Intensive Care Unit Management Attitudes Questionnaire*, der wiederum auf dem *Cockpit Management Attitudes Questionnaire* basiert, also aus der Luftfahrt stammt (Kho et al. 2005). Der SAQ (s. Kasten) wird breit im stationären und auch im ambulanten Bereich eingesetzt (Palmieri 2010, Nguyen et al. 2015). Es gibt eine Fassung für Intensivstationen (Huang et al. 2007) und andere Fachgebiete (Übersicht z.B. bei Palmieri 2010). In Deutschland basiert die Entwicklung des Frankfurter Fragebogens für Hausärzte auf diesem Konzept (Hoffman et al. 2009). Die **Übereinstimmung zwischen HSOPSC und SAQ** ist relativ gut (Etchegaray und Thomas 2012).

<b>Safety Attitudes Questionnaire (SAG)</b> (Sexton et al. 2006, aus Schrappe 2016B)	
<b>Dimensions</b>	<b>Definition</b>
1. Teamwork climate	Perceived quality of collaboration between personnel
2. Job satisfaction	Positivity about work experience
3. Perceptions of management	Approval of management actions
4. Safety climate	Perception of strong and proactive organizational commitment to quality
5. Working conditions	Perceived quality of the work environment and logistical support
6. Stress recognition	Acknowledgement of how performance is influenced by stressors

Das **Patient Safety Culture/Climate in Healthcare Organizations (PSCHO)** Instrument wurde in Kalifornien vom *Center of Health Police* und dem *Primary Care Outcomes Research Center (CHP/PCOR)* in Stanford und dem *Patient Safety Center of Inquiry* in Palo Alto entwickelt (Singer et al. 2003). Ausgehend von einem einfachen Konzept (Kaskade: Organisation – Unit – Person – Outcome) wurde das Erhebungsinstrument in 105 Krankenhäusern in den USA validiert (Singer et al. 2007) (s. Kasten). In zwei wichtigen Studien wurde die Patientensicherheitskultur mit Routinedaten-gestützten Patientensicherheitsindikatoren in Zusammenhang gebracht (s. Kap. 6.2.5.) (Singer et al. 2009A), und es wurden Patientensicherheitskultur mit der Organisationskultur in Beziehung gesetzt (s. Kap. 5.3.2.) (Singer et al. 2009B).

Aus dem *Modified Stanford Patient Safety Culture Survey (MSI)* wurde der **Canadian Patient Safety Climate Survey (Can-PSCS)** entwickelt (Ginsburg et al. 2014), der 19 Items in sechs Dimensionen beinhaltet:

- (1) *organisational leadership support for safety*
- (2) *incident follow-up*
- (3) *supervisory leadership for safety*
- (4) *unit learning culture*
- (5) *enabling open communication I: judgement-free environment*
- (6) *enabling open communication II: job repercussions of error*

<p><b><u>Patient Safety Culture/Climate in Healthcare Organizations</u></b> (PSCHO) (Singer 2009B, aus Schrappe 2016B)</p>	
<b>Organizational Dimensions</b>	<ul style="list-style-type: none"> <li>Senior management engagement</li> <li>Organizational resources for safety</li> <li>Overall emphasis on safety</li> </ul>
<b>Work-Unit Dimensions</b>	<ul style="list-style-type: none"> <li>Unit safety norms</li> <li>Unit recognition and support for safety</li> </ul>
<b>Interpersonal Dimensions</b>	<ul style="list-style-type: none"> <li>Fear of shame</li> <li>Fear of blame</li> </ul>
<b>Other</b>	<ul style="list-style-type: none"> <li>Provision of safe care</li> <li>Group culture</li> <li>Entrepreneurial culture</li> <li>Production-oriented culture</li> </ul>

Weiterhin ist die **Safety Organizing Scale** (SOS) zu nennen, die relativ verhaltensorientiert und offen gestaltet ist, wie bereits aus den genutzten Dimensionen zu erkennen ist (n. Palmieri 2010)

1. *Preoccupation with failure*
2. *Reluctance to simplify*
3. *Sensitivity to operations*
4. *Commitment to resilience*
5. *Deference to expertise*

Ähnlich offen geht das **Manchester Patient Safety Framework (MaPSaF)** vor. Es bezieht sich auf die Arbeit von Kirk et al. (2007) (s. Kasten) und steht in der Tradition der eher qualitativen Fragenkataloge von Vincent et al. (2014) (s. Kap. 4.3.3.) und Pronovost et al. (2006B) (s. Kap. 6.3.1.). Die neun Kriterien (s. Kasten) wurden in Nordengland in Delphitechnik bei Personen aller Berufsgruppen, die in der ambulanten Gesundheitsversorgung tätig waren, erarbeitet und validiert. Theoretisch geht dieses Instrument auf die Arbeiten von Westrum (2004) und Parker und Hudson (2001) zurück,

die Organisationen mit Schwerpunkt auf der Behandlung von Information in drei bzw. fünf Gruppen differenzieren. Ziel des Instrumentes ist weniger die Erhebung eines quantitativen Maßes, sondern es soll ein Verfahren zur Verfügung gestellt werden, das es Organisationen erlaubt, sich in einer Art Selbstbewertung über ihr Sicherheitsverhalten Rechenschaft abzulegen. Es wurde in mehreren Bereichen des Gesundheitswesen erfolgreich eingesetzt (Ashcroft et al. 2005, Mannion et al. 2009, Verbakel et al. 2015B, Wallis und Dovey 2011) und in einem Konsensus-Prozess innerhalb der LINNEAUS-Collaboration von allen Instrumenten für den ambulanten Bereich am höchsten bewertet (Parker et al. 2015).

### **Dimensions of patient safety culture in primary care**

(Kirk et al. 2007, aus Schrappe 2016B)

#### **Overall commitment to quality**

How much is invested in developing the quality agenda?  
What is seen as the main purpose of policies and procedures?  
What attempts are made to look beyond the organisation for collaboration or innovation?

#### **Priority given to patient safety**

How seriously is the issue of patient safety taken within the organisation?  
Where does the responsibility lie for patient safety issues?

#### **Perceptions of the causes of patient safety incidents and their identification**

What sort of reporting mechanisms are there?  
How are reports of incidents viewed?  
As an opportunity for blame or improve?

#### **Investigating patient safety incidents**

Who investigates incidents and how are they investigated?  
What is the aim of the organisation?  
Does the organisation learn from the event?

#### **Organisational learning following a patient safety incident**

What happens after an incident?  
What mechanisms are in place to learn from the incident?  
How are changes introduced and evaluated?

#### **Communication about safety issues**

What communication systems are in place?  
What are their features?  
What is the quality of record keeping to communicate about safety?

#### **Personnel management and safety issues**

How are safety issues managed in the workplace?  
How are staff problems managed?  
What are recruitment and selection procedures like?

#### **Staff education and training about**

How, why and when are education and training programmes about safety issues patient safety developed?

#### **Teamworking around safety issues**

How and why are teams developed?  
How are teams managed?  
How much teamworking is there around patient safety issues?

#### 3.3.4.4. Trigger-gestützte Instrumente

Der Prototyp der Trigger-gestützten Messinstrumente ist das **Harvard Medical Practice Study-Design (HMPS-Design)**, das erstmals in der *California Medical Insurance Feasibility Study* (Mills 1978) eingesetzt und sich in der Folge über die HMPS-Studien (Brennan et al. 1991, Leape et al. 1991, Localio et al. 1991) und die *Utah-Colorado* Studie (Thomas et al. 2000) zu einem der wichtigsten Standards der epidemiologischen Untersuchungen in großen (meist nationalen) Kohorten entwickelt hat (z.B. Rafter et al. 2017). Es basiert auf einem zweistufigen retrospektiven *Chart Review* (Analyse der Krankenhausakten). Die genannten, in den USA durchgeführten Studien zeigten mit Ereignishäufigkeiten von UE bei Krankenhauspatienten zwischen 2,9 - 3,7% die geringsten Werte. Sie unterschieden sich von späteren Studien in methodischer Hinsicht durch ihren haftungsrechtlichen Kontext und ein fehlendes Screening auf nosokomiale Infektionen. Weiterhin waren die Beobachtungszeiträume unterschiedlich definiert, Häufigkeiten wurden teilweise als Punkt- und teilweise als Periodenprävalenzen abgebildet und die Ergebnisse wurden nicht einheitlich berichtet (Lessing et al. 2012).

Ganz entscheidend ist der positive prädiktive Wert der ersten Stufe (Screening mittels Kriterienkatalog), der in den frühen Studien mit 21% sehr niedrig war (Brennan et al. 1989, s. auch Bates et al. 1995C). Außerdem spielt die Konkordanz der ärztlichen Experten in der zweiten Stufe eine Rolle, die in den frühen Studien mit 10,0% sogar noch unterhalb der Dissensrate (12,9%) lag (Localio et al. 1996). In späteren Studien mit einem höheren Anteil von UE wurde eine intensiviertere Schulung der ärztlichen Gutachter in der zweiten Analysestufe vorgenommen (z.B. Baines et al. 2013). In einem sehr wichtigen Systematischen Review wird die Reliabilität als hoch eingeschätzt, die Validität (z.B. gemessen an einem externen Kriterium) ist allerdings nicht untersucht (Hanskamp-Sebregts et al. 2016). In neuerer Zeit ist die Diskussion um die Wertigkeit des HMPS-Designs sehr aktuell geworden, als die dritte niederländische UE-Studie nach dem HMPS-Design wiederum keine durchgreifende Verbesserung der Häufigkeit von UE zeigte (Zegers et al. 2009, Baines et al. 2013, 2015) und man die Hypothese diskutierte, dass dies (auch) an der Erhebungsmethode liegen könnte (Wong et al. 2015, Shojania und Marang-van der Mheen 2015).

Man hatte früh begonnen, die erste Stufe der Analyse (das sog. Screening) durch Trigger mit einer höheren Sensitivität zu verbessern, die sich gut für das Training der Reviewer eignen (Classen et al. 2011). Durch den Trigger werden Krankenakten identifiziert, die dann einer weiteren Auswertung zugeführt werden. Das bekannteste Instrument ist das sog. **Global Trigger Tool (GTT)** des *Institute of Healthcare Improvement* (IHI 2018), das 53 einzelne Trigger umfasst und in vier großen Studien erprobt wurde (Department of Health and Human Services 2008, 2010, Landrigan et al. 2010, Classen et al. 2011, s.

auch Systematischer Review von Hanskamp-Sebregts et al. 2016). Es wird eine Einteilung der UE nach Schweregrad vorgenommen, wobei Grad F (*temporary harm, requiring a prolonged hospital stay*), Grad G (*permanent harm*), H (*life-sustaining intervention required*) und I (*contributing to death*) die wichtigsten sind. Es wurde in abgewandelter Form auch in der Pädiatrie angewandt (Stockwell et al. 2015). Die Zahl der festgestellten UE stieg dadurch rapide an und lag deutlich über den Ergebnissen anderer Methoden. (z.B. PSI-Set der AHRQ, s. Zusammenstellung bei James 2013). Solche Trigger-gestützten Verfahren sind auch in der ambulanten Versorgung (de Wet et al. 2017, kritisch: Davis et al. 2018) oder im lokalen Umfeld (z.B. Intensivstation) einsetzbar und erlauben hier bereits die Evaluation von Interventionen (Jain et al. 2006).

Eine weitere Möglichkeit, die Sensitivität der ersten Analysestufe zu verbessern, bestand natürlich in der **Computer-Unterstützung**. Bereits im Jahr 1991 war in *Salt Lake City* ein Computer-gestütztes Erfassungssystem für *adverse drug events* bei der Antibiotika-Therapie entwickelt worden (Evans et al. 1991), das an knapp 80.000 Patienten eingesetzt wurde (Evans et al. 1994) und sogar zu einer deutlichen Reduktion dieser Ereignisse führte (Evans et al. 1998). Im gleichen Jahr veröffentlichten Ashish Jha und David Bates eine vergleichende Untersuchung zur Detektion von *adverse drug events*, die zwar zeigte, dass Computer-alerts gegenüber den *chart reviews* weitere Informationen beisteuern, diese aber nicht untereinander deckungsgleich sind (Jha et al. 1998). Ein ähnliches Ergebnis fanden Samore et al. (2004) bei der Untersuchung von *Adverse Medical Device Events* (AMDEs, ebenfalls *Salt Lake City*, s. Kap. 5.6.2.1.), die nicht nur eine unterschiedliche Häufigkeit bei der Nutzung von CIRS (1,6/1000 Pat.), Computer-alerts (27,7) und Abrechnungsdaten (64,6) ergab, sondern auch zeigte, dass zwischen den Methoden keine Deckungsgleichheit besteht. Auch in einer anderen Studie, in der das *Global Trigger Tool* mit den Ergebnissen aus den elektronischen Krankenakten verglichen wurde, waren nur 46 der insgesamt 183 entdeckten UE mit beiden Methoden gleichzeitig entdeckt worden (O'Leary et al. 1998). Da in der ersten Stufe des HMPS-Design eine hohe Übereinstimmung vorausgesetzt werden muss, ist dieser Weg nicht weiterverfolgt worden.



#### 3.3.4.5. *Big Data*: Die Analyse unstrukturierter Daten

Eines der wichtigsten Merkmale von *Big Data*-Analysen liegt in der Verwendung von unstrukturierten Daten (Schrappe 2016C, Vollmar et al. 2017), die in verschiedenen Formen vorliegen können (s. Kap. 3.3.5.). Nicht alle unter diesem Begriff der Unstrukturiertheit zu fassenden Daten prägen diese Eigenschaft in gleicher Weise aus, Behandlungsunterlagen weisen z.B. gerade dann eine Struktur auf, gerade wenn sie digital als elektronische Krankenakten erhoben oder in diese überführt wurden. Die letztere Form hat jedoch den Vorteil, ein anderes Charakteristikum von *Big Data* zu erfüllen, nämlich die große Datenmenge (*volume*), die zur Analyse bereit steht (Kush und Goldman 2014).

Die Analyse großer Mengen unstrukturierter Daten können aus zweierlei Blickwinkeln vorgenommen werden, aus klinisch-epidemiologischer Sicht (s. Kap. 3.3.3.2.) und aus der Sicht der generierenden Verfahren (s. Kap. 3.3.3.4.). Im ersten Fall bedarf es einer Falldefinition und der Definition von Einflussfaktoren (z.B. bekanntes unerwünschtes Arzneimittelereignis und Exposition durch die Einnahme dieses Medikamentes, Hripcsak et al. 2003). In der Diskussion ist der Einsatz von selbstlernenden Analyseinstrumenten (Künstliche Intelligenz), die solche Zusammenhänge erkennt, ohne dass eine solche Hypothese vorgegeben wird (Binder und Blettner 2015). Das *Institute of Medicine* machte bereits im Jahr 2004 auf die Option des *Data Mining* aufmerksam (IOM 2004, S. 206):

*„Free-text searching of the electronic medical record is used to identify for each medication taken an indication of its known adverse reactions. For the drug type “diuretic,” fatigue is a potential adverse reaction and “drowsiness,” “drowsy,” and “lassitude” are some of the synonyms used instead of the word “fatigue”.*“

Eine Strukturierung des Vorgehens kann aber nicht vollständig unterbleiben, denn die Faktoren müssen in irgendeiner Form benannt werden (Schulz und Lopez-Garcia 2015); insofern ist die Annahme einer *End of Theory* irreführend (Anderson 2008). Es ist völlig richtig darauf hinzuweisen, dass die Gefahr besteht, bei einer kritiklosen Anwendung solcher Methoden zu Pseudo-Korrelationen und falschen Zusammenhängen zu kommen, die den Patienten mehr Schaden zufügen als dass sie Schaden verhindern (Antes 2018). „Zutreffend jedenfalls ist, dass Big Data Hypothesen und Modelle nicht verwendet und Kausalitätsüberlegungen nicht einbezieht. Aus Sicht der Wissenschaft ist Big Data daher ein rein exploratives datengesteuertes Verfahren mit mäßigem Evidenzgrad“ (Wegscheider und Koch-Gromus 2015).

