



# Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten – Begriff, Studientypen und ethische Herausforderungen

Martin Jungkunz, Anja Köngeter, Eva C. Winkler, Katja Mehlis und Christoph Schickhardt

## 1 Einleitung

Jeder Besuch beim Haus- oder Facharzt und jeder Klinikaufenthalt ist mit der Erzeugung von Daten verbunden. Laborergebnisse, Medikationspläne, Diagnosen und vieles mehr werden in zunehmendem Maße digital verarbeitet und gespeichert, und dienen Dokumentations-, Behandlungs- und Abrechnungszwecken. Zudem bergen Daten aus der Krankenversorgung auch Potenziale für den wissenschaftlichen Fortschritt im biomedizinischen Bereich und

---

M. Jungkunz (✉)

Nationales Centrum für Tumorerkrankungen (NCT), Sektion für Translationale Medizinethik, Universitätsklinikum Heidelberg, Heidelberg, Deutschland

E-Mail: [martin.jungkunz@med.uni-heidelberg.de](mailto:martin.jungkunz@med.uni-heidelberg.de)

A. Köngeter

Nationales Centrum für Tumorerkrankungen (NCT), Sektion für Translationale Medizinethik, Universitätsklinikum Heidelberg, Heidelberg, Deutschland

E-Mail: [koengeter@med.uni-heidelberg.de](mailto:koengeter@med.uni-heidelberg.de)

E. C. Winkler

Nationales Centrum für Tumorerkrankungen (NCT), Sektion für Translationale Medizinethik, Universitätsklinikum Heidelberg, Heidelberg, Deutschland

E-Mail: [eva.winkler@med.uni-heidelberg.de](mailto:eva.winkler@med.uni-heidelberg.de)

K. Mehlis

Nationales Centrum für Tumorerkrankungen (NCT), Sektion für Translationale Medizinethik, Universitätsklinikum Heidelberg, Heidelberg, Deutschland

E-Mail: [katja.mehlis@med.uni-heidelberg.de](mailto:katja.mehlis@med.uni-heidelberg.de)

C. Schickhardt

Nationales Centrum für Tumorerkrankungen (NCT), Sektion für Translationale Medizinethik, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Deutschland

E-Mail: [Christoph.Schickhardt@med.uni-heidelberg.de](mailto:Christoph.Schickhardt@med.uni-heidelberg.de)

können damit zu einer Verbesserung der Patientenversorgung beitragen. Diese Potenziale lassen Stimmen laut werden, Daten aus der klinischen Versorgung systematisch datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten für eine Sekundärnutzung zuzuführen. In den letzten Jahren haben sich unter anderem in den USA und in Großbritannien mehrere Initiativen mit dem Ziel einer Sekundärnutzung von Daten aus der Krankenversorgung für die Forschung formiert (Warren-Gash 2017; Committee on the Learning Health Care System in America 2013; Schilsky et al. 2014). Auch in Deutschland besteht ein reges Interesse an der Sekundärnutzung von Daten aus der Krankenversorgung. Dies zeigt sich z. B. in der vom Bundesministerium für Bildung und Forschung geförderten Medizininformatikinitiative (Bundesministerium für Bildung und Forschung 2015, siehe hierzu auch den Beitrag von Wiebke Lesch in diesem Band), aber auch in Vorstößen des Bundesministeriums für Gesundheit, Versorgungsdaten der Krankenkassen standardmäßig für die Forschung nutzbar zu machen (Digitale-Versorgung-Gesetz – DVG (Fn.: BT-Drs. 19/13438)).

Trotz der offensichtlichen Relevanz des Themas wurden begriffliche Fragen der Sekundärnutzung von Daten aus der Krankenversorgung – im Folgenden als „klinische Daten“ bezeichnet – bisher in der Literatur wenig erörtert. Mit dem vorliegenden Artikel soll ein Beitrag zur Schließung dieser Forschungslücke geleistet werden. So besteht zum einen das Desiderat einer begrifflichen Klärung dessen, was unter Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten zu verstehen ist. Dieser Aufgabe widmet sich der erste Abschnitt des vorliegenden Artikels sowohl in Bezug auf den Begriffsinhalt als auch seinen Umfang. Es werden im Anschluss drei Anwendungsfelder für die Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten aufgezeigt. Im nächsten Abschnitt analysieren wir neben möglichen Nutzenpotenzialen auch die Herausforderungen für die Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten. Für eine verantwortungsvolle Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten ist zudem eine Abschätzung möglicher negativer Folgen für Patienten<sup>1</sup> wie auch für Ärzte, Pflegepersonal und Kliniken unabdingbar. Aus diesem Grund zeigen wir wichtige Risiken und ethische Implikationen auf. Im Anschluss geben wir einen Ausblick auf mögliche Maßnahmen der Risikoreduzierung.

---

## 2 Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten<sup>2</sup>

Die American Medical Informatics Association (AMIA) definiert die Sekundärnutzung von klinischen Daten als “non-direct care use of [personal health information (PHI)] including but not limited to analysis, research, quality/safety measurement, public health,

---

<sup>1</sup> Aus Gründen der besseren Lesbarkeit verwenden wir im Folgenden ausschließlich das generische Maskulinum wobei selbstverständlich stets alle Geschlechter bezeichnet werden.

<sup>2</sup> Die vorliegende Analyse stellt eine Präzisierung eines Konzepts dar, welches an anderer Stelle veröffentlicht wurde (Jungkunz et al. 2021). Im vorliegenden Artikel bauen wir auf den wichtigsten Ideen dieser Publikation auf und erweitern sie um einige wichtige Aspekte.

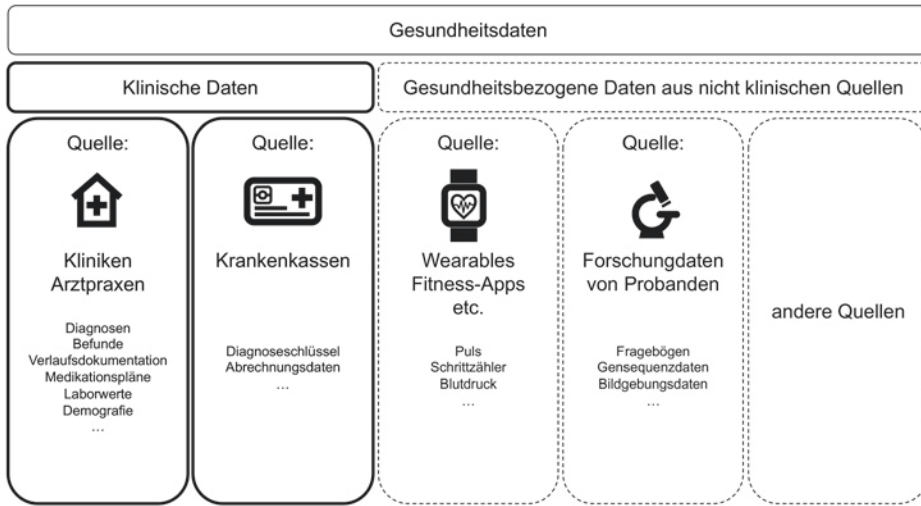
payment, provider certification or accreditation, and marketing and other business including strictly commercial activities” (Safran et al. 2007). Ähnlich allgemein äußern sich weitere Publikationen. So identifiziert ein systematisches Literaturreview von Robertson et al. die Bereiche „Forschung“, „Qualitätssicherung und Sicherheit in der Versorgung“, „Finanzverwaltung“ und „Bildung“ als Anwendungsfelder der Sekundärnutzung klinischer Daten (Robertson et al. 2016). Andere Autoren benennen ähnliche allgemeine Kategorien (Heatherly et al. 2013; Ienca et al. 2018; American Medical Informatics Association 2007). Eine detaillierte begriffliche Definition und Beschreibung der mit Sekundärdaten-nutzung durchführbaren Forschungs- oder Lernaktivitäten fehlt bisher. Diese Lücke füllen wir im nächsten Abschnitt durch Beschäftigung mit dem Begriff, dem Umfang und den Anwendungsfeldern der „Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionelle Forschungs- oder Lernaktivitäten“. Zuerst analysieren wir den Inhalt (Sinn) des Begriffs „Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten“ und zeigen im Anschluss daran auf, welche Arten von Studien unter dem Begriff subsumiert werden können (Umfang). Im weiteren Verlauf des Artikels nehmen wir auf den Begriff mit dem Akronym „SeConts“ Bezug, das auf den englischen Begriff „Secondary use of clinical data in data-gathering, non-interventional research or learning activities“ zurückgeht (Jungkunz et al. 2021).

## 2.1 Sinn des Begriffs SeConts

Wenn wir von Sekundärnutzung klinischer Daten in datensammelnden, nicht interventionellen Forschungs- oder Lernaktivitäten (SeConts) sprechen, so stellt sich zuallererst die Frage, was mit dem Begriff der **klinischen Daten** gemeint ist. Klinische Daten sind nach unserer Begriffsdeutung ausschließlich Daten, die im Kontext der Krankenversorgung in Krankenhäusern und Arztpraxen erhoben werden. Dies können zum einen Daten sein, die direkt mit der Behandlung zusammenhängen, wie z. B. Diagnosen, Befunde, Verlaufsdokumentation, Medikationspläne oder Laborwerte. Zum anderen geht es bei klinischen Daten auch um Daten aus dem Abrechnungswesen von Versorgungsleistungen, die z. B. Alter, Beschäftigungsstatus und andere soziodemographische Informationen beinhalten. Bei allen genannten Formen von Daten ist es nach unserem Verständnis des Begriffs „klinische Daten“ unerheblich, ob die Daten aus der Administration der Kliniken selbst bezogen werden, zukünftig über eine selbstverwaltete elektronische Patientenakte oder von den Krankenkassen. In allen Fällen sprechen wir von klinischen Daten. Als Überkategorie zum Begriff der klinischen Daten steht derjenige der Gesundheitsdaten<sup>3</sup>, welcher zwar auch klinische Daten beinhaltet, jedoch in seinem Umfang über

---

<sup>3</sup>Gesundheitsdaten sind „personenbezogene Daten, die sich auf die körperliche oder geistige Gesundheit einer natürlichen Person, einschließlich der Erbringung von Gesundheitsdienstleistungen, beziehen und aus denen Informationen über deren Gesundheitszustand hervorgehen;“ (Artikel 4, Abs. 15, DSGVO).