Im zweiten Fall, der Nutzung als generierendes Verfahren, kann der *Big Data*-Ansatz jedoch durchaus eine wichtige Informationsquelle darstellen. Analog zum CIRS im institutionellen Kontext besteht die Möglichkeit, die von Patienten über ihre persönlichen

digitalen Instrumente zur Verfügung gestellten Informationen zur Verbesserung der Patientensicherheit einzusetzen (Garbe und Pigeot 2015). Der wichtigste Grund besteht darin, dass man den Weg bis zur Etablierung einer klinisch-epidemiologischen Falldefinition abkürzen kann, denn hier sind mehrere Schritte vorgeschaltet (Ereignis muss relativ häufig sein, damit sie erkannt und eine Falldefinition erstellt werden kann). Allerdings setzt dies voraus, dass man ähnlich wie beim CIRS Ressourcen dafür einsetzt, die Warnhinweise zu systematisieren und auszuwerten (Bahk et al. 2015, Wasson et al. 2007).

Wenn man also *Big Data*-Analysen als Messinstrumente betrachtet, dann greift sie in starkem Maß auf *Patient Reported Outcomes* zurück (IOM 2012, S. 45). Wie in Kap. 3.3.4.3. bereits ausgeführt, setzen sich PROM bzw. PRO-AE sowohl aus strukturierten als auch aus unstrukturierten Daten zusammen (Banerjee et al. 2013).

### 3.3.5. Datenquellen

Nach Klärung der Ziele einer Erhebung, des Erkenntnisinteresses und der Messinstrumente muss man als letzten Schritt die Datenquellen identifizieren und kritisch bewerten, mit denen man arbeiten möchte. Auf einige Aspekte wurde schon eingegangen:

- **Krankenakten:** Die Krankenakten-Analyse durch interne oder externe Reviewer ist die Goldstandard-Methode und wurde in den historischen Studien nach dem *Harvard Medical Practice Study* (HMPS)-Design angewandt, wobei ein mehr oder weniger explizites Trigger-Instrument vorgeschaltet wurde (s. Kap. 3.3.4.4.).
- Auf die ***Patienten-Reported Outcome Measures*** bzw. die PRO-AE wurde unter dem Thema „Befragungen“ bereits in Kap. 3.3.4.3. eingegangen. Diese Daten stellen eine eigene Datenquelle dar, die gerade auch im Zusammenhang mit *Big Data* noch an Bedeutung gewinnen wird.
- ***Big Data*:** Obwohl der Begriff es eigentlich nahelegt, dass es sich um eine spezielle Art der Datenquelle handelt, liegt der Schwerpunkt ganz eindeutig auf der Bedeutung als Messinstrument, denn der Umgang mit den drei „V“ (*variety* für Unstrukturiertheit, *volume* für die große Datenmenge und *velocity* für rasche Ausweitung) steht im Vordergrund. Diese drei Faktoren wirken zusammen, weil bei der hohen *Noise-Signal Ratio* (Rauschen) unstrukturierter Daten und einer raschen quantitativen Zunahme der Datenmenge nicht zu erwarten ist, dass sich das Rauschen vermindert, sondern es wird exponentiell zunehmen (Psaty und Breckenridge 2014). Trotzdem wird dieser (generierende) Zugang an Bedeutung zunehmen, und aus Sicht der Patientensicherheit muss man sich damit

beschäftigen. Allerdings sollte streng beachtet werden, ob sich paradoxe Effekte einstellen (damit ist zu rechnen).

Auf zwei Themenbereiche soll hier ausführlicher eingegangen werden:

- Routine- bzw. Abrechnungsdaten und
- Kombinationsmethoden.

Die **Routine- bzw. Abrechnungsdaten** stellen ein ganz entscheidendes Thema dar, denn z.B. das AHRQ-PSI-Set basiert auf Abrechnungsdaten, und auch in Deutschland haben Gesetzgeber, Gemeinsamer Bundesausschuss und Kostenträger, ausgelöst durch den Bürokratie-Vorwurf, einen starken Hang zu Routinedaten. Um einen Eindruck davon zu erhalten, was Abrechnungsdaten können und wozu sie nicht geeignet sind, sei hier ein Ausschnitt der klassischen Veröffentlichung von Patrick Romano et al. aus dem Jahr 2003 wiedergegeben (Tab. 3.3.-5), die mit einer frühen Form des AHRQ PSI-Set (vgl. Tab. 3.3.-3 in Kap. 3.3.3.3.) an mehr als einer Million Patienten aus 1000 Krankenhäusern durchgeführt wurde:

PSI	n=	Rate	Bemerkung
<i>Death in low mortality DRG</i>	5912	0,0043%	Alle Patienten
Belassener Fremdkörper	2284	0,024%	Nur operierte Patienten
Nosokomiale Infektion	54490	0,193%	Alle Patienten

Tab. 3.3.-5: Auszug aus Romano et al. (2003) zu den PSI Nr. 2 und 5 (nosokomiale Infektionen sind heute nicht mehr enthalten)

Diese Ergebnisse zeigen (vgl. auch Zhan und Miller 2003, Rivard et al. 2008, Carey et al. 2010, vgl. Kap. 4.3.)

- die Stärke von Abrechnungsdaten: die Dokumentation von sehr seltenen, abrechnungsrelevanten *Sentinel Events* (belassene Fremdkörper im OP-Gebiet),
- eine weitere Stärke von Abrechnungsdaten, sie können nämlich hervorragend als Aufgreifkriterium für nachfolgende Analysen (z.B. *Peer Review*) dienen (z.B. *death in low mortality DRG*), stellen jedoch kein quantitatives Maß für diese Ereignisse dar, und

- die entscheidende Schwäche von Abrechnungsdaten: sie sind außerordentlich insensitiv bzgl. der wichtigsten Komplikationen, z.B. nosokomialen Infektionen, die hier mit rund 0,2% ungefähr um den Faktor 20 unterschätzt werden (4% müssen angenommen werden).

Dieser Befund zieht sich durch die gesamte Literatur (eine Zusammenstellung s. Schrappe 2015, S. 102ff). So wurden in der sehr bekannten Untersuchung von Classen et al. (2011) mit dem *Global Trigger Tool* (s. Kap. 3.3.3.4.) 354/393 UE detektiert, mit dem AHRQ-Set jedoch nur 35 (und mit CIRS 4).

In einer eigenen Untersuchung, in der trainierte Untersucher (geblindet, Untersuchung pilotiert) in den Abrechnungsdaten und parallel in den Krankenakten von 3000 zufällig ausgewählten Patienten in drei Krankenhäusern verschiedener Versorgungsstufe nach UE suchten, wurde ebenfalls das *Underreporting* durch die Abrechnungsdaten bei nosokomialen Infektionen (i.v. Katheter, Pneumonie, postoperative Wundinfektionen), außerdem beim sekundären Nierenversagen und beim „nosokomialen“ Myokardinfarkt festgestellt (Maass et al. 2015, gefördert durch das Bundesministerium für Gesundheit, s. Tab- 3.3.-6).

<b>Patient Safety Indicator</b>	<b>at risk (n=)</b>	<b>Chart Review</b>	<b>DRG</b>
Dekubitus Grad II-IV	2374	71	47
Dekubitus Grad III-IV	2374	11	22
i.v. Katheter-Infektion	2090	32	2
Postop. Respiratorisches Versagen	221	3	4
Postop. tiefe Beinvenenthrombose	1498	8	6
Nosokomiale Pneumonie	2876	90	23
Sekundäres akutes Nierenversagen	2907	170	53
Myokardinfarkt 12h nach Aufnahme	2917	24	5
Postoperative Wundinfektion	1413	58	31

Tab. 3.3.-6: Vergleich von *Chart Review* (Analyse der Krankenakten) und Abrechnungsdaten (DRG) in drei deutschen Krankenhäusern (3000 Patienten) (aus: Maass et al. 2015, gefördert durch das Bundesministerium für Gesundheit)

Die wissenschaftliche Literatur zum Thema der Abrechnungsdaten ist außerordentlich umfangreich und sollte in Deutschland evtl. noch einmal durch einen Systematischen Review aufgearbeitet werden (s. Kap. 5.3.8.); sie kann hier aus Platzgründen nicht

vollständig dargestellt werden. Unter dem Strich bleibt kein Zweifel, dass Routine- bzw. Abrechnungsdaten primär die Abrechnungslogik und nicht die Qualitäts- oder Patientensicherheitsthematik abbilden, und dies wird international durch die bekanntesten Autoren auf diesem Gebiet auch immer wieder betont (z.B. Austin, McGlynn und Pronovost 2016). Gerade bei einem Wechsel in der Dokumentationsstrategie z.B. von freiwilliger Meldung zur abrechnungsbezogenen Meldung wie bei den Dekubitalgeschwüren wird dies schlaglichtartig deutlich, wenn man die höhergradigen Dekubitusraten bei jüngeren Patienten in der freiwilligen Dokumentation (niedrig) und in den Abrechnungsdaten (hoch) vergleicht, denn diese sind gerade bei jungen Patienten wegen der nicht so häufigen Mehrfachdiagnosen abrechnungsrelevant (Darstellung in Schrappe 2015 S. 104).

Abschließend soll noch auf die **Kombinationsmethoden** eingegangen werden, die mehrere Datenquellen gleichzeitig nutzen. Wong et al. (2015) kombinierten die zeitnahe Analyse der Krankenunterlagen mit Interviews der an der Behandlung Beteiligten (s. Editorial von Thomas 2015). Am wichtigsten ist zweifelsohne der Ansatz des *Medicare Patient Safety Monitoring System* (MPSMS), das bereits im Kapitel über die PSI-Sets dargestellt wurde (Kap. 3.3.3.3.). Dieses breit eingesetzte Instrument basiert auf einer Kombination von administrativen Daten mit ein extern durchgeführten Krankenaktenanalyse (Hunt et al. 2015, Classen et al. 2016). Solchen Kombinationsinstrumenten wird die Zukunft gehören, und man sollte dringend Versuche unternehmen, eine solche Methodik auch in Deutschland zu implementieren.

### 3.3.6. Zusammenfassung und erste Empfehlung

Es kann kein Zweifel daran bestehen, dass Auseinandersetzungen über die adäquate Erhebungsmethodik in einem höchst widersprüchlichen und interessengeleiteten Umfeld stattfinden. Bei der Planung und Interpretation von Messungen ist daher ein transparentes Vorgehen von größter Wichtigkeit. Ein **klares Konzept der Darstellung** geht von folgenden vier Schritten aus (vgl. auch Kap. 5.3.8.), im Einzelnen der

- Bestimmung des Ziels der Erhebung, gefolgt von
- dem Erkenntnisinteresse,
- der Messinstrumente und
- der Wahl der verwendeten Datenquellen.

Diese Aufzählung entspricht auch dem praktischen Vorgehen bei der Planung und Durchführung entsprechender Untersuchungen, denn die Wahl des Blickwinkels bedingt die Wahl des Messinstrumentes und führt bei der Auswahl der zur Verfügung stehenden

Datenquellen. Dieses Vorgehen widerspricht der häufig verwendeten Praxis, dass man zunächst die vorhandenen Datenquellen („was haben wir denn schon?“) betrachtet und daraus zunächst die Methodik und zuletzt die zu beantwortenden Fragestellungen ableitet. Diese Aussage ist in hohem Maße „politisch“, denn ein Datenverfügbarkeits- und Methoden-orientiertes Herangehen hat zwei Konsequenzen:

- zahlreiche, oft die relevantesten Fragestellungen werden von vorneherein ausgeschlossen,
- die Untersuchungen führen zu vorhersehbar negativen Ergebnissen, wenn z.B. irrelevante Datenquellen oder invalide Erhebungsmethoden verwendet werden.

Das typische Beispiel hierfür ist die Erhebung von Abrechnungsdaten, mit denen nicht nur die UE-Raten in den meisten Fällen deutlich unterschätzt werden, sondern auch durch Umsetzungsfaktoren wie z.B. die langen Rückkopplungslatenzen effektive Lernschritte am *sharp end* unmöglich gemacht werden (s. Classen et al. 2011, s. Kap. 3.3.5.).

Sinnvoll ist also ein **Ziel-orientiertes Verfahren**, das die Auswahl der relevanten Problematik an den Anfang stellt und danach in erster Linie den Blickwinkel klärt, unter dem diese Problematik untersucht werden soll, so dass man in der Folge die Erhebungsmethodik und adäquate Datenquellen auswählen kann (s. Kasten) . Zum Thema der Ziel-Orientierung ist auf Kap. 2.3.3. zu verweisen, im Mittelpunkt steht die Überwindung des normativen, „sich selbst erklärenden“ Verständnisses des Themas Patientensicherheit – Themen zur Verbesserung der Patientensicherheit müssen in Konkurrenz zu anderen Zielen priorisiert und begründet werden, und auch innerhalb des Themenspektrums muss eine zielorientierte Auswahl und Schwerpunktsetzung erfolgen. Hinsichtlich der Zielorientierung sind zu unterscheiden

- die **Perspektivdimension**: analog zur gesundheitsökonomischen Methodik werden 7 Perspektiven unterschieden (z.B. Patienten- oder Anbieterperspektive);
- die **Strukturdimension**: Weiterentwicklung oder Hemmung der wichtigsten strukturellen Systemprobleme;
- die **Bedarfsdimension**: die Frage, ob die prävalenten Probleme der Morbidität adressiert werden (z.B. Demographie, chronische Erkrankungen).

Ist dieser Vorgang abgeschlossen, kann und muss man sich mit der Wahl dem Erkenntnisinteresse der Erhebung befassen, also mit der Frage beschäftigen, welchen Aufgaben die Beobachtung bzw. Erhebung von Daten zur Beschreibung des Problems erfüllen soll, in Einzelnen

- ob es sich um eine **wissenschaftliche Untersuchung** z.B. im Rahmen einer Evaluationsstudie handelt, die unter Kontrolle der Umgebungsfaktoren stattfindet und komplexe Messvorgänge mit höchstmöglicher Sensitivität und Spezifität umfasst („*scientific pathway*“, s. Kap. 3.3.3.1.),

- ob es sich um eine gezielten Erfassung unter Zuhilfenahme von **klinisch-epidemiologischen Falldefinitionen** handelt, die valide Daten mit hinreichender Sensitivität und Spezifität für bestimmte Problembereiche liefert, so wie es im Bereich der *Infection Control*-Ansätze gehandhabt wird („*epidemiological pathway*“, s. Kap. 3.3.3.2.),
- ob es sich um einen **Monitoring-Ansatz** handelt, der große Versorgungsbereiche mit sensitiv eingestellten Methoden (unter Verzicht auf hohe Spezifität) beobachtet, so wie es durch (Patientensicherheits-)Indikatoren geschieht („*monitoring pathway*“, s. Kap. 3.3.3.3.), oder
- ob es um die Behebung von „doppelt-blinden Flecken“ (*Unknown Unknowns*) handelt, so wie es **generierende Verfahren** (z.B. CIRIS, zum Begriff s. Schrappe 2015, S. 80) leisten können, die gerade im Zusammenhang der Thematik Patientensicherheit eine so große Bedeutung erlangt haben (sog. „*detection loop*“, s. Kap. 3.3.4.4.).

Der wichtigste Punkt besteht hier in der Unterscheidung zwischen klinisch-epidemiologischen Falldefinitionen und Patientensicherheitsindikatoren (PSI). PSI sind nicht zur quantitativen Erfassung geeignet, sondern zur Beobachtung und möglichst vollständigen Identifikation Unerwünschter Ereignisse. Die Validität von PSI liegt daher in der Vorhersagefunktion, während klinische Falldefinitionen das valide erkennen müssen, was sie zählen sollen. Insgesamt ist bei der Thematik der PSI zu erkennen, wie fortgeschritten die Dissoziation zwischen dem Konzept von Patientensicherheit, das sich in der internationalen Diskussion längst von der Ansicht losgelöst hat, dass Patientensicherheit das (Nicht-)Vorliegen von Unerwünschten Ereignissen wiedergeben würde (s. auch Kap. 2.6.2.), und den verwendeten Erhebungsmethoden ist. In der Erhebung werden immer noch „Indikatoren“ verwendet, die Patientensicherheit als Komplikationsrate oder Häufigkeit von UE verstehen. Dies wird auch in den PSI-Sets deutlich (z.B. PSI90-Set der AHRQ), die gleichwohl in Kap. 3.3.3.3. ausführlich dargestellt werden. Am fortgeschrittensten ist der MPSMS-Indikatorenset, der auch die konservative Medizin mit einschließt und zumindest von der Zielorientierung her breiter aufgestellt ist. Allerdings beinhaltet das MPSMS keine Angaben zu Vermeidbarkeit; die Weiterentwicklung zum *Quality and Safety Reporting System* (CSRS) ist daher mit Interesse zu verfolgen.

Bezüglich der **Messinstrumente** sind neben einige allgemeinen Fragen wie der Verwendung retrospektiver vs. prospektiver Designs, der Überschätzung der Risikoadjustierung etc. besonders

- die direkte Beobachtung,
- die Befragungen (*Patient Reported Outcomes*, Befragungen der Mitarbeiter zur Patientensicherheitskultur etc.),

- die Trigger-gestützten Messinstrumenten wie z.B. das *Harvard Medical Practice Study*-Design und das *Global Trigger Tool*, und
- die Methode der auf unstrukturierte Daten zugeschnittenen *Big Data*-Analyse zu unterscheiden.

Bei den **Datenquellen** muss an sich vor allem mit der Rolle der Abrechnungsdaten auseinandersetzen. Zahlreiche international und auch in Deutschland durchgeführt, vergleichende Untersuchungen zeigen deutlich, dass Abrechnungsdaten ein entscheidendes Defizit in der Sensitivität aufweisen, so dass sie nur in Ausnahmefällen (z.B. abrechnungsrelevanter *sentinel events*) sinnvoll zur Erhebung von Unerwünschten Ereignissen (durch klinisch-epidemiologische Falldefinitionen) oder gar als Indikatoren zu verwenden sind, wo ja besonders die Sensitivität gefordert ist.

Es ist aus dieser Sicht klar zu erkennen, dass die Zukunft in Kombinationsmethoden liegt, in denen Krankenaktenanalysen und (kritisch) Abrechnungsdaten kombiniert werden und man vermehrt auf *Patient Reported Outcomes* zurückgreift. Das *Medicare Patient Safety Monitoring System* (MPSMS) der *Centers of Medicare and Medicaid Services* (CMS) der USA ist derzeit das attraktivste Konzept, das in diese Richtung weist, und sollte in einer angepassten Form in Deutschland pilotiert und eingesetzt werden. Eine konkrete Empfehlung für die weitere Ausgestaltung wird in Kap. 5.3. vorgenommen (s. Kap. 5.3.8.).

#### **Zur Erhebungsmethodik: Standardisiertes Vorgehen empfohlen!**

An dieser Stelle wird zum ersten Mal im Weißbuch zusammenfassend zur Erhebungsmethodik Stellung genommen. Das Ziel dieses einen „Roten Fadens“ im Weißbuch (es gibt noch andere) besteht darin, von der Datenverfügbarkeitsorientierten Betrachtung von Patientensicherheit wegzukommen und dafür zu einem **Ziel- oder Problem-orientierten Vorgehen** zurückzukehren. Der Grund liegt darin: eine vernünftige Fragestellung ist der beste Weg zu aussagekräftigen Ergebnissen, vor allem wenn man sich noch zu einer adäquaten **Erkenntnisperspektive** (z.B. klinisch-epidemiologische Erhebung von *Status Quo*-Daten (2. Schritt), zu einem **Messinstrument** (z.B. Befragungen, Schritt 3) und zu einer differenzierten Nutzung von **Datenquellen** (Schritt 4) durchgerungen hat (s. Kap. 5.3.8.).



### **3.4. Der zentrale Begriff der Haftungslücke (*litigation gap*)**

Vor dem Hintergrund der Diskussion der Erhebungsmethodik im vorangegangenen Kapitel wird die Relevanz der Daten aus Datenbanken und Fallsammlungen mit juristischem oder versicherungsrechtlichem Hintergrund deutlich. Im Gegensatz zu epidemiologischen Daten, bei denen ein Populationsbezug besteht („Fälle auf 100 Patienten“) und man z.B. aus Untersuchungen an Krankenakten valide Häufigkeitsaussagen ableiten kann, handelt es sich bei den Daten aus dem juristischen Umfeld um Registerdaten mit nicht spezifiziertem Populationsbezug, da die Aufnahme eines Falls in die Datenbank vor allem von sekundären Umständen abhängt (z.B. Klagehäufigkeit, gerichtliche Anerkennung, Anerkennung einer Versicherungsleistung). Systematisch sind diese Daten zu den generierenden Verfahren zu rechnen, können (neben der haftungsrechtlichen Bedeutung) als Aufgreifkriterien dienen und wertvolle Hinweise für Fehlentwicklungen und Bereiche mit hohem Handlungsbedarf geben. Im Einzelfall bei *sentinel events* sind sie sogar Anlass für Einzelfallanalysen, wie dies im entsprechenden Kataster der *Joint Commission* gehandhabt wird.

Juristische und Daten aus dem haftungsrechtlichen Umfeld stammen in erster Linie von den Schiedsstellen der Landesärztekammern, von den Krankenkassen und von den Haftpflichtversicherern. Am meisten Informationen liegen zu den Daten aus den Schiedsstellen und von den Krankenkassen (Medizinischer Dienst, MDK) vor, die jährlich veröffentlicht werden.

Von den Schiedsstellen der Landesärztekammern werden Daten seit dem Erhebungsjahr 2006 im *Medical Error Reporting System* (Schaffartzik und Neu 2008) gesammelt und systematisch ausgewertet. Im Jahr gehen zwischen 10.000 und 12.000 Meldungen bei den Landesärztekammern ein, von denen rund 75% bearbeitet werden können und rund 25% als Behandlungsfehler gewertet werden. Im Jahr 2016 wurden von den Schlichtungsstellen in 1845 von 7639 bearbeiteten (von wiederum 11.500 gemeldeten) Fällen ein Behandlungsfehler erkannt (24,2%) in weiteren 400 Fällen lag ein Behandlungsfehler vor, der jedoch keine Kausalität für das unerwünschte Ereignis aufwies (Korzilius 2017). 166 Patienten mit einem Behandlungsfehler erlitten einen schweren Dauerschaden, 96 Patienten verstarben. Schwerpunkte liegen in der Unfallchirurgie und der (ambulanten) Diagnostik.

Diese Daten liegen in der gleichen Größenordnung wie die Daten des Medizinischen Dienstes des Spitzenverbandes der Krankenkassen (MDS 2017). Im Jahr 2016 hatte die Gesamtanzahl der Begutachtungen im direkten Vergleich zum Vorjahr leicht zugenommen (2016: n=15094, 2015: n=14.828), die Anzahl der festgestellten Fehler

(2016: n=4072, 2015: n=4.428), der Fehler mit Schaden (2016: n=3564, 2015: n=4.046) und der Fehler mit gutachterlich bestätigter Kausalität (2016: n=2948, 2015: n=3.156) hat leicht abgenommen. Besonders bemerkenswert ist die Übereinstimmung mit den Schiedsstellen-Angaben in der Häufigkeit der Behandlungsfehler (23,4% der gemeldeten Fälle). Insgesamt kann man in Deutschland von ca. 20.000 Fällen mit nur 500-600 Todesfällen pro Jahr ausgehen, die durch die Schiedsstellen oder juristisch weiter verfolgt werden.

Wie wichtig diese Klärung der Nomenklatur ist, zeigt die Darstellung der Deutschen Krankenhaus Gesellschaft, die auf der Basis der Haftpflichtversicherungsfälle der Ecclesia-Gruppe eine Häufigkeit von regulierten Behandlungsfehlern im Zeitraum von 1996 bis 2014 von 0,023% berechnete (Deutsch. Ärztebl. 21.4.2017). Hier bleibt nicht nur unberücksichtigt, dass die Meldung als Behandlungsfehlerverdacht nicht als epidemiologisches Kriterium gelten kann, sondern dass zusätzlich die Klärung der Anerkennung als Behandlungsfehler, der von den Haftpflichtversicherern reguliert werden, eine große Rolle spielt, so dass diese Zahlen wohl als Information für die Kalkulation von Versicherungsprämien etc. dienen kann, aber nicht als Grundlage für die populationsbezogene Häufigkeit von vermeidbaren unerwünschten Ereignissen.

Der Unterschied zwischen Angaben aus juristischer Perspektive auf der einen Seite und epidemiologisch gewonnenen Daten auf der anderen Seite ist wissenschaftlich gut untersucht. Bemerkenswerterweise ergeben sich international, also unabhängig von der Gestaltung der Rechtsbeziehungen, Ergebnisse in der gleichen Größenordnung. Zusammenfassend kann man festhalten, dass Daten, die vor einem juristischen oder versicherungsrechtlichen Hintergrund gewonnen werden, die Häufigkeit von Schäden ungefähr um den Faktor 1:30 unterschätzt. Die Ursache für diese Unterschätzung der Situation liegt darin, dass nur zwischen 1 % und 5 % aller Schäden einer rechtlichen Klärung zugeführt werden (sogenanntes *disclosure gap*). Wegen der Wichtigkeit dieser Thematik sollen die vorliegenden Studien hier kurz zusammenfassend dargestellt werden.

Die älteste Untersuchung stammt von Danzon (1985). In der sehr sorgfältigen Aufarbeitung der ersten *Malpractice Crisis* Anfang der 70er Jahre, die aus einem ökonomischen Blickwinkel vorgenommen wird, geht die Autorin von den Daten der *California Medical Insurance Feasibility Study* der *California Medical Association* und der *California Hospital Association* aus dem Jahr 1974 aus (s. Kap. 1.2.2., s. Mills 1978). In einer von 126 Krankenhausaufnahmen kommt es zu einem Behandlungsfehler im Sinne des *negligent adverse event* (s. Kap. 3.2.), bei dem also zusätzlich zur Vermeidbarkeit (Fehlerbedingtheit) noch eine Sorgfaltsverletzung festgestellt wurde. Eine juristische Aufarbeitung wurde nur in einem von 10 dieser Fälle eingeleitet. Wenn man berücksichtigt, dass die Kriterien für *negligent adverse events* sehr viel enger gefasst sind

als für Vermeidbare Unerwünschte Ereignisse (hier keine Sorgfaltsverletzung) oder gar Unerwünschte Ereignisse, weil zusätzlich zur Fehlerbedingtheit noch die Sorgfaltsverletzung vorliegen muss, kann man bereits in dieser frühen Studie eine Haftungslücke von deutlich unter 1:10 annehmen.

In einer der ersten ethnographischen Beobachtungsstudien zu UE in der stationären Versorgung (Andrews et al. 1997), in der 480/1047 (45,8%) Patienten mindestens ein UE und 185/1047 (17,7%) Patienten mindestens ein schweres UE erlitten, ließen nur 13/1047 (1,2%) Patienten ein UE juristisch abklären, darunter 11 Patienten von den 480 Patienten (2,3%), bei denen ein UE in der Studie festgestellt worden war. Nach 4 Jahren war in 3 Fällen eine Kompensation erfolgt, in 8 Fällen wurde die Untersuchung fallen gelassen.

Auch in der *Harvard Medical Practice Study* wurde die Frage intensiv untersucht (HMPS III, Localio et al. 1991). In der 31429 Patienten umfassenden Stichprobe aus New York fanden sich 51 Klagen (0,13%), davon wurden 8 durch die 280 Patienten geführt, die nach der Studie ein NAE aufwiesen (1,53%). Diese niedrige Rate steht im Gegensatz zur unspezifizierten Häufigkeit von Klagen (n=3570), die in der HMPS auf der Basis der NAE im Staat New York (n=27179) berechnet wurde (13%), aber zahlreiche Patienten betraf, bei denen nach epidemiologischen Kriterien kein NAE vorgelegen hatte. Diese Studie zeigt deutlich die Notwendigkeit auf, sauber zwischen epidemiologischer Methodik und juristischen Daten zu trennen.

In einer Nachuntersuchung der 1992 durchgeführten Utah-Colorado-Studie (Thomas et al. 2000) wurde in der Analyse von 14700 Krankenakten nach dem HMPS-Design (s. Kap. 3.3.4.4.) bei 587 UE festgestellt, von denen durch einen mehrstufigen Prozess 161 als *negligent adverse events* (NAE), also bedingt durch eine Sorgfaltsverletzung, eingeschätzt wurden (Studdert et al. 2000). Die Häufigkeit von NAE lag bei 0,9% aller Krankenhausaufnahmen. In nur 2,5% der Fälle mit einem NAE wurde eine juristische Klärung angestrebt (bei NAE mit schwerwiegenden Folgen in 3,8%), so dass in 97% der Fälle eine solche Klärung gar nicht erst versucht wurde. Negative Prädiktoren in einer gematchten Multivarianzanalyse waren Versichertenstatus (Medicare oder Medicaid), Alter über 75 Jahre oder niedriges Einkommen.

In einer Befragung von Ärzten (n=831) und Öffentlichkeit (n=1207) über die Erfahrungen mit UE bzw. „Fehlern“ in der Gesundheitsversorgung (Blendon et al. 2002) gab ein hoher Prozentsatz der Befragten an, entweder selbst oder im familiären Umfeld solche erlebt zu haben (n=290 bzw. n=507). Nur 2% (Ärzte) bzw. 6% (Öffentlichkeit) versuchten, die Sachlage juristisch zu klären.

Eine entsprechende Untersuchung in New Zealand (Davis et al. 2002B) ist deshalb bemerkenswert, weil sie in einem Gesundheitssystem mit einem *No-Fault-Ansatz*

(Schadensersatz und Schmerzensgeld unabhängig von der Feststellung eines ursächlichen Zusammenhangs, s. Kap. 5) durchgeführt wurde. Die Häufigkeit von Klagen kann nur indirekt abgebildet werden: die in Vorstudien festgestellte Häufigkeit von 2% NAE bei stationären Patienten (Davis et al. 2002A) wären im Untersuchungszeitraum in der untersuchten Auckland-Region in New Zealand 3000 Klagen zu erwarten gewesen, tatsächlich sind jedoch nur 150 Klagen eingereicht worden (5%).

Ein ähnlicher Ansatz wird in der Analyse von Krankenhaushaftpflicht-Versicherungsdaten verfolgt, der sich auf nosokomiale Infektionen bezieht und die bislang beste Abschätzung des *litigation gap* in Deutschland erlaubt (Mönch et al. 2011). Nach den vorliegenden epidemiologischen Daten zur Häufigkeit nosokomialer Infektion wären in den von der ECCLESIA-Gruppe betreuten Krankenhäusern (14% der Krankenhausbetten) in der 11 Jahre umfassenden Zeitspanne ca. 230.000 vermeidbare nosokomiale Infektionen zu erwarten gewesen, es wurden jedoch unter dem Stichwort „Hygienefehler“ nur 450 Fälle der Versicherung gemeldet (0,2%). Diese Studie bezieht sich zwar nur auf die Untergruppe der nosokomialen Infektionen, ist aber insofern von Bedeutung, weil sie als Grundgesamtheit weder UE noch Behandlungsfehler (NAE), sondern Vermeidbare UE verwendet.

In der Gesamtsicht (s. Tab. 3.4.-1) ist zunächst festzuhalten, dass eine juristische Klärung nur in einer Minderzahl der Fälle mit für den Patienten negativen Ereignissen angestrebt bzw. unternommen wird. Am eindrücklichsten erscheinen die Studien, die sich auf *Negligent Adverse Events* (NAE) beziehen, also auf eine Größe, die als epidemiologische Annäherung an den Begriff des Behandlungsfehlers verstanden werden kann, denn hier ist ja bereits die Sorgfaltsverletzung festgestellt, und man sollte am ehesten in dieser Untergruppe vermuten, dass der Weg zu einer juristischen Klärung nicht weit ist. Natürlich weisen die hier zusammengestellten Studien eine erhebliche Variabilität auf, was durch die sehr unterschiedlichen Settings und Untersuchungsmethoden bedingt sein mag, aber die meist genannte Größenordnung eines *Litigation Gap* von 97% erscheint weiterhin zutreffend, entsprechend einem Faktor von UE zu haftungsrechtlicher Klärung von 30:1.

Jahr*	Datenbasis	Bezugsgröße			Quelle
		UE**	VUE**	NAE**	
1974	<i>California Medical Insurance Feasibility Study (USA)</i>			10%	Danzon 1985
1983	Beobachtungsstudie an 1047 stationären Patienten	2,3%			Andrews et al. 1997
1984	<i>Harvard Medical Practice Study III</i>			1,53%	Localio et al. 1991
1992	Nachuntersuchung von 14700 Krankenakten der Utah-Colorado-Studie			2,5%	Studdert et al. 2000
1995	<i>Adverse Events in New Zealand Public Hospitals Study</i>			5%	Davis et al. 2002B
2002	Befragung	6%***			Blendon et al. 2002
1996-2006	Zu erwartende vermeidbare nosokomiale Infektionen, ECCLESIA-Versicherungsdaten in Deutschland		0,2%		Mönch et al. 2011

Tab. 3.4.-1: Relative Häufigkeit der juristischen Klärung von UE als Annäherung für die Größe der Haftungslücke (*litigation gap*) in der Übersicht. \*Durchführung der Studie, \*\* UE Unerwünschtes Ereignis, VUE Vermeidbares Unerwünschtes Ereignis, NAE *Negligent Adverse Event* (epidemiologische Erfassung von Behandlungsfehlern, zur Terminologie s. Kap. 3.2.), \*\*\* Angabe bezieht sich auf die befragte Öffentlichkeit, 2% unter den befragten Ärzten; gefragt wurde nach „errors“, was hier als UE interpretiert wird.