**Abb. 1.** Klinische Daten und andere Gesundheitsdaten

diese hinausgeht. So werden z. B. auch Daten als Gesundheitsdaten angesehen, die mit sogenannten Wearables (tragbare Geräte wie Fitness-Armbänder) und Gesundheitsapps etc. gewonnen werden und einen (expliziten) Gesundheitsbezug haben, jedoch nicht im klinischen Kontext erhoben werden. Verschiedene Arten von Gesundheitsdaten inklusive der für diesen Artikel relevanten klinischen Daten werden in Abb. 1 dargestellt.

Der Begriff der **Sekundärnutzung** klinischer Daten impliziert die Existenz einer vorhergehenden Primärnutzung<sup>4</sup>. Die Primärnutzung wurde bereits mit der Erläuterung der Herkunft klinischer Daten angedeutet. Klinische Daten dienen in ihrer ursprünglichen (primären) Verwendung der Patientenversorgung bzw. – im Falle von Abrechnungs- und Verwaltungsdaten – der Organisation und Durchführung jener Versorgung. Im Rahmen von SeConts werden keinerlei neue Daten erhoben. Vielmehr werden bereits aus der

<sup>4</sup>Eine derartige Unterscheidung zwischen primärer und sekundärer Nutzung ist nicht unumstritten, insofern sie eine Trennung zwischen Forschung und Versorgung impliziert. Diese Trennung wird jedoch mancherorts kritisiert (Kass et al. 2013; Faden et al. 2013). Man spricht sich vielmehr – gerade im Rahmen der Vision eines sogenannten „Lernenden Gesundheitssystems“ (Learning Healthcare System (LHCS)) – dafür aus, Forschung und Versorgung näher zusammenwachsen zu lassen. Die Trennung von Primär- und Sekundärnutzung scheint im Rahmen eines LHCS auf den ersten Blick nicht mehr angemessen. Es bleibt jedoch zu bedenken, dass die Erhebung klinischer Daten auch bei einer engeren Verzahnung von Versorgung und Forschung weiterhin das primäre Ziel der Patientenversorgung im Blick hat. Darüber hinaus wird in ethischen und rechtlichen Kodizes immer noch die Unterscheidung zwischen Versorgung und Forschung beibehalten und wird in den kommenden Jahren wahrscheinlich nicht aufgegeben.

**Tab. 1.** Anwendungsfelder von SeConts und Studientypen

Anwendungsfelder von SeConts	Studientypen:
Mikroebene: nicht-interventionelle klinische Forschung	<ul style="list-style-type: none"> <li>• Modellierung von klinischen Studien in-silico</li> <li>• vergleichende Wirksamkeitsstudien</li> <li>• Studien zur Arzneimittelsicherheit und -wirksamkeit</li> <li>• Studien zu Risikofaktoren verschiedener Krankheiten</li> <li>• ...</li> </ul>
Mesoebene: Aktivitäten zur Kontrolle und Erforschung der Versorgung (oder bestimmter Versorgungssituationen)	<ul style="list-style-type: none"> <li>• Infektionskontrolle in Kliniken</li> <li>• Früherkennung z. B. von multiresistenten Keimen</li> <li>• ...</li> </ul>
Makroebene: Public Health Forschung	<ul style="list-style-type: none"> <li>• Infektionskontrolle auf gesellschaftlicher Ebene</li> <li>• Epidemiologische Studien unterschiedlicher Art</li> <li>• Outcome-Forschung</li> <li>• Versorgungsforschung</li> <li>• Registerstudien</li> <li>• ...</li> </ul>
kommt in allen Anwendungsfeldern vor	<ul style="list-style-type: none"> <li>• Explorative Datennutzung (Hypothesengenerierung, Machbarkeitsstudien, Studienrekrutierung)</li> <li>• Medizininformatikforschung</li> <li>• ...</li> </ul>

Versorgung vorhandene **Daten gesammelt** und zu einem anderen Zweck als dem der Versorgung verwendet. Dies bedeutet, dass die sekundäre Nutzung **keinerlei zusätzliche Intervention** am Patienten voraussetzt oder mit sich bringt, welche dem Zweck der Erhebung von Daten zu Forschungszwecken dienen würde. Dieses Verständnis von Sekundärnutzung schließt nicht aus, dass die zu nutzenden Daten im Zuge von Interventionen erzeugt wurden, welche zu Diagnose- oder Therapie Zwecken am Patienten vorgenommen wurden.

Wenn wir von der Sekundärnutzung klinischer Daten im Rahmen von **Forschungs- oder Lernaktivitäten** sprechen, so sind die Begrifflichkeiten Forschung und Lernen – im technischen Sinne eines Erkenntnisgewinns durch die Auswertung klinischer Daten – schwer trennbar und die Grenzen zwischen ihnen fließend. In beiden Fällen geht es um die Generierung verallgemeinerbaren Wissens unter alleinigem Rückgriff auf die Sammlung und Nutzung klinischer Daten. Im Bereich der Forschung liegt der primäre Fokus zunächst auf dem Erkenntnisgewinn und eher indirekt und in der Folge der wissenschaftlichen Veröffentlichung der Ergebnisse auf der Verbesserung der medizinischen Versorgung.

Unter Lernaktivitäten verstehen wir Aktivitäten, die sich von Forschung dahingehend unterscheiden, dass das primäre Interesse in der unmittelbaren Sicherung oder Verbesserung einer konkreten Versorgungssituation besteht. Lernaktivitäten können auf eine bestimmte Klinik oder Station bezogen sein, wie z. B. bei der Infektionskontrolle einer

bestimmten Abteilung. Ebenso können sich Lernaktivitäten auf ganze Gesellschaften beziehen, wie dies z. B. das Robert Koch-Institut für Deutschland in Hinblick auf die Infektionskontrolle zur Aufgabe hat. Selbstverständlich können auch Lernaktivitäten zu Erkenntnissen führen, die Gegenstand einer wissenschaftlichen Publikation werden. Ebenso ist im Falle von Erkenntnissen, welche über Lernaktivitäten einzelner Kliniken gewonnen werden, eine Weitergabe des Wissens auch in anderer als wissenschaftlicher Form möglich. Eine solche Weitergabe kann etwa in Klinikverbänden geschehen, wenn z. B. Erkenntnisse zur Verbesserung des Infektionsmanagements einer einzelnen Klinik an den restlichen Verbund weitergegeben werden. Das primäre Ziel von Lernaktivitäten bleibt jedoch die zeitnahe Kontrolle und Verbesserung einer konkreten Versorgungssituation. Dieser Fokus bedeutet nicht, dass Lernaktivitäten direkt denjenigen Patienten zugutekommen, deren Daten verwendet werden. Ein derartiger direkter Nutzen für den Patienten ist eher unwahrscheinlich, da auch Qualitätssicherungs- und Verbesserungsprozesse innerhalb einzelner Kliniken eine gewisse Implementierungszeit beanspruchen.

Fassen wir die vorangehende begrifflichen Erläuterungen zusammen, lässt sich Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten (SeConts) beschreiben als *Nutzung von Daten aus der medizinischen Versorgung für Forschungs- oder Lernaktivitäten, ohne körperliche Eingriffe oder zusätzliche Maßnahmen zur Datengenerierung, zum Zweck der Verbesserung des biomedizinischen Wissens und der medizinischen Versorgung, jedoch nicht zum Eigennutzen des Patienten.*

## 2.2 Umfang des Begriffs SeConts: Studientypen

Es gibt ein weites Spektrum an Forschungs- oder Lernaktivitäten, welche unter SeConts subsumiert werden können. Beispiele wären hier die Modellierung von klinischen Studien in-silico (Weiner et al. 2008), vergleichende Wirksamkeitsstudien (comparative effectiveness studies) (Bronsert et al. 2013; Institute of Medicine 2009), Studien zur Arzneimittelsicherheit und -wirksamkeit (Kuter et al. 2013) oder auch Studien zu Risikofaktoren verschiedener Krankheiten (Mathews et al. 2013; Conway et al. 2007). Ebenso ist es möglich, individuelle Heilversuche (z. B. durch einen Off-Label Use von Medikamenten) retrospektiv zu analysieren (Oshikoya et al. 2019). Weiterhin kann SeConts der Infektionskontrolle in Kliniken (Samore et al. 1997; Evans et al. 1992) oder der Früherkennung z. B. von multiresistenten Keimen (Pittet et al. 1996) dienen. Auf gesamtgesellschaftlicher Ebene sind gesellschaftliche Infektionskontrolle (Smith et al. 2007) und epidemiologische Studien unterschiedlicher Art (Robinson et al. 2001; Mitchell et al. 2014) sowie Outcome-Forschung (Been et al. 2015) oder Versorgungsforschung (Hay und Hay 1992) mit klinischen Daten durchführbar. Ebenso können hier die bereits gut etablierten sogenannten Registerstudien (Stattin et al. 2010), die eher eine Methode als einen Studientyp beschreiben, genannt werden.

Betrachtet man den jeweiligen Forschungsgegenstand der unterschiedlichen genannten Arten von SeConts, so ist eine Kategorisierung in drei übergeordnete Bereiche möglich.

Der Forschungsgegenstand kann entweder der individuelle Patient (Mikroebene), die klinische Versorgungseinheit (Mesoebene) oder die Gesellschaft (Makroebene) sein. Entsprechend dieser Ebenen ergibt sich eine Einteilung von SeConts in nicht-interventionelle klinische Forschung (Mikroebene), Aktivitäten zur Kontrolle, Verbesserung und Erforschung der Versorgung (oder bestimmter Versorgungssituationen) (Mesoebene) und Public Health Forschung (Makroebene). Darüber hinaus ist in allen drei Bereichen eine explorative Verwendung der Daten möglich. Hierunter fällt insbesondere die Hypothesengenerierung – für welche manche Autoren (Braun und Dabrock 2016; Lee et al. 2015) ein großes Potenzial in der Datenanalyse durch Künstliche Intelligenz (KI) bzw. Maschinellem Lernen (ML) sehen. Über die hypothesengenerierende Verwendung (ob mit oder ohne KI bzw. ML) hinaus ermöglicht die Sekundärnutzung klinischer Daten die Durchführung von Machbarkeitsstudien<sup>5</sup> im Vorfeld geplanter Studien und ggf. sogar der Identifizierung möglicher Teilnehmer für geplante bzw. laufende klinische Studien<sup>6</sup> (Kopcke et al. 2013). Neben den genannten Beispielen gibt es weitere Studientypen, welche mehreren Anwendungsfeldern zugeordnet werden können, wie z. B. Forschung in der Medizin-informatik, welche u. a. an Diagnostiksoftware zur Unterstützung von Ärzten oder an einer Verbesserung der Datensicherheit in Kliniken arbeitet und somit in die Anwendungsfelder „nicht-interventionelle klinische Forschung“ und „Aktivitäten zur Kontrolle, Verbesserung und Erforschung der Versorgung (oder bestimmter Versorgungssituationen)“ fallen kann (Xu et al. 2010; Yeniterzi et al. 2010). Die genannten Anwendungsfelder von SeConts inklusive möglicher Studientypen sind in Tab. 1 zusammengefasst.