### 3.5. Zurechenbarkeit und Vermeidbarkeit

Neben dem Aspekt des *litigation gap* gibt es noch einen weiteren Punkt, der in diesem methodisch ausgerichteten Kapitel 3 zur Sprache kommen muss, und der in der allgemeinen Diskussion oftmals nicht genügend Beachtung findet: die Zurechenbarkeit von Ereignissen zur Behandlung und die Vermeidbarkeit (Fehler-Bedingtheit) dieses Ereignisses. Es handelt sich hier um zwei unterschiedliche Aspekte, und die Diskussion hierzu wird gewichtiger durch die Forderung, bestimmte UE als vollständig vermeidbar anzusehen (*Non-Payment-for-Non-Performance*, Belmont et al. 2011).

Wie oben bereits ausgeführt (s. Kap. 3.2.), formuliert der *To Err Is Human*-Report des IOM explizit: „*An adverse event attributable to error is a preventable adverse event*“ (IOM 1999, S. 24). Vermeidbare unerwünschte Ereignisse (VUE) sind daher auch methodisch schwieriger zu erfassen als negative bzw. unerwünschte Ereignisse ohne Bezug zu Fehlern (zu den Begriffen s. Abb. 3.2.-1). In Anlehnung und Erweiterung der Terminologie der Klinischen Epidemiologie muss man bei der Feststellung eines VUE folgende Schritte beachten:

1. es liegt ein **negatives Ereignis** vor, das sich darin zeigt, dass der Patient einen schlechteren gesundheitlichen Zustand hat als zu einem definierten früheren Zeitpunkt, und/oder dass der Nutzen einer Behandlung nicht verwirklicht werden kann, und/oder dass ein negativer Nutzen vorliegt (Beispiel: Fortschreiten einer Tumorerkrankung, aber auch: Komplikation eines Eingriffs, Anwendung einer ungesicherten Behandlungsmethode);
2. es liegt ein **zuschreibbares (attributable) negatives Ereignis** vor, das im Laufe und bedingt durch eine Behandlung auftritt und (soweit unbeabsichtigt) im Rahmen der Patientensicherheits-Terminologie als Unerwünschtes Ereignis anzusprechen ist, da die Kriterien der Negativität und des Behandlungsbezugs erfüllt sind (Beispiel: Auftreten einer Beatmungspneumonie);
3. es liegt ein **vermeidbares zuschreibbares negatives Ereignis** vor, das auf einen Fehler zurückgeht und als Vermeidbares Unerwünschtes Ereignis (VUE) oder *preventable adverse event* bezeichnet wird: „*an injury (or complication) that results from an error or systems failure*“ (*Massachusetts Coalition for the Prevention of Medical Errors*, Davis et al. 2001, S. 5).

Dieser Bezug zur Klinischen Epidemiologie durch den Begriff der Zurechenbarkeit von Ereignissen ist für die Patientensicherheitsforschung – und praxis von großer Bedeutung, denn hierdurch kann sie auf deren Erfahrungsschatz bei der Bewertung von durch die Behandlung bedingten Ereignissen zurückgreifen. Dies zeigt sich z.B. bei dem Begriff der *attributable mortality*, die nicht die Mortalität in einer Population (z.B. die Patienten einer

Intensivstation), sondern die auf eine Behandlungsmethode (z.B. künstliche Beatmung) zurückzuführende Sterblichkeit beschreibt. Erstere schließt auch die Sterblichkeit durch den Krankheitsverlauf (z.B. Multiorganversagen bei Polytrauma) mit ein, Letztere umfasst nur die Sterblichkeit z.B. durch Komplikationen. Die der Behandlung zurechenbare Mortalität stellt die Differenz zwischen Gesamtmortalität und krankheitsbedingter Mortalität dar. Bezieht man diese Betrachtung nicht nur auf die Sterblichkeit, sondern bezieht sie ganz allgemein auf Unerwünschte Ereignisse (UE), dann kann man diese beschreiben als:

$$UE = E - E_{Kr}$$

wobei E. alle negativen Ereignisse und  $E_{Kr}$  alle krankheitsbezogenen negativen Ereignisse (z.B. Krankheitsprogression) bezeichnet.

Bereits die Unterscheidung zwischen Krankheits- und Behandlungs-bezogenen negativen Ereignissen ist *in praxi* alles andere als einfach umzusetzen, stellt jedoch z.B. bei der Erfassung von nosokomialen Infektionen die zentrale Herausforderung dar, denn diese Infektionen sind allein durch ihren Behandlungsbezug definiert. In der ersten niederländischen, im HMPS-Design durchgeführten Studie zur Häufigkeit von UE im Krankenhaus wurde von den Reviewern in der zweiten Stufe eine abgestufte Einschätzung vorgenommen (s. Kasten).

Die Situation wird jedoch noch komplizierter, wenn man zusätzlich zum Behandlungsbezug das Auftreten eines Fehlers im Behandlungsprozess beschreiben will und somit ein vermeidbares, also ohne das Auftreten des Fehlers sich nicht verwirklichendes unerwünschtes Ereignis erfassen möchte. Im Vergleich zu klassischen klinisch-epidemiologischen Studien bedingt dies eine detaillierte Analyse des Ereignisses, des Krankheits- und Behandlungszusammenhangs sowie der Behandlungsplanung in Einheit mit der Behandlungsdurchführung. Die letzten beiden Entitäten, die Behandlungsplanung und die Behandlungsdurchführung, beziehen sich auf die Definition des Fehlers, der entweder in der Wahl des Behandlungsplans (Irrtum) oder in der Durchführung (Wahrnehmungs- oder Ausführungsfehler, Versehen oder Patzer) liegt (s. Begriff Fehler, Kap. 3.2.).

**Adverse events: Zurechenbarkeit zur Behandlungsmethode**

- 1 *(Virtually) no evidence for management causation*
- 2 *Slight to modest evidence of management causation*
- 3 *Management causation not likely (less than 50/50, but "close call")*
- 4 *Management causation more likely (more than 50/50, but "close call")*
- 5 *Moderate to strong evidence of management causation*
- 6 *(Virtually) certain evidence of management causation*

Causation scores of 4–6 were classified as AEs. Zegers et al. 2009

Es ist in der Literatur unbestritten, dass die Feststellung der Vermeidbarkeit nicht nur schwer umzusetzen, sondern auch regelmäßig von einer hohen Rate nicht-konkordanter Beurteilungen bei den Reviewern begleitet ist (Sari et al. 2007, Baines et al. 2013, Manaseki-Holland et al. 2017).

In der internationalen Literatur gibt es zu diesem Thema mehrere methodische Annäherungen:

„*Preventability of an AE was assessed as an error in health care management due to failure to follow accepted practice at an individual or system level*“ (Davis et al 2001),

und zwar auf der Ebene des betreuenden Arztes (Typ 1), durch ein anderes Mitglied des Behandlungsteams (Typ 2) oder durch das System (Typ 3) (Barnes et al. 2006). Etwas einfacher lautet die Definition von Michel et al. 2004:

„*Preventable adverse events were those that would not have occurred if the patient had received ordinary standards of care appropriate for the time of the study.*“

Ähnlich äußert sich die WHO (2009, S. 17):

“*Preventable is being accepted by the community as avoidable in the particular set of circumstances.*“

In den *New Zealand Public Hospitals* Studien wird eine Abstufung der Vermeidbarkeit (*preventability*) entsprechend der Sicherheit, mit der UE vermeidbar sind, vorgeschlagen (Davis et al. 2003), zur besseren Illustration werden in Tab. 3.4.-1 auch die absoluten und relativen Häufigkeiten zitiert (analog in den niederländischen Studien, s. Zegers et al. 2009).

Preventability Score		Preventability Level	n	%
1	Virtually no evidence	None	319	37,5
2	Slight to modest evidence	Low	143	16,8
3	Close call, <50:50	Low	73	8,6
4	Close call, > 50:50	High	135	15,9
5	Moderate/strong evidence	High	132	15,5
6	Virtually certain evidence	High	48	5,6
All AE			850	100

Tab. 3.4.-1: Vermeidbarkeit von UE, Umsetzung und Bewertung in der *New Zealand Public Hospitals* Studie II (Davis et al. 2003)



Die genannten Vorschläge zur Operationalisierung der Vermeidbarkeit in epidemiologischen Erhebungen beschränken sich jedoch mehr auf eine Abstufung des Zusammenhangs und nicht auf eine inhaltliche Argumentation. Die Austin Bradford Hill Kriterien (Poots et al. 2017) gehen allerdings deutlich darüber hinaus (s. Kasten). Die ersten drei Kriterien betreffen den statistischen Zusammenhang, außerdem die Frage der Reproduzierbarkeit und der eventuellen Alternativen. Weiterhin sollte der Zusammenhang zeitlich in Frage kommen, eine „Dosis-Abhängigkeit“ aufweisen, plausibel und mit der Theorie vereinbar sein, außerdem durch andere Zusammenhänge gestützt und evtl. durch Experimente abgesichert sein (Hill 1965).

**Kausalität: die Austin Bradford Hill-Kriterien** (Poots et al. 2017)

- 1 *Strength of association*
- 2 *Consistency of association*
- 3 *Specifity of association*
- 4 *Temporality*
- 5 *Biological Gradient*
- 6 *Plausibility*
- 7 *Coherence*
- 8 *Analogy*
- 9 *Experiment*

Man erkennt erst in diesem Zusammenhang, auf welches „epidemiologisches Wagnis“ sich die Initiatoren der frühen Studien in den USA (z.B. *Harvard Medical Practice Study*) eingelassen haben, als sie sich vorgenommen haben, neben der Zurechenbarkeit und der Vermeidbarkeit noch (drittens) die Bestimmung der Sorgfaltsverletzung zu schultern, um auch Behandlungsfehler (*negligent adverse events*) epidemiologisch zu erfassen.

## 3.6. Neuere Studien unter besonderer Beachtung der Situation in Deutschland

### 3.6.1. Allgemeine Vorbemerkungen

In diesem Weißbuch Patientensicherheit wird ein Verständnis von Patientensicherheit vertreten, das den linearen Ansatz überwindet und Patientensicherheit nicht mehr alleine durch Ergebnis-Parameter beschreibt (*end result*-Ansatz), sondern durch Indikatoren, die die Kompetenz der Organisationen und des Systems darstellen, Sicherheit zu verwirklichen. Entsprechende Ansätze werden in Kap. 5.3.8. diskutiert. Trotzdem ist der lineare Ansatz – dies wurde immer wieder angemerkt (z.B. Kap. 2.4.4.) - in manchen Situationen unersetzbar. Ein Beispiel hierfür ist die Diskussion um den *Status quo* und die Priorisierung durch die Zahl von (Vermeidbaren) Unerwünschten Ereignissen. Wenn wir in Deutschland jedes Jahr rund 20.000 vermeidbare Todesfälle allein in der Krankenhausversorgung vermuten müssen, ist dies als eindeutiges Indiz für den bestehenden Handlungsbedarf zu werten. Dass dieser Handlungsbedarf nicht von allen Seiten gleichermaßen gesehen wird, zeigt die wechselvolle Geschichte der Rezeption dieser Erkenntnisse (z.B. Geraedts 2014).

Im Vordergrund stehen klinisch-epidemiologische Daten, die nach klaren wissenschaftlichen Vorgaben international und in Deutschland erhoben wurden (zur Abgrenzung zu juristischen Daten s. Kap. 3.4., zu einzelnen methodischen Problemen s. Kap. 3.3.). Im Umgang mit diesen epidemiologischen Erkenntnissen sind im Verlauf der letzten 20 Jahre grob drei Phasen zu unterscheiden:

- In der **initialen Phase** in den Jahren vor und im Zeitraum um die Veröffentlichung des *To Err Is Human*-Reports durch das *Institute of Medicine* (IOM) in den USA (Kohn et al. 1999) wurde durch erste Studien (führend die Harvard Studie II und die Utah-Colorado-Studie (Leape et al. 1991, Thomas et al. 2000)) die Größenordnung des Problems deutlich, sowohl in der wissenschaftlichen *community*, den Akteuren im Gesundheitswesen und der Öffentlichkeit. Unter der wissenschaftlichen Koordination von Linda Kohn, Janet Corrigan und Molla Donaldson, die praktisch die Gesamtheit der wissenschaftlichen Experten auf diesem Gebiet hinter sich versammeln konnten, sparte das IOM nicht an deutlichen Worten:

*“When extrapolated to the over 33.6 million admissions to U.S. hospitals in 1997, the results of the study in Colorado and Utah imply that at least 44,000 Americans die each year as a result of medical errors. The results of the New York Study suggest the number may be as high as 98,000. Even when using the lower estimate, deaths due to medical errors exceed the number attributable to the 8th leading cause of death. More people die in a given year as a result of medical errors*

*than from motor vehicle accidents (43,458), breast cancer (42,297), or AIDS (16,516)” (Kohn et al. 1999, S. 1).*

- In der **zweiten Phase** führten die epidemiologischen Ergebnisse, untermauert durch zahlreiche weitere Studien, in den USA und in anderen Ländern – so auch in Deutschland – zu einer Diskussion über die möglichen Konsequenzen. Die Berufsgruppen und Institutionen der Gesundheitssysteme mussten sich aufgrund der öffentlichen und politischen Resonanz zur Thematik positionieren und kamen häufig zu einer konstruktiven Einstellung. So umfasste das Aktionsbündnis Patientensicherheit e.V. bei seiner Gründung praktisch alle Akteure des deutschen Gesundheitssystems, untermauert durch zahlreiche persönliche Gründungsmitgliedschaften auf Leitungsebene (Rothmund 2005A). Es wurden auf breiter Front nationale Aktionspläne entwickelt, die in den meisten Fällen die epidemiologische Basis der Vorarbeiten akzeptierten und an Präventionsstrategien und anderen Maßnahmen zur Verbesserung der Situation arbeiteten. Da auch in Deutschland zu diesem Zeitraum bereits zahlreiche Studien durchgeführt worden waren, konnte auf eine nochmalige Vergewisserung (wie in den meisten Ländern) verzichtet werden.

- in der **dritten Phase**, die bis heute andauert, gewinnt die wissenschaftliche Basis mit hohem Tempo an Tiefe und Breite, vor allem aber wird in die Differenziertheit der Kenntnisse investiert. Globale Studien zur Häufigkeit von UE und VUE stehen nicht mehr im Mittelpunkt, weil der Kenntnisstand zunehmend als abgesichert gilt, dafür werden die epidemiologische Daten zu einzelnen Untergruppen der UE (z.B. Medikation, Infektionen, einzelne Fächer) vertieft, und es werden solche Daten im Rahmen von Interventionsstudien gewonnen, was sie im Sinne der Konstruktvalidität besonders wertvoll erscheinen lässt (z.B. vorher-nachher-Vergleich). Allerdings gibt es auch fortbestehende Lücken, die gefüllt werden müssen (z.B. ambulante Versorgung, Medizinprodukte etc.).