Was hingegen nicht zum Umfang unseres Begriffs von SeConts gezählt werden kann, ist Forschung, welche klinische Daten mit Forschungsdaten kombiniert oder Daten aus anderen Quellen (z. B. von Wearables oder Fitness Apps) auswertet. Ein Beispiel hierfür wären Genomweite Assoziationsstudien (GWAS), welche aktuell häufig eine Kombination aus klinischen Daten (Phänotypen) und Forschungsdaten (Genotypen) darstellen<sup>7</sup>. Ebenso kann Biobankforschung nicht unter SeConts gefasst werden. Zwar ist es möglich, dass Biomaterial im Zuge der Behandlung und somit ohne zusätzliche Eingriffe entnommen werden kann (Kaye et al. 2012). Jedoch sind Entnahme und Sammlung des

---

<sup>5</sup>Machbarkeitsstudien dienen dazu, sich einen ersten Überblick über die Datenlage bzw. über die klinische Population zu verschaffen, welche es zu erforschen gilt. So können klinische Daten z. B. dahingehend untersucht werden herauszufinden, ob die in der Studie zu untersuchende Patientengruppe am Standort realistisch rekrutierbar ist oder ob eine Fragestellung überhaupt untersuchbar wäre.

<sup>6</sup>Klinische Daten können anhand bestimmter, für eine geplante Studie relevanter, Ein- und Ausschlusskriterien durchsucht werden um geeignete, potenzielle Studienteilnehmer zu identifizieren die im nächsten Schritt, sollte ein Re-Kontakt rechtlich erlaubt sein, für eine mögliche Studienteilnahme kontaktiert werden können.

<sup>7</sup>Aktuell sind Ganzgenomsequenzierungsdaten kein Teil der Standardbehandlung und werden vorwiegend in Forschungsstudien erhoben. Dies könnte sich möglicherweise in Zukunft ändern, was Ganzgenomsequenzierungsdaten zu klinischen Daten machen würde.

Materials ein ethisch und rechtlich relevanter Extraschritt, der nicht zu Behandlungs-, sondern zu Forschungszwecken unternommen wird. Dies gilt besonders auch für die eventuell aus dem Biomaterial zu erzeugenden Daten, mit denen später Forschung betrieben wird, da diese Daten eben nicht bereits aus dem Versorgungskontext vorliegen, sondern aus dem Biomaterial nur zu Forschungszwecken gewonnen werden.<sup>8</sup>

---

### 3 Nutzenpotenziale der Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten

Darüber, dass SeConts große Nutzenpotenziale beinhaltet, sind sich zahlreiche Autoren einig. Diese Nutzenpotenziale liegen in sehr unterschiedlichen Aspekten begründet (siehe hierzu auch den Beitrag von Anja Köngeter in diesem Band). Aus **forschungsökonomischer Sicht** ist die Sekundärnutzung klinischer Daten insofern äußerst interessant, als sie Studien ermöglicht, welche ohne bzw. mit nur minimalem Erhebungsaufwand durchführbar sind (Meystre et al. 2017). Im Rahmen dieser Studien können – durch eine mögliche Aggregation von Daten aus unterschiedlichen Kliniken – große Stichproben untersucht werden. Ebenso ist es möglich, Patienten, z. B. bei chronischen Erkrankungen, über einen langen Zeitraum hinweg anhand ihrer in der Versorgung erzeugten Daten zu betrachten und somit große Langzeitstudien (z. B. die Erforschung von Langzeiteffekten bestimmter Interventionen) durchzuführen (Martin-Sanchez et al. 2017).

Aus **forschungsmethodischer Sicht** zeichnet sich SeConts durch ein hohes Maß an externer Validität aus. Daten aus traditionellen Studiendesigns werden häufig in einer mehr oder weniger artifiziellen Umgebung erzeugt, was sinnvoll ist, um äußere Störfaktoren zu kontrollieren, jedoch auch gewisse Probleme mit sich bringen kann. So spricht man in diesem Zusammenhang vom sogenannten Hawthorne Effekt. Dieser beschreibt das Phänomen, dass Probanden sich aufgrund des Wissens, dass sie sich in einer Studie befinden, anders verhalten, als sie dies normalerweise tun würden (Adair 1984; Benedetti et al. 2016). Ein derartiger Effekt ist bei SeConts nicht zu erwarten, da die (retrospektiv genutzten) Daten einer „natürlichen“ Umgebung entspringen und

---

<sup>8</sup>Uns ist bewusst, dass der Ausschluss von Biobanken aus unserem Begriff von SeConts auch anders gesehen werden kann. Allerdings wird nach aktuellem Stand in Deutschland Biomaterial im Rahmen der Versorgung höchstens zum Zwecke der weiteren Befunderhebung für wenige Tage aufbewahrt. Die langfristige Aufbewahrung in Biobanken dient der Forschung, nicht der Versorgung. Zudem liegen im Falle von Biomaterial keine Daten vor, sondern diese müssen erst (im Rahmen der Forschungsnutzung) erzeugt werden. Sowohl die Tatsache der Aufbewahrung, die nur zu Forschungszwecken geschieht, als auch der Extraschritt der Datenerzeugung, welcher ebenfalls zu Forschungszwecken geschieht, sind Aspekte, die von unserem Begriff von SeConts explizit ausgeschlossen werden.



somit die Patienten zwar eventuell unvollständig (hierzu mehr im nächsten Abschnitt), aber dennoch in einer alltäglichen (Klinik-)Wirklichkeit abbilden (Martin-Sanchez et al. 2017). Ebenso unterliegt SeConts dadurch, dass die Daten direkt aus der Versorgung stammen, einer geringeren Wahrscheinlichkeit, bestimmten Formen von Verzerrungen (Bias) zu unterliegen (Lucero et al. 2015; Brakewood und Poldrack 2013), z. B. derart, dass im Fall von klinischen Studien bestimmte Gruppen über- bzw. unterrepräsentiert sind. So sind z. B. Frauen in Studien zu nicht-geschlechtsspezifischen Krebsformen häufig unterrepräsentiert (Jagsi et al. 2009). Ähnliches gilt für Ältere und ethnische Minderheiten in anderen Bereichen der klinischen Forschung (Sardar et al. 2014; Vitale et al. 2017). Eine Über- oder Unterrepräsentation kann sehr unterschiedliche Gründe haben, wie z. B. eine unterschiedliche Bereitschaft zur Studienteilnahme in bestimmten Gruppen von Patienten, aber auch die Verwendung von Ein- und Ausschlusskriterien, welche bestimmte Patientengruppen aus methodischen Gründen ausschließt. Dass SeConts allerdings nicht nur derartige Verzerrungen vermeidet, sondern andererseits auch anfällig für andere Arten von Verzerrungen sein kann, werden wir im nächsten Abschnitt zeigen.

Aus **forschungsethischer Sicht** besteht ein grundsätzlicher Vorteil, wie bereits angedeutet, darin, dass SeConts aufgrund ihrer Nicht-Interventionalität keinerlei direkte physische Risiken für Probanden mit sich bringt. Dies ist auch deshalb relevant, da die Nicht-Interventionalität die Möglichkeit eröffnet, Fragestellungen nachzugehen, deren Erforschung durch gezielte Interventionen in interventionellen Studien ethisch problematisch wäre (Lee 2017). Ein Beispiel hierfür ist die Erforschung bestimmter medizinischer Interventionen bei vulnerablen Gruppen wie z. B. Kindern oder Schwangeren. Eine Forschungslücke und damit ein Wissensdefizit in Bezug auf vulnerable Gruppen äußert sich in einem Fehlen von Behandlungsmethoden und Medikamenten, die speziell für diese Gruppen entwickelt und getestet wurden. Wissen aus bereits erfolgter Behandlung, vermittelt über die SeConts, könnte hier helfen, die Evidenzlage und somit auch die zukünftige Behandlung dieser Gruppen zu verbessern.

Auf die gleiche Weise können systematische Analysen von Heilversuchen ermöglicht werden. Bei Heilversuchen handelt es sich um ärztliche Heilbehandlungen, die vom medizinischen Standard abweichen, etwa mit einem noch nicht etablierten Medikament oder einer noch nicht etablierten Methode, z. B. bei Off-Label-Use von Medikamenten. Im Rahmen eines individuellen Heilversuchs bei Patienten, für die bereits alle therapeutischen Möglichkeiten ausgeschöpft sind, kann die Durchführung bestimmter Interventionen als „letzte Möglichkeit“ gerechtfertigt sein. Eine kumulierte retrospektive Analyse derartiger Heilversuche anhand der erzeugten klinischen Daten hat das Potenzial, Hypothesen zu generieren, welche die Entwicklung einer zukünftigen Therapie anstoßen.

Ein weiterer Bereich, in dem SeConts neue Möglichkeiten für die biomedizinische Forschung eröffnet, ist derjenige der **seltenen Erkrankungen**. Die Erforschung seltener Erkrankungen ist aufgrund der per Definition geringen Fallzahlen zum einen nur sehr schwierig durchführbar, zum anderen für nicht universitäre Forschungseinrichtungen

**Tab. 2.** Nutzenpotenziale von SeConts

Art des Nutzenpotenzials	Quelle des Nutzenpotenzials
forschungsökonomisch	große Stichproben, hohe zeitliche Abdeckung bei geringem Ressourceneinsatz
forschungsmethodisch	hohes Maß an externer Validität/ Repräsentativität
forschungsethisch	Möglichkeit der Erforschung von Fragen, die aus ethischen Gründen mit interventionellen Studien nicht erforscht werden dürfen (Beforschung vulnerabler Gruppen, Auswertung experimenteller Therapien)
auf einzelne Erkrankungen bezogen	Ermöglichung von Studien an seltenen Erkrankungen

aus dem Bereich der Industrie aufgrund der zu erwartenden geringen Absätze äußerst unattraktiv. Dies kann zu einer – im Verhältnis zu anderen Erkrankungen – schlechteren Versorgung der Betroffenen führen. Die Möglichkeit der Kumulation der Behandlungsdaten von Patienten aus unterschiedlichen Krankenhäusern bringt die Hoffnung mit sich, diese Forschungslücken schneller zu schließen und die Entwicklung von Behandlungsmethoden voranzutreiben, da hier schneller hohe Fallzahlen mit relativ geringem organisatorischen und finanziellen Aufwand erreicht werden können (Elger et al. 2010). Tab. 2 fasst die genannten Nutzenpotenziale von SeConts zusammen.