Von der Forschungsstelle des Aktionsbündnis Patientensicherheit e.V. wurde in den Jahren 2005 bis 2009 (unterstützt vom Bundesministerium für Gesundheit) mehrere Systematische Reviews zur Epidemiologie von UE und VUE (Schrappe et al. 2006, 2008) sowie der damit verbundenen Mortalität (Schrappe et al. 2007) erarbeitet und auch international veröffentlicht (Lessing et al. 2010). Im erweiterten, zweiten Review wurden 241 Studien eingeschlossen (Publikation bis 6/2007), die Primärdaten zur Häufigkeit von unerwünschten Ereignissen, Schäden, Behandlungsschäden, Fehlern und Beinaheschäden enthielten und mit einer klar benannten Erhebungsmethode an einem definierten Patientenkollektiv durchgeführt worden waren (s. Abb. 3.6.1.-3). Es wurden keine Studien eingeschlossen, die nur ein einzelnes Behandlungsverfahren oder ein einzelnes Medikament beschrieben, wohl aber Studien, die sich auf eine komplette

Gruppe von Verfahren (z.B. alle unerwünschten Arzneimittelereignisse oder die Gesamtheit der nosokomialen Infektionen) bezogen. Der größte Teil dieser Studien stammte aus den Vereinigten Staaten von Amerika, Deutschland fand sich mit 12 Studien auf dem sechsten Platz. Die meisten der 241 Studien wurden in stationären Einrichtungen durchgeführt, ein weiterer Schwerpunkt lag im Arzneimittelbereich. Im Krankenhausbereich lag die Größenordnung bei

UE: zwischen 5% und 10%,

VUE: zwischen 2% und 4%,

Behandlungsfehler bei 1% und

Vermeidbare Mortalität bei 0,1%.

Diese Zahlen wurden von der Öffentlichkeit allgemein akzeptiert (FAZ vom 24.4.2007). Wichtig ist der Hinweis, dass es sich bei der Angabe zu den Behandlungsfehlern um eine epidemiologische Operationalisierung des juristischen Begriffs handelt, der sich am Begriff des *negligent adverse event* orientiert (s. Kap. 3.2.), so wie er in den frühen US-amerikanischen Studien eingesetzt wurde (*California Medical Insurance Feasibility Study* (Mills 1978), *Harvard Medical Practice Study II* (Leape et al. 1991), *Harvard Medical Practice Study III* (Localio et al. 1991), *Utah-Colorado Studie* (Thomas et al. 2000)).

Bereits damals war aufgrund der Ergebnisse über die Häufigkeit allein von nosokomialen Infektionen (3-5 % aller Patienten; Rueden et al. 1996), arzneimittelbedingten UE (0,17-6,5%; Lazarou et al. 1998), Stürzen (2,4%), Dekubiti (4,2%; beide Dassen 2004) und auf Medizinprodukte zurückgehenden Ereignisse (bis zu 8%; Samore et al. 2004) klar erkennbar, dass die oben genannten Größenordnungen eher als konservative Näherungswerte zu verstehen waren und die Problematik mit Sicherheit nicht überschätzten. Um ein Beispiel aus dem Bereich der *sentinel events* zu nennen: man vermutet in Deutschland maximal 100 Fälle von Seiten- und Eingriffsverwechslungen pro Jahr, im *Minnesota Report* in den USA werden aber 50 *wrong site surgeries* pro Woche (!) berichtet, was bei uns ca. 17 Fälle pro Woche oder 900 Fälle im Jahr entsprechen würde (Chassin 2013). Wir müssen davon ausgehen, dass wir uns nach wie vor nur mit der Spitze des Eisbergs beschäftigen – denn das eigentliche Problem tritt erst bei der näheren Betrachtung der Sachlage und außerdem der Betrachtung von besonders „verletzlichen“ Patientengruppen zutage, z.B. älteren und chronisch erkrankten Patienten.

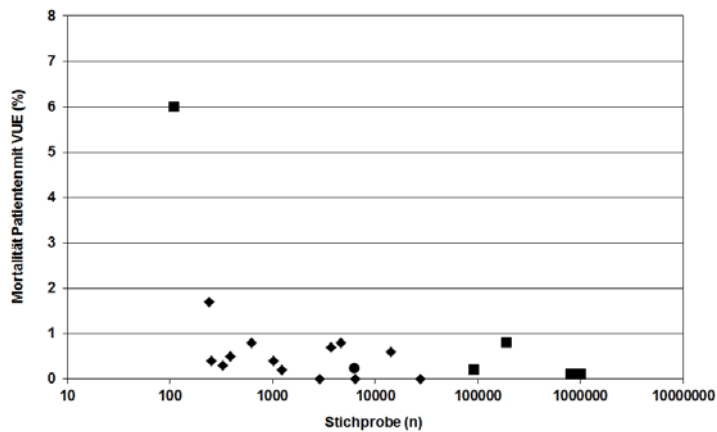
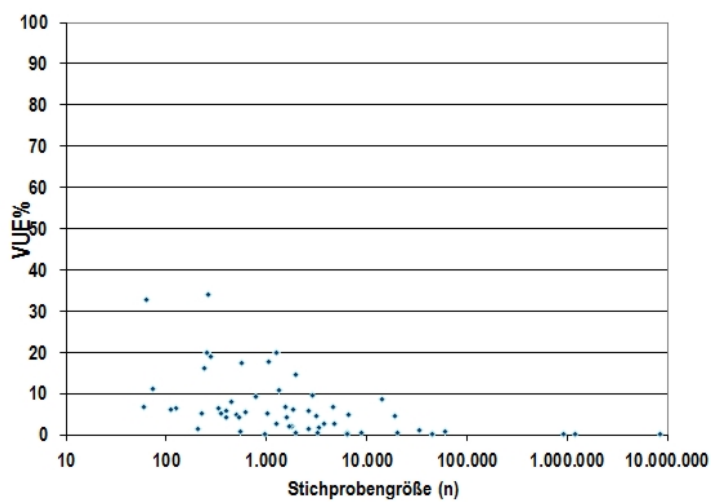
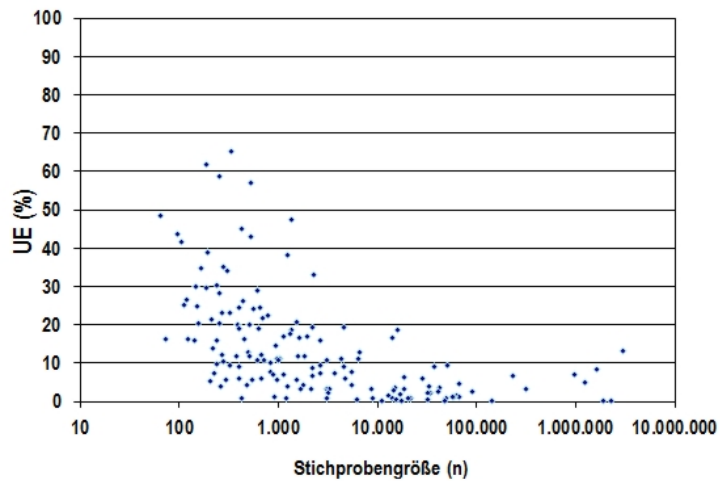


Abb. 3.6.-1 (oben): Häufigkeit von Unerwünschte Ereignissen (UE) in % der Patienten mit mindestens einem Ereignis, in Abhängigkeit von der Stichprobengröße, 153 Studien, modif. aus Schrappe et al. 2008

Abb. 3.6.-2 (Mitte): Häufigkeit von Vermeidbaren Unerwünschte Ereignissen (VUE) in % der Patienten mit mindestens einem Ereignis, in Abhängigkeit von der Stichprobengröße, 58 Studien, modif. aus Schrappe et al. 2008

Abb. 3.6.-3 (unten): Mortalität für Patienten an VUE über alle Studiengruppen (n=19 Studien).

(◆)=Kohortenstudien, bei denen ein kausaler Zusammenhang zwischen VUE und Sterblichkeit nachgewiesen wurde (n=13), (■) und (●) = Fallkontrollstudien an VUE (n=6), modif. nach Schrappe et al. 2007

Jeweils halblogarithmische Darstellung über die Stichprobengröße.

Die epidemiologischen Daten wurden genau auf Einflussfaktoren untersucht. Es ergab sich kein Einfluss auf die Häufigkeit durch Faktoren wie Land der Untersuchung, Art des Unerwünschten Ereignisses (z.B. durch alle UE, Arzneimittel oder nosokomiale Infektionen; letztere waren die einzigen beiden Untergruppen, die in die Analyse aufgenommen worden waren), stationäre vs. ambulante Versorgung. Untersuchungen mit der Datenquelle Routinedaten würde in größeren Studien eingesetzt und erbrachten niedrigere Befunde. Aus Deutschland wurden 12 Studien eingeschlossen (s. Tab. 3.6.-1), die die gleiche Häufigkeitsverteilung wie die Gesamtheit der 241 Studien zeigte (Schrappe et al. 2018).

Rang	Land	Anzahl
1	USA	88
2	Kanada	20
	Frankreich	20
3	Australien	16
4	Spanien	15
5	Großbritannien	13
6	Deutschland	12
7	Schweiz	7

Tab. 3.6.-1: Systematischer Review an 241 Studien zur Häufigkeit von UE, Verteilung über die Länder, in denen die Untersuchungen durchgeführt wurden (Schrappe et al. 2008)

Genauer wurde in einer späteren Publikation (Lessing et al. 2010) die Streuung untersucht, die im *funnel plot* (Trichtergraphik) deutlich von der Zahl der eingeschlossenen Patienten abhing und sich in Studien ab 1000 Patienten in der genannten Größenordnung stabilisierte (s. Abb. 3.6.-1 und -2). Dieser Effekt ist bei epidemiologischen Studien zu erwarten und kann als indirektes Zeichen für die Validität der Studien und des Reviews gewertet werden.

Aus einer Untergruppe von 51 Studien des ersten Reviews wurden Daten zur **Mortalität** extrahiert (Schrappe et al. 2007). Es handelte sich um 39 Kohortenstudien und 12 Fallkontrollstudien. In 45/51 Studien wurde die Mortalität bezogen auf die Grundeinheit der Gesamtpopulation untersucht, so dass Raten bzw. Proportionen zu bilden waren. In 43 Studien wurden Krankenhauspatienten beobachtet, 31 davon gehörten zu der wichtigen Gruppe der 36 Kohortenstudien über UE. Der mediane Stichprobenumfang betrug 4031 Patienten, 18 Studien wurden in den USA, 2 in Deutschland durchgeführt. Unerwünschte Ereignisse (UE) wurden in 43 und vermeidbare unerwünschte Ereignisse (VUE) in 19 Studien erhoben (11 beide). Die Ergebnisse der 19 Studien zu VUE als Todesursache zeigen, dass man mit einer Risikoangabe von 0,1% aller Krankenhauspatienten, an einem VUE zu versterben, die Situation sicher nicht übermäßig dramatisiert, sondern eher konservativ einschätzt (s. Abb. 3.6.-3).

Wenn man diese Ergebnisse zusammenfasst und heutzutage von jährlich knapp 20 Mill. Krankenhauspatienten ausgeht, ergeben sich allein im Krankenhausbereich

- 1.000.000 bis 2 Mill. unerwünschten Ereignissen (UE),
- 400.000 bis 800.000 vermeidbare unerwünschte Ereignisse (VUE),
- 200.000 Behandlungsfehler (*negligent adverse events*) und
- 20.000 auf VUE zurückgehende (also vermeidbare) Todesfälle.

Besonders die vermeidbaren Ereignisse sind von Bedeutung, da sie auf die Feststellung eines Fehlers zurückgehen (zur Methodik s. Kap. 3.5.). Wenn man Studien heranzieht, die genauere Erhebungsmethoden (z.B. die begleitende Beobachtung durch geschultes Personal) verwenden, kommt man auf deutlich höhere Werte (vgl. Andrews et al. 1997, vgl. Kap. 3.3.4.).

Die Aktualisierung des epidemiologischen Kenntnisstandes in den nächsten Kapiteln erfolgt nicht als Systematischer Review. Stattdessen werden gezielt Studien zusammengestellt, die

- nach dem HMPS-Design durchgeführt wurden (Kap. 3.6.2.1.),
  - auf bestimmte Subkollektive und spezifische Datenquellen zurückgehen (Kap. 3.6.2.2.),
  - als Systematische Reviews Studien zusammenfassen (Kap. 3.6.2.3.),
  - eine besonders hohe Aussagekraft als Interventionsstudien haben (Kap. 3.6.2.4.)
- und die
- entweder in Deutschland durchgeführt wurden oder gut auf Deutschland übertragbar erscheinen (Kap. 3.6.3.).

Abschließend wird der Handlungsbedarf skizziert und später in Kap. 5.3.8. wieder aufgenommen.

### 3.6.2. Internationale Studien seit 2007

#### 3.6.2.1. Trigger-gestützte Studien

Bei den Trigger-gestützten Studien handelt es sich um zwei- oder mehrstufige Verfahren (s. Kap. 3.3.4.4.). In mehreren Ländern wurden auch in den letzten Jahren noch derartige Studien durchgeführt, um die nationale Situation genauer beschreiben zu können. Zunächst wird hier auf die Studien eingegangen, die analog zum *Harvard Medical Practice Study Design* (HMPS-Design) durchgeführt wurden (s. Kap. 3.3.4.4., s. Tab. 3.6.-2).

In **Schweden** wurden in einem dreistufigen Verfahren UE in 12,3% der 1967 zufällig ausgewählten Patienten gefunden, 70% vermeidbar, 5 der 169 VUE führten zum Tode (ca. 0,25% der eingeschlossenen Patienten). Die VUE hatten eine Verlängerung der Verweildauer von 6 Tagen zur Folge (Soop et al. 2009).

In **Großbritannien** wurde eine zweite, monozentrische Studie (zur ersten Studie s. Vincent et al. 2001) nach einem HMPS-ähnlichen Verfahren durchgeführt, die bei 1006 zufällig ausgewählten Patienten eine Rate von UE von 8,7% (31% vermeidbar, Verlängerung der Verweildauer 8. Tage, 10% der UE waren mitverantwortlich für den Tod der Patienten (Vermeidbarkeit nicht berichtet)) ergab (Sari et al. 2007)..

Eine sorgfältig geplante Studie in der **Republik Irland** an 1574 Patienten aus 8 Krankenhäusern (stratifiziert nach Größe und Region) erbrachte eine UE-Rate von 12,2% (70% vermeidbar, 6,7% mitverantwortlich für den Tod der Patienten (Vermeidbarkeit nicht berichtet)). Die Liegezeit verlängerte sich bei Patienten mit UE um 6,1 Tage, entsprechend zusätzlicher Kosten von €5550 pro Ereignis (Rafter et al. 2017).

In den **Niederlanden** wurden drei Studien nach dem HMPS-Design durchgeführt. In der **ersten** Studie (Zegers et al. 2009) wurde an 3943 zufällig ausgewählten, entlassenen Patienten eine Rate von UE in der Höhe von 5,7% gefunden (VUE 2,3%, entspr. 39,6% der UE; 7,4% trugen zum Tod bei (Vermeidbarkeit nicht berichtet)). In einem gleichzeitig untersuchten Kollektiv von 3983 verstorbenen Patienten war nur in 10,7% ein UE aufgetreten, allerdings war in 4,1% der Verstorbenen ein VUE mitverantwortlich für den Tod der Patienten (Zegers et al. 2009). Die Autoren rechnen auf der Grundlage dieser Daten mit zwischen 1482 und 2032 zusätzlichen (vermeidbaren) Todesfällen in den Niederlanden pro Jahr. In der **zweiten** Studie lag bei 3996 Patienten die Rate von UE bei 6,2% und der VUE bei 1,6% (Baines et al. 2003), in der **dritten** Studie an 4048 Patienten bei 5,7% bzw. 1,6% (Baines et al. 2015). Die Verlaufsbeurteilung ist nicht ganz einfach verständlich, da die Kollektive jeweils einer neuen Analyse unterzogen wurde (hier jeweils



die aktuellen Werte in der Originalpublikation; zur Diskussion s. Shojania und Thomas 2013B, Shojania und Marang-Van de Mheen 2015).

Autoren	Land	Pat.	Mind. 1 UE	VUE/UE	Mort.	Div.
Soop et al. 2009	Swe	1967	12,3%	70%	ca. 0,25% (vermeidbar)	Verl. VWD 6 Tg.
Sari et al. 2007	GB	1006	8,7%	31%	10% der UE*	Verl. VWD 8 Tg.
Rafter et al. 2017	Irl.	1574	10,3%	70%	6,7% der UE*	Verl. VWD 6,1 Tg.
Zegers et al. 2009	NL	3943	5,7%	39,6%	7,4% der UE*	-
Banies et al. 2013	NL	3996	6,2%	ca. 25%	-	-
Banies et al. 2015	NL	4048	5,7%	ca. 20%	-	-

Tab. 3.6.-2: Studien analog zum HMPS-Design, UE Unerwünschtes Ereignis, VUE Vermeidbares UE, Krhs. Krankenhaus, k.A. keine Angabe, Swe Schweden, Irl. Rep. Irland, Verl. VWD Verlängerung der Verweildauer. \*Vermeidbarkeit der UE, die zu den Todesfällen beitrugen, nicht berichtet.