#### 4 Herausforderungen für die Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten

Neben den genannten Nutzenpotenzialen sind Herausforderungen von bzw. für SeConts zu nennen (siehe hierzu auch den Beitrag von Anja Köngeter in diesem Band). Diese teilen wir in Probleme mit Datenqualität und -vollständigkeit, Interoperationalisierbarkeit und Verzerrung (Bias) ein.

**Datenqualität und Vollständigkeit:** Im Gegensatz zu Forschungsdaten werden klinische Daten per Definition mit dem primären Zweck der Patientenversorgung erhoben. Die Tatsache, dass Daten einen primären Zweck erfüllen, führt dazu, dass sie nicht optimal auf den sekundären Nutzen angepasst sind. So sind gewisse Informationen, welche im Forschungskontext wichtig sind, für den eigentlichen Erhebungszweck – den klinischen Alltag – mitunter weniger relevant und werden deshalb nur lückenhaft erhoben (Sturmer et al. 2011; Hersh et al. 2013). Derartige Lücken bestehen mitunter auch in Bezug auf Informationen zu etwaigen konfundierenden Faktoren, Outcome Werte o.ä., welche im klinischen Alltag oft nicht in der gleichen Ausführlichkeit kontrolliert und somit dokumentiert werden können wie es für die Forschung notwendig wäre (Meystre et al. 2017). Doch auch wenn ein Datensatz alle (für Forschung wesentlichen) Daten beinhaltet, birgt sein Ursprung im klinischen Alltag gewisse Schwierig-

keiten hinsichtlich seiner Verwendung für die Forschung. So liegen klinische Daten aktuell noch immer häufig in unstrukturierter Weise vor, d. h. in natürlicher Sprache ohne eine systematische Codierung (z. B. Arztbriefe). Dieser Umstand bewirkt, dass eine effiziente Auswertung nur über den Umweg des sogenannten Natural Language Processing möglich ist (Murff et al. 2011; Xu et al. 2010). Wenngleich diese Technologie in den letzten Jahren große Fortschritte gemacht hat, so ist sie immer noch in der Entwicklung und potenziell fehleranfällig (Sheikhalishahi et al. 2019).

Eine weitere große Herausforderung für SeConts sind unterschiedliche Standards der Dateneingabe und -codierung oder gar unterschiedliche Erhebungsinstrumente in unterschiedlichen Kliniken und Arztpraxen (Meystre et al. 2017), welche die **Interoperationalisierbarkeit** von Daten aus mehreren Ursprungsorten erschwert (Geissbuhler et al. 2013; Ancker et al. 2011).<sup>9</sup> So existieren zahlreiche unterschiedliche klinische Informationssysteme (KIS) und klinische Routinen der Dokumentation, deren Daten vereinheitlicht werden müssen, sollen sie in einer Studie zusammengeführt werden. Die ex post Herstellung der Interoperationalisierbarkeit von Daten aus unterschiedlichen Quellen führt notwendigerweise zu gewissen Aufbereitungskosten und kann ihrerseits wiederum Fehler produzieren (Lee 2017). Zusätzlich zu den technischen Faktoren der Interoperationalisierbarkeit der Daten ist anzunehmen, dass ein potenzieller Unterschied in der Erfahrung der Diagnostiker und Behandler bei der Diagnosestellung und Dokumentation (Hersh et al. 2013) sich in geringer und vor allem nicht kontrollierbarer Inter-Rater-Reliabilität (d. h. der Zuverlässigkeit der Vergleichbarkeit von Informationen die von unterschiedlichen Personen dokumentiert werden) und somit in Schwierigkeiten der Vergleichbarkeit der aus der Diagnostik gewonnenen Daten niederschlägt.

Die bisher genannten Schwierigkeiten bei SeConts bzgl. Datenqualität, -vollständigkeit und Interoperationalisierbarkeit zeigen sich exemplarisch in einer Studie von Botsis et al. (Botsis et al. 2010). Hier wurde mit Hilfe elektronischer Krankenakten die Überlebenszeit von Patienten mit einem Pankreaskarzinom ausgewertet. Es zeigte sich, dass nur 33 % der vorhandenen Datensätze für die finale Analyse verwendet werden konnten, da in 67 % der Datensätze an relevanten Stellen Informationen fehlten. Vorhandene Informationen wurden zudem oft als ungenau eingestuft und mussten aufwändig aus unterschiedlichen Quellen zusammengestellt werden. Dies führte mitunter zu Inkonsistenzen in den Daten, die z. B. daraus resultierten, dass Patienten in unterschiedlichen Datenquellen unterschiedliche Diagnosen zugewiesen wurden.

Mit der Gefahr von **Verzerrungen (Bias)** hängt eine weitere Schwierigkeit der Sekundärnutzung ebenfalls mit der unsystematischen bzw. nicht an einer bestimmten Forschungslogik orientierten Art der Erhebung zusammen (Terris et al. 2007; Rusanov

---

<sup>9</sup>Im Rahmen der Medizininformatikinitiative wird aktuell an der Erstellung von Standards gearbeitet, welche eine Interoperationalisierbarkeit ermöglichen soll (vgl. hierzu auch den Beitrag von Wiebke Lesch in diesem Band).

et al. 2014; Prada-Ramallal et al. 2019; Jones et al. 2017; Lee 2017). Mögliche Verzerrungen können verschiedene Ursachen haben. Es zeigt sich z. B., dass die klinische Dokumentation von Patienten mit schwerem Krankheitsverlauf potenziell ausführlicher vorliegt als diejenige von Patienten mit milden Verläufen. Dies führt bei einer Sekundärnutzung in der Forschung zu einer Verzerrung, die sich dahingehend äußert, dass – um die Vollständigkeit der verwendeten Datensätze zu gewährleisten – Patienten mit schwereren Verläufen überproportional häufig in entsprechenden Studien vertreten sind (Rusanov et al. 2014).

Außerdem können nicht erhobene konfundierende Faktoren Verzerrungen erzeugen. Zu konfundierenden Faktoren können z. B. spezielle Rahmenbedingungen in einer bestimmten Klinik gehören, welche zu einer Überrepräsentation bestimmter Patienten führen können. So kann eine bestimmte Station z. B. einen besonders guten Ruf haben, was dazu führen kann, dass sich dort besonders viele Patienten mit schweren Verläufen vorstellen, in der Hoffnung, ihnen würde dort besonders gut geholfen. Diese Information ist nicht notwendigerweise in den klinischen Daten sichtbar, kann aber eine Verzerrung in den Daten mit sich bringen.

Zusätzlich besteht die Gefahr des sogenannten „Upcodens“ (Meystre et al. 2017), d. h. der ungerechtfertigten oder zumindest „großzügigen“ Diagnosestellung. Im Rahmen eines in vielen Ländern gängigen Abrechnungssystems für Behandlungsleistungen werden diagnosebezogene Fallgruppen (diagnosis related groups (DRG)) als Grundlage für die Berechnung der von den Krankenkassen erstatteten Mittel verwendet. Jeder Behandlungsfall wird entsprechend einer Pauschale vergütet, die anhand der Diagnose definiert ist. „Upcoden“ wird in einem DRG Abrechnungssystem potenziell<sup>10</sup> aus ökonomischer Motivation durchgeführt um eigentlich nicht indizierte, aber profitablere Behandlungen durchführen und abrechnen können (Steinbusch et al. 2007). Derartiges „Upcoden“ kann die Validität der Daten mitunter stark verzerren.

Ein Teil der genannten Einschränkungen und Herausforderungen für die Sekundärnutzung klinischer Daten für SeConts sind technischer und organisatorischer Natur und werden aller Voraussicht nach in den nächsten Jahren bzw. Jahrzehnten gelöst werden. Mit Investitionen in die digitale medizinische Infrastruktur und nationalen wie internationalen Harmonisierungsanstrengungen können Probleme der Interoperationalisierbarkeit der Daten, der Unstrukturiertheit bzw. der Umgang mit derselben und der Unvollständigkeit der Datensätze zumindest prinzipiell gelöst werden. Andere Einschränkungen hingegen, wie z. B. die Frage der Inter-Rater-Reliabilität oder auch verschiedene Formen einer möglichen Verzerrung der Daten sind Probleme, die bei der Bewertung und Durchführung stets beachtet werden müssen, um das Potenzial

---

<sup>10</sup>Upcoding muss nicht notwendigerweise aus Gründen der persönlichen Bereicherung oder zum finanziellen Vorteil der Klinik geschehen. Es ist auch denkbar, dass Behandler ihren Patienten eine Therapie ermöglichen wollen, die sie für sinnvoll erachten, obwohl die dafür abrechnungsrechtlich notwendigen Voraussetzungen nicht vollständig erfüllt sind.

**Tab. 3.** Herausforderungen für SeConts

Art der Herausforderungen	Ursache
Datenqualität und Vollständigkeit	<ul style="list-style-type: none"> <li>• unvollständige Datensätze</li> <li>• Unstrukturiertheit der Daten</li> </ul>
Interoperationalisierbarkeit	<ul style="list-style-type: none"> <li>• unterschiedliche Standards der Dateneingabe und -codierung</li> <li>• nicht kontrollierbare Inter-Rater-Reliabilität</li> </ul>
Verzerrung (Bias)	<ul style="list-style-type: none"> <li>• unterschiedliche Dokumentations- und Erhebungsstandards</li> <li>• unbekannte konfundierende Faktoren</li> <li>• „Upcoding“</li> <li>• ...</li> </ul>

von SeConts bestmöglich ausschöpfen zu können. Tab. 3 fasst die genannten Herausforderungen zusammen.

## 5 Mögliche Risiken der Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten für Patienten und weitere Stakeholder

### 5.1 Die Gefahr von Datenpannen

Nachdem im vergangenen Abschnitt verschiedene technische, organisatorische und datenimmanente Herausforderungen von SeConts benannt wurden, sollen im Folgenden mögliche Risiken analysiert werden, welche SeConts für die beteiligten Stakeholder, insbesondere die Patienten, mit sich bringen kann. Die Durchführung von SeConts ist in verschiedenen Dimensionen auf einer breiten Skala vorstellbar: (1) punktuell mit (spezi-fischer) informierter Einwilligung, d. h. im Rahmen einzelner Studien, welche Patienten um die Einwilligung zur Datennutzung bitten; (2) semi-systematisch mit breiter Einwilligung, d. h. Patienten willigen in zukünftige SeConts ein, die zum Zeitpunkt der Einwilligung noch nicht feststeht; (3) systematisch ohne Einwilligung, d. h. Daten werden standardmäßig und ohne vorige Einwilligung der Forschungsnutzung (SeConts) zugeführt, so lange Patienten nicht widersprechen (Opt-Out). Zusätzlich sind weitere Einwilligungsmodelle denkbar. Während unser Text sich auf Nutzenpotenziale, Herausforderungen und Risiken für verschiedene Stakeholder bezieht, die grundsätzlich für alle Dimensionen der möglichen Durchführung von SeConts relevant sind, geht es nicht um die verschiedenen denkbaren Einwilligungsmodelle, sodass wir auch keine ethische Bewertung derselben durchführen.

Bei SeConts handelt es sich um einen Vorgang, bei dem besonders sensible Daten betroffen sind. Zwar sind personenbezogene Daten stets als schützenswert einzustufen.

Nicht umsonst betont das Bundesverfassungsgericht, dass im Kontext der elektronischen Datenverarbeitung kein Datum belanglos ist (BVerfG, Urt. v. 15.12.1983 – 1 BvR 209 u.a./83, BVerfGE 65, 1). Jedoch zeichnen sich klinische Daten durch ein besonders hohes Maß an Sensibilität aus (Rumbold und Pierscionek 2018), handelt es sich doch um Informationen über Patienten, welche besonders persönlicher Natur sind und potenziell Rückschlüsse auf viele Bereiche des Lebens (z. B. Lebensstil, Sexualverhalten, erwarteter zukünftiger Gesundheitsstatus etc.) des Einzelnen zulassen. Nicht zuletzt zählen Informationen dieser Art zu den „besonderen Kategorien von personenbezogenen Daten“ gemäß Art. 9 Abs. 1 DSGVO und werden zusätzlich von der ärztlichen Schweigepflicht geschützt.

Speicherung, Transfer und Verarbeitung von Daten birgt stets die Gefahr von Datenpannen. Mit dem Begriff der Datenpanne sind hier alle Arten von Ereignissen gemeint, bei denen die Vertraulichkeit und der geplante Schutz der Daten verletzt werden, sei es durch technisches Versagen, menschliche Unachtsamkeit oder durch vorsätzliche unethische oder kriminelle Handlungen. Das Spektrum möglicher Folgen derartiger Datenpannen mit klinischen Daten, z. B. nach einem Datendiebstahl, ist weit und könnte von Belästigung durch personalisierte Werbung bis hin zu Identitätsdiebstahl, Diskriminierung, Stigmatisierung oder Erpressung reichen (Benitez und Malin 2010; Kaplan 2016; Parker und Aggleton 2003; Laurie et al. 2014; Weichert 2018; Bundesamt für Gesundheit 2017).

Auch wenn uns im deutschsprachigen Raum keine systematischen Untersuchungen zu Datenpannen in Form konkreter Datendiebstähle in Krankenhäusern bekannt sind<sup>11</sup>, so zeigen Untersuchungen z. B. aus den USA, dass dort bereits Daten aus elektronischen Patientenakten entwendet wurden. Liu und Kollegen geben an, dass zwischen 2010 und 2013 in amerikanischen Krankenhäusern insgesamt 949 unbefugte Zugriffe auf Informationen aus insgesamt fast 30 Mio. Patientenakten dokumentiert wurden. 552 der Datenpannen wurden explizit als Diebstahl klassifiziert (Liu et al. 2015). Eine Studie von Floyd et al. zeigt anhand von belegten Datenpannen (inklusive Datendiebstählen) in US-Krankenhäusern zwischen 2012 und 2015, dass die Mehrheit der betreffenden Daten aus Namen, Geburtsdatum, Sozialversicherungsnummer und ähnlichem besteht. Es wurden jedoch auch Daten wie Diagnosen und Behandlungsinformationen entwendet (Floyd et al. 2016). Die jeweilige Art des Missbrauchs der betroffenen Daten in Folge eines Datendiebstahls ist schwer zu erforschen. Eine genaue Nachverfolgung der missbräuchlichen Verwendungen und damit verbundener Folgen für die Betroffenen ist grundsätzlich schwierig. Mit Blick auf die Existenz zahlreicher (illegaler) Portale im sogenannten Darkweb, auf denen mit persönlichen Daten gehandelt wird, liegt die Vermutung nahe, dass auch gestohlene klinische Daten dort angeboten werden (Floyd et al. 2016).

---

<sup>11</sup>In den letzten Jahren wurden allerdings auch im deutschsprachigen Raum Fälle von Datenpannen inklusive Datendiebstahl bekannt (Bahners 2018; Ärzteblatt 2020; Deutschlandfunk 2018).

## 5.2 Risiken für Patienten

Im Folgenden widmen wir uns zunächst den möglichen **Risiken**, die in erster Linie die **Patienten** selbst betreffen aber potenziell auch Auswirkungen für andere Stakeholder haben (zum Thema „Risiken“, siehe auch den Beitrag von Anja Köngeter in diesem Band). Hierbei legen wir den Schwerpunkt auf diejenigen Risiken, die direkt oder indirekt mit der Sicherheit der Daten in Verbindung stehen, zeigen aber auch Risiken auf, die im Rahmen von SeConts sogar ohne Datenpannen entstehen können. Dabei ist zu bedenken, dass es, wie es auf dem Gebiet der Beschäftigung mit neuartigen Entwicklungen oft vorkommt, keine einschlägigen empirisch gesicherten Kenntnisse gibt. Für eine bessere Einschätzung möglicher Risiken in Bezug auf den Missbrauch gestohlener Daten ist eine Verbesserung der Forschungslage erforderlich. Eine Pflicht für öffentliche Einrichtungen wie Krankenhäuser, Datenschutzpannen zu melden (vgl. Art. 33 DSGVO) stellt hierfür eine geeignete Grundlage dar. Die wissenschaftliche Aufbereitung derartiger Informationen ist im deutschen Kontext derzeit noch äußerst begrenzt. Die Wahrscheinlichkeit einer Datenpanne oder gar eines Datendiebstahls im Kontext von SeConts ist zudem äußerst schwierig abzuschätzen, da empirische Daten aktuell wenn, dann lediglich für die Krankenversorgung, nicht aber für SeConts vorliegen (siehe Abschn. 5.1). Gleiches gilt in verstärktem Maße für mögliche Arten des Missbrauchs, welche aus einer Datenpanne folgen können. Aus diesem Grund untersuchen wir mögliche Risiken aus Sicht eines „Negativszenarios“, dessen Eintrittswahrscheinlichkeit im Moment nur schwer einzuschätzen ist. Voraussetzung dafür, dass Daten missbraucht werden und dem Einzelnen schaden können, ist die Möglichkeit, Personen mit den, in den Daten enthaltenen, persönlichen Informationen zu verknüpfen (Re-Identifikation). Zwar ist es in der biomedizinischen Forschung üblich, mit pseudonymisierten (d. h. de-identifizierten) Daten zu arbeiten,<sup>12</sup> doch macht dies eine Re-Identifikation nicht unmöglich (zur datenschutzrechtlichen Relevanz pseudonymisierter Daten, vgl. den Beitrag von Markus Spitz in diesem Band). Möglichkeit und Aufwand einer solchen Re-Identifikation schwankt je nach Inhalt, Struktur sowie dem jeweiligem Verwendungskontext des Datensatzes (Jungkunz et al. 2021). Doch drohen mögliche Risiken für Patienten im Zuge von Datenpannen nicht allein im konkreten Fall einer Datenpanne mit gezieltem Datenmissbrauch, welcher faktisch spürbare negative Auswirkungen auf die betroffenen Personen hat. Bereits die Kenntnis einer Datenpanne kann für die einzelne betroffene Person aufgrund der Ungewissheit über das Schicksal ihrer persönlichen Daten und darüber, ob diese irgendwo und irgendwann

---

<sup>12</sup>In nur sehr wenigen Fällen ist eine direkte Identifikation einzelner Patienten notwendig (z. B. in Longitudinalstudien oder aber beim Zusammenführen verschiedener Datensätze von verschiedenen Standorten. Selbst in diesen Fällen ist es jedoch in der Regel möglich, mit pseudonymisierten Daten zu arbeiten, z. B. mit der Hilfe von Datentreuhändern o. ä.).

gegen sie verwendet werden, potenziell erheblich belastend sein (de Bruin 2010; Laurie et al. 2014).

Ausgehend von den genannten möglichen Risiken (Identitätsdiebstahl, Diskriminierung, Stigmatisierung oder Erpressung), welche für Patienten von einer möglichen Datenpanne ausgehen, lassen sich im Rahmen eines „Negativszenarios“ weitere mögliche unerwünschte Folgen vorstellen. Im Falle einer systematischen Durchführung von SeConts ohne Einwilligung (siehe Abschn. 5.1) könnte z. B. die Gefahr bestehen, dass Patienten, die sich Gedanken über mögliche Datenschutzrisiken machen, Vertrauen in das Gesundheitssystem verlieren (Platt und Kardia 2015), was ein Risiko darstellt, welches nicht allein auf Patienten beschränkt ist. Der Vertrauensverlust in das Gesundheitssystem könnte sich in einem Vertrauensverlust gegenüber dem Arzt niederschlagen und somit das Arzt-Patienten-Verhältnis negativ beeinflussen (Kelley et al. 2015). Ein Vertrauensverlust könnte auf der Sorge vor Einbußen der Vertraulichkeit der Daten beruhen (zum Thema Angst um die Vertraulichkeit, vgl. (King et al. 2012)). Studien zeigen, dass Patienten aufgrund mangelnden Vertrauens in die Sicherheit bei der Speicherung und Weitergabe ihrer Behandlungsdaten Informationen vorenthalten haben (Agaku et al. 2014; Campos-Castillo und Anthony 2015). Eine systematische Durchführung von SeConts ohne Einwilligung könnte diesen Effekt verstärken, so dass Patienten aus Angst um die Vertraulichkeit ihrer Daten im schlimmsten Fall behandlungsrelevante Informationen für sich behalten könnten (Kaplan 2016). Auch ist denkbar, dass Patienten aus Angst um die Vertraulichkeit ihrer Daten notwendige Arztbesuche nicht wahrnehmen.