Breit eingesetzt wurde auch der **Global Trigger Tool** (s. Kap. 3.3.4.4., Ergebnisse s. Tab. 3.6.-3). In einer frühen Studie mit 278 Patienten wurde bei 41 Patienten mindestens ein UE (15%) und bei 3 Patienten ein UE festgestellt, das als *contributing to death* (Grad I, 1%) eingeschätzt wurde (Department of Health and Human Services 2008). In der Nachfolgestudie wurde bei 838 Medicare-versicherten Patienten (erste Stufe: Abrechnungsdaten ohne *Present on Admission* und/oder Review durch Pflege, zweite Stufe für 420 Patienten: Review durch Ärzte) ein UE bei 13,5% der Patienten (44% vermeidbar, 7,4% VUE, 29/1000 Patiententage, 12/128 UE trugen zu Todesfällen bei) gefunden (Department of Health and Human Services 2010). In der sehr prominenten Studie von Classen et al. (2011) wurden in den Unterlagen von 795 zufällig ausgewählten Patienten aus drei Krankenhäusern in den USA mindestens ein UE-Häufigkeit bei 33,2% der Patienten (91 UE auf 1000 Patiententage) festgestellt. Da mehrere Patienten mehrere UE erlitten, lag die Häufigkeit aller UE bei 45 auf 100 Patienten. Die Vermeidbarkeit wurde nicht untersucht. In der Studie von Landrigan et al. (2010), die sich mit dem zeitlichen Verlauf von 2002 bis 2007 und dem Vergleich von internen vs. externen Reviewern

beschäftigte, wurde an 2341 Patienten aus 10 Krankenhäusern in North Carolina mindestens ein UE in 18,1% der Patienten (25,1 UE auf 100 Patienten, 56,5 UE auf 1000 Patiententage) festgestellt, im zeitlichen Verlauf sank die Zahl der VUE von 10,2 auf 6,5 pro 100 Patienten. 14 der 588 UE trugen zum Tod der Patienten bei. In der Studie von Stockwell et al. (2015) in der Pädiatrie wurde in 600 zufällig ausgewählten Patienten in 6 Kliniken in den USA eine Rate von mindestens einem UE in 24,3% der Patienten (40 UE auf 100 Patienten, 54,9 UE auf 1000 Patiententage) festgestellt, 108 UE (45%) waren vermeidbar.

Autoren	Land	Pat./Krhs.	Mind. 1 UE	UE/100 Pat.	UE/1000 Pat.-Tg.	VUE	Mort.	Div.
DHHS 2008	USA	278 / 24	15%	k.A.	k.A.	k.A.	1%*	-
DHHS 2010	USA	838 (420) / k.A.	13,5%	36	29	44% d. UE, 7,4%	12/128 UE (1,5%*)	-
Classen et al. 2011	USA	795 / 3	33,2%	45	91	k.A.*	8/393 UE (2%)*	Geschätzte**** vermeidbare Mortalität 1%
Landrigan et al. 2010	USA	2341 / 10	18,1%	25,1	56,5	63,1% der UE	9/364 VUE	Geschätzte**** vermeidbare Mortalität 0,4%
Stockwell et al. 2015	USA	600 / 6	24,3%	40	54,9	45% der UE	1 Fall	Todesfall nicht vermeidbar

Tab. 3.6.-3: Studien mit dem *Global Trigger Tool*, DHHS *Department of Health and Human Services* (USA), UE Unerwünschtes Ereignis, VUE Vermeidbares UE, Krhs. Krankenhaus, k.A. keine Angabe, \*Vermeidbarkeit nicht berichtet, \*\*bzw. 100% der UE als VUE bezeichnet; \*\*\*in James 2013 und Makary und Daniel 2016 9 Todesfälle berichtet, die aber aus der Publikation von Classen et al. 2011 nicht zu entnehmen sind; \*\*\*\* durch MS (8 UE auf 795 Pat. (UE = VUE!), 9 Todesfälle durch VUE auf 2341 Pat.)

### 3.6.2.2. Systematische Reviews

Studien zur Häufigkeit von unerwünschten Ereignissen liegen seit mehr als 20 Jahren vor. **Systematische Reviews** zu diesen Studien sind in den letzten Jahren seltener durchgeführt worden, weil der Kenntnisstand allgemein als gegeben angesehen wird (s. Tab. 3.6.-4). Ein Systematischer Review aus dem Jahr 2008 zeigte bei knapp 75000 Krankenhaus-Patienten eine Häufigkeit von UE von 9.2%, 43,5% davon waren vermeidbar (VUE), die Sterblichkeit lag bei 7,4% (de Vries et al. 2008). Ein Systematischer Review aus der Arbeitsgruppe um den bekannten Patientensicherheitsforscher Charles Vincent aus dem Jahr 2013 zu 14 Studien an 16424 chirurgischen Patienten aus England kommt zu höheren Ergebnissen als das APS: 14,4% erleiden ein UE, 5,2% ein VUE und 0,52% versterben (Anderson et al. 2013). Wie unten genauer ausgeführt, ist es nicht ohne weiteres möglich, von der Vermeidbarkeitsrate der UE auf die Vermeidbarkeitsrate der Todesfälle zu schließen.

Autoren	Studien	Mind. 1 UE	VUE	Mort.	Div.
De Vries et al. 2008	8	9,2%	43,5% d. UE	7,4% der UE*	-
Anderson et al. 2013	14	14,4%	37,9% d. UE	3,6% der UE*	-
De Boeker et al. 2013	6	8,92%	18-54% d. UE	k.A.	Nur chir. Pat., nur UAE
James 2013	4	5,7%	39,6% d. UE	7,4% der UE*	Eher narrativ

Tab. 3.6.-4: Systematische Reviews zu Studien analog zum HMPS-Design, UE Unerwünschtes Ereignis, VUE Vermeidbares UE, UAE Unerwünschtes Arzneimittelereignis, Krhs. Krankenhaus, k.A. keine Angabe, \*keine Angabe zur Vermeidbarkeit.

Der Review von James (2013) wird hier zwar aufgeführt und ist auch als sehr einflussreich zu bewerten (s. Kap. 3.6.2.3.), es handelt sich dabei aber nicht im eigentlichen Sinne um einen Systematischen Review (Einschlusskriterien nicht genannt etc.), sondern ähnlich wie bei Makari und Daniel (2016) um eine gezielte Zusammenstellung von Studien. Es werden aber wichtige Aspekte beigetragen, z.B. das Fehlen von *errors of omission* in den meisten Studien (vgl. Kavanagh et al. 2017).

### 3.6.2.3. Mortalität

Zur Mortalität gab es in den USA in den letzten Jahren eine intensiv geführte Debatte, in der zwei ganz gegensätzliche Positionen aufeinandertrafen: die eine Position sieht die gegenwärtig Zahlen zur vermeidbaren (!) Mortalität durch Unerwünschte Ereignisse, die ja in der Größenordnung auf den Zahlen aus *To Err Is Human* beruhen, als deutlich zu niedrig an, die andere Position ist der Meinung, dass die Zahlen viel zu hoch seien, weil die Vermeidbarkeit von Todesfällen besonders schwer zu erheben ist.

Zur Untermauerung der ersten Position spielen sowohl die Publikation von James (2013) als auch diejenige von Makary und Daniel (2016) im BMJ eine große Rolle. Beide Autoren bezweifelten die Häufigkeitsangaben der frühen HMPS-Studien und des IOM-Reports von 1999, es seien zwischen 44.000 und 98.000 vermeidbare Todesfälle pro Jahr in den USA zu erwarten. Nach ihren Schätzungen seien die Zahlen um den Faktor 4 nach oben zu korrigieren (Größenordnung 200.000 bis 400.000). Abgesehen von der Tatsache, dass die Gesamtzahl aller in den US-amerikanischen Krankenhäusern pro Jahr verstorbenen Patienten zwischen 700.000 und 800.000 liegt (Hall et al. 2013) und es daher zweifelhaft ist, ob bis zu einer Hälfte aller Todesfälle im Krankenhaus als vermeidbare UE einzustufen sein können, muss man anerkennen, dass hier berechnete Fragen gestellt werden (z.B. dass in keiner Studie die unterlassenen Maßnahmen mitberücksichtigt werden (James 2013)). Die *Leapfrog*-Organisation (s. [www.leapfrog.org](http://www.leapfrog.org), Moran und Scanlon 2013) hat in Zusammenarbeit mit dem *John Hopkins Armstrong Institute for Patient Safety and Quality* eine Studie zum Thema veröffentlicht, in der man sich der Meinung der beiden oben genannten Autoren(gruppen) anschließt (Austin und Derk 2016). Die Studie bezieht sich auf die *attributable mortality* (zum Begriff s. Kap. 3.5.) von ausgewählten *Outcomes* aus dem *Leapfrog*-PSI-Set (s. Kap. 3.3.3.3.), die aus der Literatur entnommen und um 50% reduziert werden, um die Vermeidbarkeit zu berücksichtigen (s.u.). Dieses Verfahren führt zu einer Zahl von 206021 vermeidbaren Todesfällen pro Jahr in den USA. Es besteht eine deutliche Differenzierung dieser derart berechneten vermeidbaren Mortalität zwischen den nach dem *Leapfrog*-Ranking besten und schlechteren Krankenhäusern, so dass pro Jahr über 33.000 Todesfälle zu vermeiden wären, wenn alle Krankenhäuser auf dem nach *Leapfrog*-Kriterien besten Niveau arbeiten würden.

Von Seiten der Patientenverbände hat man sich daher auch sehr konstruktiv mit diesen Veröffentlichungen auseinandergesetzt (Kavanagh et al. 2017) und insbesondere betont, dass bei der Beurteilung der Vermeidbarkeit durch (interne oder externe) Reviewer nicht berücksichtigt werden dürfe, dass der Tod nach einer begrenzten Zeitspanne (z.B. einer Woche) auch ohne den besagten Fehler sowieso eingetreten wäre.

Dieses Argument ist ernst zu nehmen. Trotzdem muss man sich mit teils erheblichen Bedenken gegen diese Erhebungen auseinandersetzen (z.B. Shojania und Dixon-Woods 2017), die argumentieren, dass die Vermeidbarkeitsrate bei Todesfällen unter der Vermeidbarkeitsrate der Gesamtheit der Unerwünschter Ereignisse zu liegen scheint (Hayward und Hofer 2001). In einer englischen Studie wurden 1000 stationäre Todesfälle retrospektiv durch Reviewer untersucht und nur bei 5,2% dieser Todesfälle Vermeidbarkeit festgestellt (Hogan et al. 2012). In einer Nachfolgestudie wurden sogar nur 3,6% der Todesfälle im Krankenhaus als vermeidbar angesehen (Hogan et al. 2015). Auch in der ersten der drei niederländischen Studien nach dem HMPS-Design hat eine Untersuchung an knapp 4000 im Krankenhaus Verstorbenen erbracht, dass ein VUE nur für 4,1% der Verstorbenen verantwortlich zu machen war (Zegers et al. 2009).

In der praktischen Konsequenz kann also festgehalten werden, dass man mit der Berechnung der vermeidbaren Mortalität aus der Vermeidbarkeitsrate aller UE vorsichtig sein sollte. Wenn 50% der UE vermeidbar sind, heißt dies noch lange nicht, dass auch 50% der Todesfälle als vermeidbar angesehen werden dürfen. Stattdessen ist es auf der Basis der bislang vorliegenden Untersuchungen nicht auszuschließen, dass die Vermeidbarkeitsrate der Todesfälle geringer als die Vermeidbarkeitsrate aller UE ist. Hier liegt auch für Deutschland ein wichtiges Forschungsfeld (die gleiche Problematik wurde auch im Review 2006/8 thematisiert (Schrappe et al. 2007)).

Wenn man sich die vorstehenden Tabellen unter diesem Gesichtspunkt anschaut, dann liegt über alle methodischen Zugänge hinweg die **Sterblichkeit der Unerwünschten Ereignisse**

- bei den HMPS-analogen Studien zwischen 6,7% und 10% (s. Tab. 3.6.-2),
- bei den GTT-Studien niedriger bei 1,5% (DHHS 2010) bzw. 2% (Classen et al. 2011) (s. Tab. 3.6.-3) und
- bei den Systematischen Reviews bei 3,6% und 7,4% (s. Tab. 3.6.-4; die Zusammenstellung der Studien von James (2003) nicht berücksichtigt),

Zwei Studien geben einen Anhaltspunkt für die **vermeidbare Mortalität**: Nur eine einzige Studie (Soop et al. 2009) berichtet direkt die vermeidbare Mortalität; sie liegt bei 0,25%. In der Studie von Landrigan et al. (2010) ist es statthaft, die vermeidbare Mortalität aus der Rate der Todesfälle wegen vermeidbarer VUE (9/364) zu berechnen (0,4% der Gesamtpopulation von 2341 Patienten). In der Studie von Classen et al. (2011) besteht eine Besonderheit, weil hier alle UE als vermeidbar eingeschätzt werden, so dass man nach den Kriterien dieser Studie (unter Vorbehalt) 8 UE auf die Studienpopulation von 795 Patienten beziehen könnte (vermeidbare Mortalität von 1%).

Zusammenfassend kann aber auf jeden Fall festgehalten werden, dass die in Deutschland vom Aktionsbündnis Patientensicherheit vertretene Ansicht, die vermeidbare

Mortalität liege bei 0,1%, eine konservative, aber belastbare Schätzung darstellt und wenn, dann vorsichtig nach oben korrigiert werden sollte. Die Rate von 0,1% entspricht in Deutschland bei rund 20 Mill. Krankenhauspatienten einer vermeidbaren Mortalität von 20.000 Patienten pro Jahr. Bei knapp 420.000 Sterbefällen im Krankenhaus (Jahr 2016, Statistisches Bundesamt 2017) ist daher jeder 20. Sterbefall im Krankenhaus in Deutschland als vermeidbar (auf einen Fehler zurückzuführen) einzustufen.

Man darf aber bei diesen feingliedrigen Überlegungen nicht aus dem Auge verlieren, dass die Mortalität im Krankenhaus und in der unmittelbar anschließenden Versorgung ein häufiges Ereignis darstellt. In den USA wurden kürzlich die Ergebnisse einer Studie an über 21 Mill. Krankenhauspatienten aus 4483 Krankenhäusern zur Frage veröffentlicht, wie sich die Krankenhausgröße und der *teaching*-Status auf die Sterblichkeit auswirkt: die unadjustierte 30-Tage Sterblichkeit für große Universitätskliniken lag bei 8,1%, für kleinere *teaching*-Krankenhäuser bei 9,2% und bei kleineren Krankenhäusern bei 9,6% (adjustiert 8,3%/9,2%/9,5%) (Burke et al. 2017). Eine Studie zum *Hospital Readmission Reduction Program* (HRRP) erbrachte eine risikoadjustierte 30Tage-Sterblichkeit von 8,4% für Patienten mit Herzinsuffizienz, 7,6% für den akuten Myokardinfarkt und 8,5% für die ambulant erworbene Pneumonie (Dharmarajan et al. 2017). In der *European Surgical Outcome Study* wurde für Europa an 46439 Patienten aus 498 Krankenhäusern in 28 europäischen Ländern eine Mortalität von 4% berichtet, wovon 73% vor ihrem Tod nicht auf Intensivstation behandelt worden waren (Pearse et al. 2012). Dies sind nur drei herausgegriffene Beispiele, die aber zeigen, dass die Mortalität im Zusammenhang mit dem Thema Patientensicherheit auch in Zukunft relevant bleiben wird. Abschließend muss allerdings nochmals mit aller Deutlichkeit darauf hingewiesen werden, dass die Erhebung der Mortalität ein lineares Sicherheitsverständnis widerspiegelt. Mortalität ist ein wichtiges klinisch-epidemiologisches Datum, aber kein sinnvoll einzusetzender Patientensicherheitsindikator (s. Kap. 3.3.3.3., s. z.B. Hogan et al. 2015).