Doch auch eine Durchführung von SeConts mit vorheriger Einwilligung könnte bei Patienten die Sorge um die Sicherheit der Daten und somit Belastungen für Patienten mit sich bringen, so wie es in jeder anderen Form der Datenverarbeitung der Fall sein kann. Im Falle von SeConts ist es zudem möglich, dass die Einwilligung des Patienten die Befürchtungen um die Sicherheit der Daten nicht adäquat abbildet, wenn eben jene Einwilligung durch soziale Erwünschtheit oder auch falsche Hoffnungen in Bezug auf den eigenen Nutzen von SeConts motiviert ist (Aitken et al. 2016) und deshalb eigene Befürchtungen hinten angestellt werden (vgl. hierzu die Debatte zum Begriff „therapeutic misconception“).

So ist es vorstellbar, dass Patienten hoffen, durch ihre Einwilligung in SeConts das Wohlwollen ihres Arztes und somit einen Vorteil für ihre Behandlung zu erlangen. Ebenso ist es möglich, dass Patienten sich von SeConts (bzw. deren Ergebnissen) einen persönlichen Vorteil erhoffen (Richter et al. 2018), der sie motiviert, der Nutzung zuzustimmen und dabei andere (entgegenstehende) persönliche Interessen und Werte nachzuordnen. SeConts ist dem Verständnis im vorliegenden Artikel zufolge nicht darauf ausgelegt, dem Patienten, der seine Daten bereitstellt, selbst zu nutzen. Es ist zwar nicht gänzlich auszuschließen, dass SeConts einen individuellen Nutzen für einzelne Patienten mit sich bringt, jedoch ist dies aufgrund der notwendigen Zeit, welche selbst die Umsetzung oder Implementierung von Kenntnissen aus *Aktivitäten zur Kontrolle und Erforschung der Versorgung (oder bestimmter Versorgungssituationen)* benötigt



(von der Translation von Forschungserkenntnissen in die Behandlung ganz zu schweigen), sehr unwahrscheinlich. So könnte SeConts bei (jungen) Patienten mit chronischen Erkrankungen mittelfristig eine Verbesserung der Behandlung hervorbringen, von denen sie auch selbst in Zukunft profitieren könnten. Aus unserer jetzigen Perspektive gehen wir jedoch davon aus, dass bei SeConts in aller Regel kein individueller Eigennutzen erwartet werden kann.

Es ist theoretisch möglich, dass Patienten durch die Rückmeldung sogenannter Zufalls- oder Zusatzbefunde von SeConts profitieren. Hierbei muss aber unterstrichen werden, dass es zu möglichen Zufallsbefunden bei SeConts, nach unserem Kenntnisstand, keine Erfahrungen und Berichte gibt. Es herrscht in der Bioethik eine lebhafte Diskussion um Zufallsbefunde aus der Forschung, d. h. aus der Analyse von Forschungsdaten, insbesondere bei Bildgebung (Erdmann 2015) und Genomik (Gitter 2019). Selbst bei genomischer Forschung ist jedoch unklar, ob Zufallsbefunde in relevanter Häufigkeit auftreten (Schuol et al. 2015). Im Gegensatz zum reinen Forschungskontext, in welchem Daten zu Forschungszwecken analysiert werden und nicht schon von einem (behandelnden) Arzt in Augenschein genommen wurden, stammen klinische Daten per Definition bereits aus klinischen Diagnosen und Therapien, für die sie erhoben wurden. Sie wurden also bereits von einem Behandler oder Diagnostiker zum Zwecke der Behandlung untersucht. Ein Zufallsbefund im Rahmen von SeConts durch Forscher ist aus unserer Sicht daher relativ unwahrscheinlich, so dass, zusammen mit dem völligen Mangel an Erfahrung, zum jetzigen Zeitpunkt wenig dafür spricht, in möglichen Zufallsbefunden einen ernstzunehmenden Grund für potenziellen persönlichen Nutzen für den Patienten zu sehen.

Etwaige Hoffnungen von Patienten, sie würden von der Zustimmung zu SeConts selbst profitieren, sind in der Regel unbegründet. Der Gefahr der falschen Hoffnungen auf einen persönlichen Nutzen ist also entgegenzuwirken, z. B. durch eine ausführliche Aufklärung von Patienten über Zweck, Möglichkeiten und Konsequenzen von SeConts.

### 5.3 Risiken für weitere Stakeholder

Neben den Risiken, die mehr oder weniger ausschließlich die Patienten betreffen, kann SeConts auch **Risiken** und Herausforderungen für Patienten und **weitere Stakeholder** mit sich bringen. Zwei derartige mögliche Risiken sollen im Folgenden angeführt werden: erstens die mögliche Belastung durch erhöhten Dokumentationsaufwand für Ärzte und Pflegepersonal und zweitens der Vergleich (von Qualität und Wirtschaftlichkeit) medizinischer Maßnahmen zwischen verschiedenen Behandlern oder Kliniken.

Werden Daten aus der Versorgung standardmäßig und systematisch der Forschung zugeführt, so verändert dies mit hoher Wahrscheinlichkeit die Dokumentationsroutinen der Versorgung selbst. Daten müssen in einer Form dokumentiert werden, welche eine spätere sekundäre Verarbeitung im Rahmen von SeConts ermöglicht. Ebenso könnte es zu der Entwicklung kommen, dass, vor dem Hintergrund der späteren Forschungs-

nutzung, eine größere Menge an Daten erhoben wird, als dies für die reine Versorgung notwendig wäre. Eine derartige Erhöhung der dokumentierten Datenmenge kann sowohl für das Personal selbst eine Belastung darstellen, als auch negative Auswirkungen auf die Patientenversorgung mit sich bringen. Zum einen trüfe ein derartiger Mehraufwand bei der Datenerhebung auf ohnehin bereits knappe zeitliche Ressourcen des klinischen Personals, was für das Personal selbst eine physische und psychische Belastung wäre. Zudem würde das für die Dokumentation aufgewendete Mehr an Zeit an anderer Stelle in der Versorgung fehlen und somit negative Folgen für Patienten mit sich bringen. Gegen die Entwicklung in Richtung dieses negativen Szenarios spricht die Möglichkeit eines anderen, positiveren Szenarios. So kann eine standardisierte Dokumentation mit standardisierten digitalen Formularen – z. B. durch die Nutzung von festen Textbausteinen, welche für die Erstellung von Arztbriefen ausgewählt werden können – auch eine Zeitersparnis mit sich bringen. Voraussetzung hierfür ist jedoch die Implementierung eines ausgereiften, komplexen und möglichst bundesweit kompatiblen digitalen klinischen Informationssystems (KIS) mitsamt der entsprechenden IT-Infrastruktur und gut geschultem Personal.

Das zweite zu schildernde konkrete Risiko von SeConts besteht in der Möglichkeit eines Vergleichs von Qualität und anderen Parametern (z. B. Kosten) medizinischer Maßnahmen einzelner Ärzte, Abteilungen oder ganzer Kliniken. Zwar ist der Vergleich von Qualität etc. eine durchaus wichtige Maßnahme zur Qualitätskontrolle und -verbesserung sowie der Patientensicherheit. Jedoch erzeugen mögliche Vergleiche zwischen Behandlern oder zwischen Kliniken mitunter auch Ängste, (ungerechtfertigt) in ein schlechtes Licht gerückt zu werden (z. B. bei Kollegen; Vorgesetzten; den Forschern, die die Daten analysieren). Eine derartige Befürchtung kann sich z. B. für ein onkologisches Spitzenzentrum ergeben, an das sich überproportional viele Patienten mit negativen Prognosen wenden, wenn dieser spezielle Umstand in einem Vergleich der Leistungen, z. B. der Outcomes, mit anderen onkologischen Kliniken nicht angemessen berücksichtigt wird.

Auch ist zu befürchten, dass sich ein Vergleich von Stationen oder Kliniken nicht nur auf die Qualität der Versorgung, sondern auch und primär auf Wirtschaftlichkeit (Profitabilität und Kosten) fokussieren und in dieser Sphäre einen Konkurrenzdruck um die „günstigste oder profitabelste Behandlung“ weiter befeuern könnte. In einem Gesundheitssystem, welches ohnehin bereits unter hohem Kosten- und Ökonomisierungsdruck leidet (Strech et al. 2008; Fernau et al. 2017; Krause et al. 2013), könnte dies einen weiteren Faktor zur Zuspitzung der Lage darstellen und durchaus negative Auswirkungen auf die Versorgungssituation selbst haben. Allerdings ist auch zu bemerken, dass ein Blick auf die Kosten von Behandlung im Rahmen eines eingepflegten Wirtschaftlichkeitsgebots stets erforderlich ist und z. B. ganz besonders an den Stellen ethisch geboten ist, an denen die Gefahr einer Überversorgung aufgrund finanzieller Interessen der Kliniken droht. Vergleiche zwischen Ärzten oder Abteilungen/Kliniken im Rahmen von SeConts haben also auch das Potenzial, zur Aufdeckung derartiger problematischer Missstände beizutragen, etwa im Falle eines Übermaßes an Operationen mit fragwürdiger Wirksamkeit aber finanziellen Vorteilen für die jeweilige Klinik (Quintel et al. 2016).

## 6 **Ausblick auf mögliche Maßnahmen zur Reduzierung der Risiken der Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten**

### 6.1 **Einschätzung der Höhe des Risikos für Patienten**

Ein wichtiger erster Schritt, um Risiken zu reduzieren, ist die Identifikation derselben und eine Einschätzung ihrer Höhe. Das Gesamtrisiko für den einzelnen Patienten im Rahmen einer konkreten Form von SeConts lässt sich zwar nicht in absoluten Zahlen darstellen. Jedoch lassen sich Aspekte von Daten und deren Nutzung identifizieren, die für eine **Einschätzung der Höhe des Risikos für Patienten** aufschlussreich sind (Jungkunz et al. 2021). Die Identifikation und Einschätzung der Höhe eines Risikos für Patienten ermöglicht im nächsten Schritt die Ergreifung von (technisch-organisatorischen) Gegenmaßnahmen. Datensätze, die aufgrund ihrer hohen Variablen- und geringen Fallzahl die Re-Identifikation von Individuen verhältnismäßig „leicht“ machen (Bender et al. 2001; Mokken et al. 1992; Dankar et al. 2012; El Emam et al. 2009), können z. B., soweit es die Forschungsfrage zulässt, anders strukturiert werden um das Re-Identifikationsrisiko zu senken.<sup>13</sup> Ebenso ließen sich – unter Vorbehalt der Kompatibilität mit der Forschungsfrage sowie der technischen Machbarkeit – die Analysen u. U. vor Ort durchführen (dezentrale Analysen), so dass nur die kumulierten Ergebnisse herausgegeben würden, die auf der Basis der Vorgaben der Forscher erstellt wurden. Die klinischen Daten selbst würden somit die datenhaltende Stelle nicht verlassen.

### 6.2 **Use and Access Committees**

Zusätzlich zur Überprüfung durch eine Ethikkommission – wie bei Studien mit klinischen (personenbezogenen) Daten üblich – sollten mögliche Risiken und Implikationen für Betroffene auch von einem „**Use and Access Committees (UAC)**“ bewertet werden. UACs stellen ein wichtiges Element einer angemessenen Governance Struktur für SeConts dar. Ihre Aufgabe ist es, den Zugriff auf klinische Daten für SeConts strukturiert und transparent zu regeln. Dabei haben sie sowohl die Datenschutzinteressen der Patienten als auch die legitimen Interessen und Befürchtungen anderer

---

<sup>13</sup>Eine mögliche Modifikation des Datensatzes kann z. B. in der Form geschehen, dass – gerade bei geringen Fallzahlen – die Anzahl der Variablen reduziert werden muss. Die Kombination vieler Variablen ermöglicht es potenziell, eine Re-Identifikation mit statistischen Mitteln vorzunehmen, da eben jene Kombination mit einer höheren Wahrscheinlichkeit einzigartig und somit potenziell einer konkreten Person zuzuordnen ist, je größer die Anzahl der Variablen (vgl. hierzu (Sweeney 2000, 2002)).

Stakeholder im Blick (Shabani und Borry 2016). UACs können z. B. eventuell auf die Konditionen des Data Transfer Agreements bzw. der Datennutzungsbestimmungen einwirken, die mit dem anfragenden Forscher abgeschlossen werden (zum Thema der vertraglichen Regelungen zwischen Datensender und Datenempfänger, siehe den Beitrag von Markus Spitz in diesem Band). Mögliche Risiken wie z. B. das oben angesprochene Risiko eines unsachgemäßen und damit potenziell „ungerechten“ Vergleichs verschiedener Kliniken oder Abteilungen könnte durch gezielte Bestimmungen, welche entsprechende Aktivitäten der datenanfragenden Forscher unterbinden, reduziert werden. Ebenso kann ein UAC prüfen, ob der jeweilige datenanfragende Forscher bzw. die in Frage kommende Studie vorab definierte Kriterien erfüllt und diese ggf. einfordern bzw. die Datenherausgabe im Zweifel untersagen. Zu diesen Kriterien gehören beispielsweise Vorgaben in Bezug auf wissenschaftliche Transparenz und Gemeinwohlorientierung (Cheah und Piasecki 2020). Wie die Anforderungen an Studien in Bezug auf Transparenz und Gemeinwohlorientierung konkret aussehen, ist eine Frage, die noch geklärt werden muss (vgl. hierzu den Beitrag von Anja Köngeter in diesem Band).

### **6.3 Aufklärung und Widerspruchsmöglichkeit**

Weitere Elemente für eine ethisch vertretbare Durchführung von SeConts finden sich auf der Ebene der Aufklärung, Einwilligung und der Gewährleistung einiger konkreter informationeller Selbstbestimmungsrechte wie insbesondere dem Recht auf Widerspruch (Opt-Out). Die Notwendigkeit der Widerspruchsmöglichkeit (Opt-Out) ergibt sich aus der Tatsache, dass eine Risikoabschätzung, wie wir sie oben (Abschn. 6.1) vorschlagen, stets nur eine ungefähre Abschätzung für einen Großteil der möglichen Patienten erfassen kann, nicht aber individuelle Variationen des Risikos. So erscheinen z. B. bestimmte Datentypen für viele Personen als relativ unbedenklich in Bezug auf ihre Sensibilität. Für einzelne Personen kann diese Einschätzung aber durchaus falsch sein. Ein Profi-Fußballspieler sieht sich bei der Weitergabe der Röntgenaufnahme einer Meniskusluxation im Kniegelenk mit Blick auf den Abschluss möglicher neuer lukrativer Arbeitsverträge anderen Risiken gegenüber als dies für die meisten anderen Personen (bei gleicher Wahrscheinlichkeit einer Re-Identifikation) der Fall ist.

Derartig spezielle Einzelfälle, welche in einer allgemeinen Risikoabschätzung nicht berücksichtigt werden (können), dürfen keinem individuell nicht tragbaren Risiko ausgesetzt werden. Eine Voraussetzung, um diesem individuellen Risiko für einzelne Patienten im Rahmen einer Widerspruchsmöglichkeit Rechnung zu tragen, ist eine verständliche Patientenaufklärung. Eine adäquate Aufklärung darüber, was mit seinen klinischen Daten geschehen soll, ermöglicht jedem Patienten, eine individuelle Risikoeinschätzung vorzunehmen und eine Entscheidung darüber zu treffen, ob er dieses Risiko tragen will oder nicht. Gleichzeitig dient die Aufklärung der Schaffung von Transparenz in Bezug auf SeConts und wirkt so falschen Hoffnungen bzgl. des Nutzens entgegen. Die Achtung der Patientenautonomie durch die Widerspruchsmöglichkeit kann in Ver-

**Tab. 4.** Mögliche Risiken von SeConts für Patienten und weitere Stakeholder

Risiko betrifft	Art des Risikos	Mögliche Gegenmaßnahme
Patient	Re-identifizierung (Verlust der informationellen Selbstbestimmung)	<ul style="list-style-type: none"> <li>• Einschätzung der Höhe des Risikos für Patienten gefolgt von Risikominimierungsstrategien (technisch-organisatorische Maßnahmen)</li> <li>• Aufklärung und Widerspruchsmöglichkeit</li> <li>• Use and Access Committees</li> <li>• Data Transfer Agreements/ Datennutzungsbestimmungen</li> </ul>
	Datenmissbrauch: Identitätsdiebstahl, Diskriminierung, Stigmatisierung oder Erpressung	
	Vertrauensverlust im Arzt-Patientenverhältnis: Unterlassung notwendiger Arztbesuche, Zurückhalten Behandlungsrelevanter Informationen	
	falsche Hoffnungen auf persönlichen Nutzen: Erwartung von den Ergebnissen der Forschung selbst zu profitieren, Hoffnung auf behandlungsrelevante Zufallsbefunde	<ul style="list-style-type: none"> <li>• Aufklärung und Widerspruchsmöglichkeit</li> </ul>
Ärzte/Institutionen (und indirekt Patienten)	erhöhter Dokumentationsaufwand	<ul style="list-style-type: none"> <li>• „arbeitsparende“ IT-Infrastruktur</li> </ul>
	Leistungsvergleich/ Ökonomisierung	<ul style="list-style-type: none"> <li>• Data Transfer Agreements/ Datennutzungsbestimmungen</li> </ul>

bindung mit der genannten Transparenz in Aufklärung und Kommunikation einem möglichen Vertrauensverlust von Patienten in ihre Ärzte, Gesundheitsinstitutionen oder in das Gesundheitssystem entgegenwirken. Die Auseinandersetzung mit der Frage ob die Widerspruchsmöglichkeit (im Sinne einer Minimalanforderung) oder aber vielmehr eine der verschiedenen Formen einer aktiven Einwilligung (z. B. Broad Consent, Tiered Consent, Meta-Consent, Dynamic Consent) als angemessen für SeConts angesehen werden kann, ist nach wie vor eine viel diskutierte Frage, der wir hier, wie bereits erwähnt, nicht nachgehen können. Tab. 4 fasst die erwähnten Risiken für Patienten und andere Stakeholder und die vorgeschlagenen Gegenmaßnahmen zusammen.

## 7 Resümee

Wir haben im vorliegenden Aufsatz einen Vorschlag gemacht, die Sekundärnutzung klinischer Daten in datensammelnden, nicht-interventionellen Forschungs- oder Lernaktivitäten (SeConts) begrifflich zu fassen als: *Nutzung von Daten aus der medizinischen Versorgung für Forschungs- oder Lernaktivitäten, ohne körperliche Eingriffe oder*

*zusätzliche Maßnahmen zur Datengenerierung, zum Zweck der Verbesserung des biomedizinischen Wissens und der medizinischen Versorgung, jedoch nicht zum Eigennutzen des Patienten.*

Im nächsten Schritt haben die wir die Vielzahl möglicher Arten von SeConts anhand konkreter Beispiele dargestellt und in Anwendungsfelder unterteilt. Als Anwendungsfelder von SeConts können *nicht-interventionelle klinische Forschung, Aktivitäten zur Kontrolle und Erforschung der Versorgung (oder bestimmter Versorgungssituationen)* und *Public Health Forschung* sowie in allen drei Bereichen die *explorative Verwendung klinischer Daten* gelten (vgl. Tab. 1).

Im Anschluss an die begriffliche Analyse von SeConts haben wir deren mögliche Nutzenpotenziale dargestellt, die wir in forschungsökonomischen, forschungsmethodischen und forschungsethischen Nutzen sowie im Nutzen für einzelne Patientengruppen sehen (vgl. Tab. 2). Gleichzeitig haben wir Herausforderungen für SeConts aufgezeigt, welche wir in der Datenqualität- und Vollständigkeit, der Interoperationalisierbarkeit und in möglichen Verzerrungen (Bias) identifizieren (vgl. Tab. 3). SeConts bringt, wenngleich auch keine physischen Risiken für Patienten, Risiken unterschiedlicher Art sowohl für Patienten als auch für andere Stakeholder mit sich. Zu diesen Risiken gehört für Patienten das Risiko der Re-Identifikation und des Datenmissbrauchs, Gefahren für das Vertrauen in der Arzt-Patientenbeziehung und das Wecken falscher Hoffnungen. Auf Seiten von Ärzten und Institutionen ist ein erhöhter Dokumentationsaufwand denkbar sowie ein erhöhter Leistungsdruck für Klinikmitarbeiter (und Kliniken) aufgrund der Möglichkeit eines Vergleiches der Daten unterschiedlicher Behandler oder Institutionen.

Mögliche Gegenmaßnahmen zu den genannten Risiken bestehen (neben der Schaffung einer „arbeitsparenden“ IT-Infrastruktur) unter anderem in einer SeConts vorgelagerten Identifikation und Einschätzung der Höhe des Risikos für Patienten. In der Folge können technisch-organisatorische Maßnahmen ergriffen werden, um das Risiko für Patienten zu verringern. Ebenso ist die Verwaltung der Daten durch ein Data Access Committee notwendig, welches sowohl die Datenschutzinteressen der Patienten als auch die legitimen Interessen und Befürchtungen anderer Stakeholder im Blick hat und den Datenzugriff regelt. Zusätzlich bleibt die Patientenaufklärung und Widerspruchsmöglichkeit ein wichtiges Element einer ethisch vertretbaren Durchführung von SeConts (vgl. Tab. 4).