#### 3.6.2.4. Interventionsstudien

Bedingt durch den international sich immer mehr ausbreitenden Einsatz von Sicherheitsrelevanten Daten zur Steuerung auf Systemebene (z.B. *Public Reporting, Pay for Performance*) liegen zunehmend Studien vor, die klinisch-epidemiologische Parameter über einen längeren Zeitraum und im zeitlichen Zusammenhang mit einer solchen Systemintervention untersuchen. Untersuchungen insbesondere im Zusammenhang mit Interventionen haben eine besonders hohe (Konstrukt-)Validität und können heute zur Abschätzung der Sicherheitsproblematik genutzt werden, während sie vor 10 Jahren noch nicht in diesem Umfang vorlagen. Allerdings sind diese Studien meist nicht geeignet, die

Vermeidbarkeit von UE zu beschreiben, da sie nicht primär auf die Epidemiologie von UE ausgerichtet sind.

Eine der Studien bezieht sich auf den Einsatz des **National Surgical Quality Improvement Program (NSQIP)** des *American College of Surgeons* in den USA, das einem *Private Disclosure* Programm entspricht (die Daten werden nicht der Öffentlichkeit, sondern nur intern zur Verfügung gestellt). Bei elektiven Operationen (von der Hernienchirurgie über Schilddrüsen-Operationen bis zur Kolon- und Ösophaguschirurgie) wurden bei knapp 350.000 Patienten Komplikationen von 2009 bis 2013 für Häuser innerhalb und außerhalb des NSQIP-Programms dargestellt (keine Überlegenheit der NSQIP-Häuser). Komplikationen traten in 4,9% und schwere Komplikationen in 2,0% der Patienten auf, 0,8% der Patienten verstarben (Etzioni et al. 2015), über die Zeit kam es in allen Gruppen zu einem kontinuierlichen Abfall.

Eine weitere Studie zum NSQIP verglich in einem *case control*-Ansatz bei insgesamt 800 Krankenhäusern mit mehr als 1,2 Mill. chirurgischen Patienten die Risiko-adjustierte 30Tage-Mortalität (4,3 vs. 4,5%), schwerwiegende Komplikationen (11,1 vs. 11,0%), Reoperationen (0,49 vs. 0,45%) und Wiederaufnahmen (13,3 vs. 12,8%) vor und nach der Aufnahme in das NSQIP-Programm (Osborne et al. 2015).

In einer Studie zur Wirkung des *Medicare*-Programms **Hospital Compare** (*Public Reporting*) wurde für drei Diagnosen in einem Kollektiv von 12 Mill. Krankenhaus-Patienten die Sterblichkeit vor und nach Programmeintritt verglichen: Myokardinfarkt 17,8% vs. 14,8%, Herzinsuffizienz 10,1% vs. 8,8% und ambulant erworbene Pneumonie 11,7% vs. 9,1% (Ryan et al. 2012).

In der **randomisierten Studie zum Einsatz eines eHealth-Instrumentes** zur Gestaltung des Entlassungsprozesses wurde in einer monozentrischen Studie in den USA der kombinierte Endpunkt *Mortalität/Readmission* bei 32,8% der Patienten in der Interventions- und 29,4% in der konventionellen Gruppe beobachtet (Santana et al. 2017).

In einer großen Studie zum *Hospital Readmission Reduction Program* (HRRP) wurde bei über 20 Mill. Patienten mit knapp 50 Mill. stationären Aufnahmen wegen Herzinfarkt, Herzversagen und ambulant erworbener Pneumonie Wiederaufnahmeraten zwischen 15,7 und 27,5% beobachtet (Desai et al. 2015). Im Zusammenhang mit dem gleichen HRRP-Programm wurde bei knapp 500.000 chirurgischen Patienten eine ungeplante Wiederaufnahmerate von 5,7% festgestellt (20% davon durch postoperative Wundinfektionen) (Merkow et al. 2015), ein Befund, der Lucian Leape immerhin zu dem Editorial „*Turning Complications into ‚Treasures‘*“ veranlasste (2015). Eine etwas ältere Studie mit knapp 60.000 chirurgischen Patienten zeigte eine 30Tage-Komplikationsrate von 22,6% (75% nach Entlassung) und eine 30Tage-Wiederaufnahmerate von knapp

12% (Morris et al. 2014). Eine oben schon erwähnte, neuere Studie zum *Hospital Readmission Reduction Program* (HRRP) erbrachte eine risikoadjustierte 30Tage-Wiederaufnahmerate von 24,6% für Patienten mit Herzinsuffizienz, 19,3% für akuten Myokardinfarkt und 18,3% für ambulant erworbene Pneumonie (Dharmarajan et al. 2017). In einer vorbildlichen Validierungsstudie für den Indikator der Wiederaufnahmerate im eigentlichen Sinn des Begriffs (s. Kap. 3.3.3.3.1.) wurde bei Medicare-Patienten eine mediane Wiederaufnahmerate von 15,5% gefunden (Krumholz et al. 2017).

Eine wichtige Rolle spielt die **Einführung von Checklisten**, gerade in der operativen Medizin. In einer sehr wichtigen, historisch kontrollierten Studie zur Wirksamkeit der *Safer Surgery Checklist* auf die peri- und postoperative Komplikationsraten in 8 verschiedenen Ländern (nicht in Deutschland) und Krankenhäusern unterschiedlicher Versorgungsstufe wurde eine Mortalität (vor/nach Intervention durch Implementierung der Checkliste) von 1,5% vs. 0,8%, andere Komplikationen (z.B. sekundäres Nierenversagen, tiefe Beinvenenthrombose und Lungenembolie) 11,0% vs. 7,0%, postoperative Wundinfektionen 6,2% vs. 3,4% und Re-Operation 2,4% vs. 1,8% festgestellt (Haynes et al. 2009, s. auch Weiser et al. 2010 und de Vries et al. 2010). In einer kanadischen Studie zum gleichen Thema wurde bei 109.341 Patienten vor und 106.370 Patienten nach der Einführung einer Checkliste eine 30Tage-Mortalität von 0,71% bzw. 0,65% und eine Komplikationsrate von 3,86% und 3,82% gefunden (Urbach et al. 2014).

Die hier zusammengestellten Studien gehören zu den größten Untersuchungen für so wichtige Parameter wie Mortalität, Wiederaufnahme und schwere Komplikationen, die verfügbar sind; wie oben angemerkt, sind sie wegen ihres Interventionscharakters von hoher Validität. Allerdings muss immer berücksichtigt werden, dass sie nicht zu epidemiologischen Zwecken durchgeführt wurden und auch nichts zur Vermeidbarkeit dieser Ereignisse beitragen können.

#### 3.6.2.5. Spezifische Datenquellen

Kurz soll an dieser Stelle auf epidemiologische Daten aus administrativen Datenquellen (Abrechnungsdaten) eingegangen werden. Wie in Kap. 3.3.5. ausführlich diskutiert, sind speziell Abrechnungsdaten bzgl. häufiger auftretender Komplikationen (z.B. nosokomialen Infektionen) sehr insensitiv und können daher hier kaum sinnvoll eingesetzt werden. Ihre Stärke liegt bei abrechnungsrelevanten *sentinel events* (z.B. belassene Fremdkörper) und in ihrer Rolle als Aufgreifkriterium in Verfahren wie dem *Peer Review*. Das in Kap. 3.3.3.3.3. aufgeführten **PSI-Set der AHRQ (PSI-90)** ist hierfür ein gutes Beispiel (s. Tab. 3.3.-3), der PSI-02 „*death in low mortality DRG*“ ist gut zu verwenden, aber schon bei der



Dekubitusrate ist wegen der Interferenz mit den Abrechnungsregeln Vorsicht angezeigt. Aus diesem Grund kann man die aus diesen Datenquellen stammenden Statistiken auch nur eingeschränkt für epidemiologische Untersuchungen verwenden.

Anders sieht es mit dem **Medicare Patient Safety Monitoring System (MPSMS)** aus (s. Kap. 3.3.3.3.3., s. Tab. 3.3.-4), das nicht nur weitaus breiter aufgestellt ist und Teile des konservativen Bereichs mit einschließt, sondern sich hinsichtlich der Datenquellen auf einer Kombination von administrativen Daten (zur Erhöhung der Reliabilität) und einer Algorithmus-gestützten Analyse von Krankenakten bezieht. Die Analyse der Krankenakten von einer externen Stelle durchgeführt. Wenn hier also eine UE-Rate von 11,1% bei der Heparin- und 4,8% bei der Antikoagulantientherapie oder von einer Rate von 11,9% beim sekundären akuten Nierenversagen nach Angiographie berichtet wird, dann kann man mit diesen Daten durchaus etwas anfangen. Daher wird dieses Verfahren in diesem Weißbuch auch empfohlen (s. Kap. 5.3.8.).

Kurz soll noch auf eine andere **Kombinationsmethode**, nämlich einer zeitnahen Analyse der Krankenunterlagen mit gleichzeitigen Interviews der an der Betreuung der Patienten beteiligten Personen, hingewiesen werden (s. Kap. 3.3.4.2.), die bei 141 Patienten in 10% mindestens ein vermeidbares UE feststellte (Wong et al. 2015).

#### 3.6.2.6. *Patient Reported Outcomes*

Die Angaben durch Patienten stellen nicht nur eine Möglichkeit dar, die epidemiologischen Aussagen zu plausibilisieren, sondern repräsentieren eine gesonderte Erkenntnisquelle, die den Fokus der bisherigen Erhebungen deutlich erweitert (s. Kap. 3.3.4.3.1.) (WHO 2017). In einer vergleichenden Untersuchung gaben 42% der Befragten und 35% der befragten Ärzte an, bei sich oder den Angehörigen einen „Fehler“ bemerkt zu haben (Blendon et al. 2002). Im neuesten Euro-Barometer (European Commission 2014) berichten 27% der befragten Patienten in Europa, selbst oder im familiären Umfeld ein UE erlebt zu haben, Deutschland liegt mit 33% oberhalb des Mittelwertes (Extremwerte zwischen Schweden 53%, Bulgarien und Österreich 11%). Die direkte, persönliche Befragung der Patienten erbringt höhere Werte als die Papier-gestützte Verfahren oder eine Hotline (O'Hara et al. 2017). Eine australische Studie an 20.000 Patienten älter als 45 Jahre erbrachte das Ergebnis, dass 7% der Befragten ein UE erlebt hatten (Walton et al. 2017). Eine US-amerikanische Studie zeigte eine Rate von 7,0% UE, die in knapp der Hälfte der Fälle einen verlängerten Krankenhausaufenthalt zur Folge hatte (Evans et al. 2006). In einer Serie internationaler Erhebungen des *Commonwealth Funds* gaben 23%, 12%, 19% und 16% der Patienten an, sie oder ihre Angehörige hätten

bei der medizinischen Behandlung einen „Fehler“ bemerkt (Schoen et al. 2005, 2007, 2009, 2011). In der letzten Aktualisierung dieser Erhebung findet sich das deutsche Gesundheitssystem auf einem Mittelfeldplatz wieder (und wie vorher in der Dimension Koordination auf dem vorletzten Platz); je 8% der deutschen Patienten gaben an, dass sie in den letzten 2 Jahren einen Fehler in der Behandlung bemerkt und ein falsches Medikament bekommen hätten und 10% gaben das Auftreten einer nosokomialen Infektion an (Davis et al. 2014). Patientenangaben haben daher auch eine zentrale Rolle in fortgeschrittenen PSI-Sets, z.B. dem *Leapfrog Healthgrade*-Instrument (s. Kap. 3.3.3.3.3.) (Leapfrog Group 2017).

In der Konsequenz kann von einer Informationsasymmetrie zuungunsten der Patienten nicht ausgegangen werden, im Gegenteil, auf der Seite der Patienten bestehen völlig adäquate Vorstellungen über die Häufigkeit von Unerwünschten Ereignissen.

### **3.6.3. Übertragbarkeit auf Deutschland und deutsche Studien**

In den Systematischen Reviews von 2006 und 2008 (Schrappe et al. 2006, 2008) wurden insgesamt 12 Studien aus Deutschland eingeschlossen (gesamt 241 Studien, s. Tab. 3.6.-1), ohne dass sich entweder im Studiendesign oder im Ergebnis ein Unterschied zu den Studien aus anderen Ländern herausgestellt hätte. Auch in der jetzt aktualisierten Analyse in den vorangehenden Kapiteln war kein Unterschied zu den unten angeführten Studien festzustellen, außerdem kann an der Übertragbarkeit der international durchgeführten Studien auf die deutschen Verhältnisse kein Zweifel bestehen. Dies gilt insbesondere für die drei niederländischen Studien, die nacheinander nach dem HMPS-Design durchgeführt worden sind (Zegers et al. 2009, Banies et al. 2013, 2015, s. Kap. 3.6.2.1.).

Eine derart umfassende Untersuchung wurde in Deutschland bislang nicht durchgeführt, da durch Systematische Review gut belegt ist, dass eine nochmalige nationale Erhebung außer einem enormen Zeitverlust für die Einleitung von Präventionsmaßnahmen keine Auswirkung hätte; international hat sich die Erkenntnis stabilisiert, dass hier auch keine zusätzlichen Informationen zu erwarten sind, und daher hat die Durchführung von nationalen Studien insgesamt auch deutlich abgenommen. Wie im nachfolgenden Kapitel und vor allem in Kap. 5.3. herausgestellt wird, liegt Handlungsbedarf derzeit vor allem in speziellen Patientengruppen und bei speziellen nosokomialen Krankheitsgruppen.

In Deutschland existieren gute Prävalenz-Untersuchungen zur Problematik der nosokomialen Infektionen, die gut als Approximation für das Gesamtproblem der Unerwünschten Ereignisse verwendet werden können. Die erste Studie von Rueden et al.

(1996) wurde bereits in die Reviews von 2006/2008 eingeschlossen. Mittlerweile sind in Abstimmung mit dem *European Centre for Disease Prevention and Control* (ECDC) in Deutschland zwei weitere Studien durchgeführt worden (Behnke et al. 2013), deren Neueste erst vor ein paar Wochen veröffentlicht wurde (Behnke et al. 2017). Die Befunde sind stabil und zeigen den Handlungsbedarf auf, insbesondere wenn man die nosokomialen Pneumonien und postoperativen Wundinfektionen mit einer Prävalenz von je über 1% betrachtet (s. Tab. 3.6.-5). Ohne dass die Vermeidbarkeit dieser Komplikationen hier im Einzelfall berichtet wurde, geben diese beiden schweren nosokomialen Infektionen hinsichtlich der vermuteten Rate von Unerwünschten Ereignissen zwischen 5 und 10% in Deutschland einen wichtigen Fingerzeig, dass nämlich bereits durch diese beiden Entitäten mehr als 2% die UE beigesteuert werden. Zur Vermeidbarkeit mögen folgende Angaben einen indirekten Hinweis geben: 2011 lag der mediane Händedesinfektionsmittelverbrauch bei 24,5ml pro Patiententag, 2016 bei 32,5ml, was (bei einem mittleren Verbrauch von 3ml pro Desinfektion) zwischen 8 und 10 Desinfektionsmaßnahmen pro Patient pro Tag für alle Mitarbeiter bedeutet, die mit diesem Patienten zu tun hatten (und für die entsprechenden Besucher).

	<b>NIDEP I*</b>	<b>NIDEP II**</b>	<b>ECDC***</b>
Patienten n=	14966	41539	64412
Krankenhäuser	72	132	218
Prävalenz NI	3,46%	5,08%	4,58%
Prävalenz NI neu aufgetreten	k.A.	3,76%	3,32%
Harnwegsinfektionen	1,46%	1,26%	1,04%
Pneumonien	0,72%	1,17%	1,16%
Postoperative Wundinfektionen	0,55%	1,31%	1,08%
Nosokomiale Sepsis (primär)	0,29%	0,26%	0,24%
<i>Clostridium difficile</i>	k.A.	0,34%	0,48%

Tab. 3.6.-5: Ergebnisse der bislang in Deutschland durchgeführten drei Studien zur Prävalenz von nosokomialen Infektionen. NI Nosokomiale Infektion, NIDEP Nosokomiale Infektionen in Deutschland – Erfassung und Prävention, ECDC *European Centre for Disease Prevention and Control*. \*Rueden et al. 1996, \*\*Behnke et al. 2013, \*\*\*Behnke et al. 2017

In einer monozentrischen Studie an der Medizinischen Hochschule Hannover erwarben 124/1047 (11,2%) der Patienten während des stationären Aufenthalts ein nosokomiale Infektion (29% postoperative Wundinfektion, 26% gastrointestinale Infektion); bei weiteren 11,2% bestand eine nosokomiale Infektion bereits bei Aufnahme (Ott et al. 2013). Die Daten zu den postoperativen Wundinfektionen können durch die Ergebnisse der oben bereits angeführten *European Surgical Outcome Study* durch Angaben zur postoperativen Sterblichkeit ergänzt werden, die für Deutschland bei 2,5% aller operierten Patienten liegt (Pearse et al. 2012). In einer sicherlich an einem Hochrisikokollektiv durchgeführten Untersuchung zur Bedeutung der Messinstrumente und Datenquellen (s. Kap. 3.3.5.) wurden in drei deutschen Kliniken unterschiedlicher Versorgungsstufe bei 492/3000 (18%) Patienten ein UE festgestellt (Maass et al. 2015).

Hinsichtlich der vermeidbaren Mortalität durch UE in Deutschland können auch am ehesten die Daten des Robert-Koch-Institutes herangezogen werden, das für das Jahr 2006 in Deutschland mit 400.000 bis 600.000 nosokomiale Infektionen rechnet, die bei 7.500 bis 15.000 Patienten die Todesursache darstellen (Gastmeier et al. 2008). Geht man davon aus, dass davon 1/3 vermeidbar gewesen wären, kommt man auf 2.500-5.000 vermeidbare Todesfälle allein durch nosokomiale Infektionen (Gastmeier et al. 2010 gehen von 25% Vermeidbarkeit aus). In einer neueren Darstellung für die Untergruppe der nosokomialen Infektionen, die in Deutschland durch mehrfachresistente Erreger auftreten, kommen die Autoren zu einer Sterblichkeit zwischen 1000 und 4000 Patienten pro Jahr, wovon ein Drittel vermeidbar sei (Gastmeier et al. 2016).

Zusammenfassend kann für die deutsche Situation davon ausgegangen werden, dass der bereits oben zitierte Sachstand

- UE: zwischen 5% und 10%,
- VUE: zwischen 2% und 4%,
- Behandlungsfehler bei 1% und
- Vermeidbare Mortalität bei 0,1%

für die Krankenhauspatienten unverändert Bestand hat. Sowohl im internationalen als auch im Rahmen des deutschen Gesundheitssystems hat sich keine Änderung der Situation im Vergleich zum Systematischen Review vor 10 Jahren ergeben, allerdings hat sich durch die zur Verfügung stehenden Interventionsstudien die Belastbarkeit der Daten verbessert. Die Daten zur vermeidbaren Mortalität müssen genau analysiert werden; die Studien, die hier aussagekräftige Angaben machen, lassen vermuten, dass die Zahlen eher oberhalb der genannten 0,1% liegen.

### **3.7. Zusammenfassung**

Quantitative Aussagen zum Thema Patientensicherheit sind nicht nur wegen der Standortbestimmung von größter Wichtigkeit, sondern vor allem als Grundlage für jeden Verbesserungsansatz. Aus diesem Grund wird in diesem Weißbuch Patientensicherheit auf die kritische Darstellung der Erhebungsmethodik ein so großes Gewicht gelegt. Eine sinnvolle Diskussion der Erhebungsmethodik kann jedoch nur auf der Basis eines adäquaten Konzeptes von Patientensicherheit vorgenommen werden. In der hier vorgelegten Analyse stellt sich allerdings heraus, dass die Erhebungsmethodik fast ausnahmslos auf einem Verständnis von Patientensicherheit beruht, das Sicherheit als *end result*, also als Endergebnis (*outcome*) eines abgelaufenen Prozesses versteht. Gemessen werden Mortalität und Komplikationsraten, aber Elemente der Definition aus Kap. 2.6.2., die die Eigenschaften der Akteure und deren Innovationskompetenz betreffen, werden fast gar nicht berücksichtigt. Am ehesten wären Befragungen von Patienten zu ihren Erfahrungen oder von Mitarbeitern zum Thema Patientensicherheitskultur in der Lage, diese Lücke zu schließen (weitere Vorschläge in Kap. 5.3.).

Im vorliegenden Kapitel 3 wird jedoch zunächst auf die Grundlagen eingegangen. Da *end results*-bezogene Erhebungskonzepte immer noch so häufig angewandt werden und sich der linearen Terminologie aus dem Prozessverständnis (s. Kap. 2.4.4.) bedienen, wird die lineare Begrifflichkeit in Kap. 3.2. nochmals rekapituliert. In der Folge wird dann die Erhebungsmethodik dargestellt, beginnend mit der Zielorientierung, der Bestimmung des Erkenntnisinteresses (z.B. wissenschaftlicher Ansatz), der Auswahl des Messinstrumentes (z.B. Befragungen) und letztlich den Datenquellen (s. Kap. 3.3.). Die **Zielorientierung** wurde in Kap. 2.3.3. bereits operationalisiert, indem drei Dimensionen vorgeschlagen wurden: die Perspektivdimension (z.B. Anbieterperspektive, Patientenperspektive), die Strukturdimension (wie steht das Thema zu den drängenden Strukturentwicklungen des Systems) und die Bedarfsdimension (wird die relevante Morbidität angesprochen etc.).

Bei der **Diskussion des Erkenntnisinteresses** (s. Kap. 3.3.3.) werden ganz zentrale Fragen beleuchtet. Nicht nur dass das Thema Patientensicherheit eine große Besonderheit mit sich bringt, indem es den generierenden Verfahren wie CIRS einen prominenten Platz einräumt (zur Behebung der doppelt blinden Flecken, den *Unknown Unknowns*), sondern es geht um große Bedeutung der klinisch-epidemiologischen Perspektive. Der Großteil der epidemiologischen Studien zur Häufigkeit von UE nimmt diese Perspektive ein, es geht darum, den *Status quo* zu beschreiben, das Problem zu quantifizieren. Das große Vorbild der Infektiologie bzw. Krankenhaushygiene hat hier

Jahrzehnte Vorsprung, und dort kann man sich vieles anschauen, was die Erstellung von klinisch-epidemiologischen Falldefinitionen angeht. Allerdings werden (gerade in Deutschland) diese Falldefinitionen nicht von Patientensicherheitsindikatoren differenziert, die ihrerseits die Aufgabe haben, größere Versorgungsbereiche einem Monitoring zu unterwerfen (Ampelfunktion). So sind die meisten Parameter, die in Deutschland (und gelegentlich auch international) als Patientensicherheitsindikatoren (PSI) Verwendung finden, gar keine Indikatoren im eigentlichen Sinn. Dies hat erhebliche Folgen, denn die statistischen Anforderungen unterscheiden sich deutlich: klinisch-epidemiologische Falldefinitionen weisen eine ausgeglichene Sensitivität und Spezifität auf, Indikatoren zu Zwecken des Monitorings dagegen eine hohe Sensitivität unter Verzicht auf zu hohe Anforderungen an die Spezifität. Die meisten PSI-Sets, dies wird ausführlich dargestellt, enthalten ausschließlich klinisch-epidemiologische Falldefinitionen, die das *Outcome* „abzählen“ (was ja durchaus wichtig sein kann), weisen aber weder eine Monitoring-Funktion für größere Versorgungsbereiche auf noch lassen sie als *end results* eine Aussage über die Verwirklichung von Sicherheit zu (s.u.). Zusätzlich sind manche „PSI-Sets“, die aus klinisch-epidemiologischen Falldefinitionen bestehen, durch eine schlechte Sensitivität in ihrer Aussagekraft geschwächt, insbesondere bei der Verwendung von Abrechnungsdaten (massive Untererfassung z.B. von Komplikationen). Viertens sind noch wissenschaftliche Fragestellungen abzugrenzen, die z.B. der Evaluation von Verbesserungsmaßnahmen dienen und aufwendige Settings notwendig machen, die die komplexe Interaktion von Beobachtung, der komplexen Intervention und des aktiven Kontext zum Gegenstand haben. Es sind also vier Perspektiven des Erkenntnisinteresses zu unterscheiden.

Die Thematik des Erkenntnisinteresses führt wieder auf das Verständnis von Patientensicherheit zurück. Versteht man unter Patientensicherheit lediglich die „Abwesenheit von Unerwünschten Ereignissen“ (s. IOM 1999), geht man also ausschließlich vom End-Ergebnis aus, dann kann man mit der Quantifizierung von Falldefinitionen zufrieden sein. Versteht man aber unter Patientensicherheit im Sinne der Definition von Kap. 2.6.2. nicht nur einen Zustand (keine Unerwünschten Ereignisse), sondern die Eigenschaft von Organisationen und System, mit Unsicherheit umzugehen, und beschreibt man mit Patientensicherheit außerdem noch die Kompetenz zur Umsetzung von Innovationen (insbesondere Struktur- und Prozessinnovationen), dann kann man mit der Betrachtung von End-Ergebnissen nicht zufrieden sein. Erhebungsmethodik und Konzept sind unmittelbar aufeinander bezogen. In Kap. 5.3.8. wird ein Erhebungskonzept, das diese Aspekte adäquat berücksichtigt, vorgestellt.

Nach Zielorientierung und Bestimmung des Erkenntnisinteresses folgt als drittes Element die **Wahl des Messinstrumentes**. Hier stehen Methoden wie die direkte Beobachtung, die Befragung, der Einsatz von Trigger-gestützten Instrumenten (der Klassiker des

*Harvard Medical Practice Study* (HMPS)-Designs, das Global Trigger Tool (GTT), das *Medicare Patient Safety Monitoring System* (MPSMS)) und die *Big Data*-Analyse zur Auswahl. Das MPSMS ist eines der fortgeschritteneren Instrumente, da es administrative Daten mit einem externen, streng standardisierten *Chart Review* verbindet; es wird derzeit zum *Quality and Safety Review System* (QSRS) weiterentwickelt, das auch die Analyse unstrukturierter Daten beinhaltet. Aber es sind auch Instrumente beschrieben, die *Chart Review* und Mitarbeiterbefragungen miteinander verbinden.

Abschließend muss die **Datenquelle** ausgewählt werden. Im Mittelpunkt steht die immer wieder präferierte Verwendung von Routine- bzw. Abrechnungsdaten, die zwar ihre Stärken bei seltenen, abrechnungsrelevanten Ereignissen haben (z.B. belassene Fremdkörper), aber - wie oben bereits angemerkt - große Defizite in häufiger auftretenden, für die Patienten sehr relevanten, in der Abrechnung aber nicht immer erfassten Komplikationen wie nosokomialen Infektionen aufweisen. In sehr vielen internationalen Studien und auch in Deutschland ist dies nachgewiesen, wird aber in Deutschland mit der Begründung des Erhebungsaufwandes immer noch weiter praktiziert. Bei auf Abrechnungsdaten basierenden „PSI-Sets“ besteht also eine doppelte Problematik: sie sind nicht valide, weil sie weder auf eine Monitoringfunktion validiert noch auf relevante Sicherheitsaspekte abgestellt sind, und sie sind außerdem nicht reliabel, weil sie die Ereignisse nicht vollständig erfassen (und können deswegen natürlich erst recht nicht valide sein).

Dieses in vier Schritten ablaufende Vorgehen widerspricht der häufig verwendeten Praxis, dass man zunächst die vorhandenen Datenquellen („was haben wir denn schon?“) betrachtet und daraus zunächst die Methodik und zuletzt die zu beantwortenden Fragestellungen ableitet. Diese Aussage ist in hohem Maße „politisch“, denn ein Datenverfügbarkeits- und Methoden-orientiertes Herangehen hat zwei Konsequenzen:

- zahlreiche, oft die relevantesten Fragestellungen werden von vorneherein ausgeschlossen,
- die Untersuchungen führen zu vorhersehbar negativen Ergebnissen, wenn z.B. irrelevante Datenquellen oder invalide Messinstrumente verwendet werden.

Es ist aus dieser Sicht klar zu erkennen, dass die Zukunft in **Kombinationsmethoden** liegt, in denen Krankenaktenanalysen und (kritisch) Abrechnungsdaten kombiniert werden und man vermehrt auf *Patient Reported Outcomes* zurückgreift. Das schon genannte *Medicare Patient Safety Monitoring System* (MPSMS) der *Centers of Medicare and Medicaid Services* der USA ist derzeit das attraktivste Konzept, das in diese Richtung weist, und sollte in einer angepassten Form in Deutschland pilotiert und eingesetzt werden (unter Beachtung der o.g. Erweiterung in das QSRs).

Im letzten Teil dieses Kapitels werden die internationalen und (soweit vorhanden) in Deutschland durchgeführten **Studien zur Epidemiologie** zusammengestellt. Dies kann nicht wie 2006/2008 in Form eines Systematischen Reviews geschehen, dafür ist aber die Validität der heute zur Verfügung stehenden Studien sehr viel besser als vor 10 Jahren. Zum einen liegen serielle Untersuchungen (z.B. in den Niederlanden), zum anderen liegen Ergebnisse von Interventionsstudien vor, die zwar keine Auskunft zur Vermeidbarkeit geben können, aber sehr verlässliche Aussagen zur Größenordnung der wichtigsten Unerwünschten Ereignisse. Zusammenfassend lassen sich die Ergebnisse wie folgt darstellen:

- zu (vermeidbaren) UE liegen sechs Studien nach dem HMPS-Design vor (Krankenhaus: mind. 1 UE bei zwischen 5,7 und 12,3% der Patienten, Vermeidbarkeit zwischen 20 und 70% der UE),
- außerdem 5 Studien mit dem *Global Trigger Tool* (Krankenhaus, mind 1 UE zwischen 13,5 und 33,2%, Vermeidbarkeit zwischen 44 und 63% der UE), und
- vier Systematische Reviews (UE zwischen 5,7 und 14,4%).

Besonders sorgfältig wird dabei auf die Frage der **vermeidbaren Mortalität** eingegangen. Diese ist epidemiologisch nicht einfach zu erfassen, da zum Einen die Behandlungsbedingtheit (Zurechenbarkeit, s. Kap. 3.5.), zum anderen die Vermeidbarkeit (verursacht durch einen Fehler) dokumentiert werden muss. Zwei große Studien erfassen diesen Wert, er liegt in einer schwedischen Studie nach dem HMPS-Design bei 0,25% aller Patienten (Soop et al. 2009) und in einer US-amerikanischen Studie bei 0,4% (Landrigan et al. 2010). In einer dritten Studie (Classen et al. 2011) werden die UE den Vermeidbaren UE gleichgesetzt, berichtet wird einen Mortalität von 1% aller Patienten.

Weiterhin wird als Näherung die Sterblichkeit der Unerwünschten Ereignisse berichtet, die jedoch nicht mit der vermeidbaren Mortalität identisch ist (Fehler werden nicht erfasst):

- bei den HMPS-analogen Studien liegt die Sterblichkeit der UE zwischen 6,7% und 10% (s. Tab. 3.6.-2),
- bei den GTT-Studien niedriger bei 1,5% (DHHS 2010) bzw. 2% (Classen et al. 2011) (s. Tab. 3.6.-3) und
- bei den Systematischen Reviews bei 3,6% und 7,4% (s. Tab. 3.6.-4; die Zusammenstellung der Studien von James (2003) wird hier nicht berücksichtigt),

Das Ergebnisse lassen sich klar zusammenfassen und den Ergebnisse der Systematischen Reviews 2006/2008 ist nichts hinzuzufügen: in Deutschland muss im Krankenhausbereich mit jährlich

- UE: zwischen 5% und 10%,
- VUE: zwischen 2% und 4%,
- Behandlungsfehler bei 1% und einer



vermeidbaren Mortalität bei 0,1%

gerechnet werden. Es gibt kein Argument gegen die Übertragbarkeit der zitierten Studien auf Deutschland, außerdem gibt es hiesige Studien z.B. zu nosokomiale Infektionen. Die genannten Zahlen waren schon damals sehr konservativ gerechnet und überschätzen die Situation auch heute nicht. Vermeidbare (auf Fehler zurückzuführende) Unerwünschte Ereignisse treten jährlich bei zwischen 400.000 und 800.000 Krankenhauspatienten auf. Die in Deutschland vom Aktionsbündnis Patientensicherheit vertretene Angabe einer vermeidbaren Mortalität von 0,1% ist belastbar und entspricht bei rund 20 Mill. Krankenhauspatienten einer vermeidbaren Mortalität von 20.000 Patienten pro Jahr, bei 420.000 Sterbefällen ist also ca. jeder 20. Sterbefall im Krankenhaus in Deutschland als vermeidbar (auf einen Fehler zurückführbar) einzustufen. Hinzu kommt, dass Ereignisse aus der ambulanten und pflegerischen Versorgung ebenso wenig in dieser Zahl enthalten sind wie die Ereignisse, die aus diagnostischen Fehlern, *errors of omission* und aus Überversorgung resultieren.