Im Zuge des zunehmenden Interesses an SeConts ist zu erwarten, dass verschiedene Personen oder Einrichtungen (Wissenschaftler, Bioethiker, Forschungsförderer, Ethikkommissionen und Use and Access Committees) die wissenschaftliche und ethische Qualität und Legitimität bestimmter SeConts Vorhaben begutachten. Wir sind der Überzeugung dass bei einer derartigen Begutachtung all die Aspekte, die wir im vorliegenden Text ansprechen, von Relevanz sind, angefangen von den Herausforderungen über die Nutzenpotenziale bis hin zu den Risiken für verschiedene Stakeholder und den möglichen Maßnahmen der Governance und Risikoreduzierung.

**Danksagung** Dieses Manuskript wurde im Rahmen des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts „Learning from Clinical Data (LinCDat)“ erstellt. Wir möchten uns bei unseren Projektpartnern Kai Cornelius und Markus Spitz (Universität Heidelberg) bedanken. Für wertvolle Hinweise möchten wir uns zudem bei Roland Jahns (Universität Würzburg) und Nikolaus Kleindienst (Zentralinstitut für Seelische Gesundheit, Mannheim) bedanken. Ebenso danken wir Vincent Lotz (Nationales Centrum für Tumorerkrankungen, Heidelberg) für Unterstützung bei der Recherche zu diesem Manuskript.

**Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – 406103282**

---

## Literatur

- Adair JG (1984) The Hawthorne effect: a reconsideration of the methodological artifact. *J Appl Psychol* 69(2):334
- Agaku IT, Adisa AO, Ayo-Yusuf OA, Connolly GN (2014) Concern about security and privacy, and perceived control over collection and use of health information are related to withholding of health information from healthcare providers. *J Am Med Inform Assoc* 21(2):374–378. <https://doi.org/10.1136/amiajnl-2013-002079>
- Aitken M, de St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S (2016) Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics* 17(1):73. <https://doi.org/10.1186/s12910-016-0153-x>
- American Medical Informatics Association (2007) Secondary uses and re-uses of healthcare data: taxonomy for policy formulation and planning
- Ancker JS, Shih S, Singh MP, Snyder A, Edwards A, Kaushal R, Investigators H (2011) Root Causes Underlying Challenges to Secondary Use of Data. In: AMIA Annual Symposium, S 57–62
- Ärzteblatt (2020) Tausende Datensätze von Patienten in Berliner Krankenhaus gestohlen. <https://www.aerzteblatt.de/nachrichten/108969/Tausende-Datensaetze-von-Patienten-in-Berliner-Krankenhaus-gestohlen>. Zugegriffen: 15. Mai 2020
- Bahnert R (2018) „medileaks.cc“: Der größte bisher bekannte Datenklau im Gesundheitswesen. <https://www.medical-tribune.de/praxis-und-wirtschaft/ehealth/artikel/medileaks-cc-der-groesste-bisher-bekannte-datenklau-im-gesundheitswesen/>. Zugegriffen: 15. Mai 2020
- Been JV, Mackay DF, Millett C, Pell JP, van Schayck OC, Sheikh A (2015) Impact of smoke-free legislation on perinatal and infant mortality: a national quasi-experimental study. *Sci Rep* 5:13020. <https://doi.org/10.1038/srep13020>
- Bender S, Brand R, Bacher J (2001) Re-identifying register data by survey data: an empirical study. *Stat J United Nations ECE* 18(4):373–381
- Benedetti F, Carlino E, Piedimonte A (2016) Increasing uncertainty in CNS clinical trials: the role of placebo, nocebo, and Hawthorne effects. *Lancet Neurol* 15(7):736–747
- Benitez K, Malin B (2010) Evaluating re-identification risks with respect to the HIPAA privacy rule. *J Am Med Inform Assoc* 17(2):169–177. <https://doi.org/10.1136/jamia.2009.000026>
- Botsis T, Hartvigsen G, Chen F, Weng C (2010) Secondary use of EHR: data quality issues and informatics opportunities. In: *Summit on Translational Bioinformatics*, S 1
- Brakewood B, Poldrack RA (2013) The ethics of secondary data analysis: considering the application of Belmont principles to the sharing of neuroimaging data. *Neuroimage* 82:671–676. <https://doi.org/10.1016/j.neuroimage.2013.02.040>

- Braun M, Dabrock P (2016) Ethische Herausforderungen einer sogenannten Big-Data basierten Medizin. *Z Med Eth* 62:313–329. <https://doi.org/10.14623/zfme.2016.4.313-329>
- Bronsert MR, Henderson WG, Valuck R, Hosokawa P, Hammermeister K (2013) Comparative effectiveness of antihypertensive therapeutic classes and treatment strategies in the initiation of therapy in primary care patients: a Distributed Ambulatory Research in Therapeutics Network (DARTNet) study. *JABFM* 26(5):529–538
- Bundesamt für Gesundheit (2017) Aktuelle Entwicklungen in der datengetriebenen Medizin und die damit verbundenen Herausforderungen und Aufgaben für das BAG. Bundesamt für Gesundheit, Bern
- Bundesministerium für Bildung und Forschung (2015) Förderkonzept Medizininformatik. Daten vernetzen – Gesundheitsversorgung verbessern. Bundesministerium für Bildung und Forschung (BMBF), Berlin
- Campos-Castillo C, Anthony DL (2015) The double-edged sword of electronic health records: implications for patient disclosure. *J Am Med Inform Assoc* 22(e1):e130–140. <https://doi.org/10.1136/amiajnl-2014-002804>
- Cheah PY, Piasecki J (2020) Data access committees. *BMC Med Ethics* 21(1):12. <https://doi.org/10.1186/s12910-020-0453-z>
- Committee on the Learning Health Care System in America IoM (2013) Best care at lower cost: the path to continuously learning health care in America. National Academies Press, Washington, D.C.
- Conway PH, Cnaan A, Zaoutis T, Henry BV, Grundmeier RW, Keren R (2007) Recurrent urinary tract infections in children. Risk factors and association with Prophylactic Antimicrobials. *JAMA* 298(2):179–186
- Dankar FK, Emam KE, Neisa A, Roffey T (2012) Estimating the re-identification risk of clinical data sets. *BMC Med Inform Decis Mak* 12(1):66
- de Bruin B (2010) The liberal value of privacy. *Law Philos* 29(5):505–534. <https://doi.org/10.1007/s10982-010-9067-9>
- Deutschlandfunk (2018) Cyberkriminalität im Gesundheitswesen. Manipulation, Erpressung, Datenklau. [https://www.deutschlandfunk.de/cyberkriminalitaet-im-gesundheitswesen-manipulation.684.de.html?dram:article\\_id=408220](https://www.deutschlandfunk.de/cyberkriminalitaet-im-gesundheitswesen-manipulation.684.de.html?dram:article_id=408220). Zugegriffen: 15. Mai 2020
- El Emam K, Dankar FK, Vaillancourt R, Roffey T, Lysyk M (2009) Evaluating the risk of re-identification of patients from hospital prescription records. *CJHP* 62(4):307
- Elger BS, Iavindrasana J, Lo Iacono L, Muller H, Roduit N, Summers P, Wright J (2010) Strategies for health data exchange for secondary, cross-institutional clinical research. *Comput Methods Programs Biomed* 99(3):230–251. <https://doi.org/10.1016/j.cmpb.2009.12.001>
- Erdmann P (2015) Zufallsbefunde aus bildgebenden Verfahren in populationsbasierter Forschung. *mentis*, Münster
- Evans RS, Burke JP, Classen DC, Gardner RM, Menlove RL, Goodrich KM, Stevens LE, Pestotnik SL (1992) Computerized identification of patients at high risk for hospital-acquired infection. *Am J Infect Control* 20(1):4–10
- Faden RR, Kass NE, Goodman SN, Pronovost P, Tunis S, Beauchamp TL (2013) An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics. *Hastings Cent Rep* 43(1):S16–S27. <https://doi.org/10.1002/hast.134>
- Fernau S, Mehlich K, Schildmann J, Krause S, Winkler EC (2017) The role of physicians in rationing Cancer care Attitudes of German oncologists. *Oncol Research Treat* 40(9):490–494
- Floyd T, Grieco M, Reid EF (2016) Mining Hospital Data Breach Records: Cyber Threats to U.S. Hospitals. Vortrag. In: 2016 IEEE Conference on Intelligence and Security Informatics (ISI), Tucson, AZ, USA



- Geissbuhler A, Safran C, Buchan I, Bellazzi R, Labkoff S, Eilenberg K, Leese A, Richardson C, Mantas J, Murray P, De Moor G (2013) Trustworthy reuse of health data: a transnational perspective. *Int J Med Inform* 82(1):1–9. <https://doi.org/10.1016/j.ijmedinf.2012.11.003>
- Gitter DM (2019) The ethics of big data in genomics: the instructive icelandic Saga of the incidentalome. *Wash U Global Stud L Rev* 18(2):351–390
- Hay JW, Hay AR (1992) Inflammatory bowel disease: costs-of-illness. *J Clin Gastroenterol* 14(4):309–317
- Heatherly RD, Loukides G, Denny JC, Haines JL, Roden DM, Malin BA (2013) Enabling genomic-phenomic association discovery without sacrificing anonymity. *PLoS One* 8(2):e53875. <https://doi.org/10.1371/journal.pone.0053875>
- Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, Lehmann HP, Hripcsak G, Hartzog TH, Cimino JJ, Saltz JH (2013) Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care* 51(8 Suppl 3):S30–S37. <https://doi.org/10.1097/MLR.0b013e31829b1dbd>
- Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E (2018) Considerations for ethics review of big data health research: a scoping review. *PLoS One* 13(10):e0204937. <https://doi.org/10.1371/journal.pone.0204937>
- Institute of Medicine (2009) Initial national priorities for comparative effectiveness research. The National Academies Press, Washington, DC
- Jagsi R, Motomura AR, Amarnath S, Jankovic A, Sheets N, Ubel PA (2009) Under-representation of women in high-impact published clinical cancer research. *Cancer* 115(14):3293–3301
- Jones KH, Laurie G, Stevens L, Dobbs C, Ford DV, Lea N (2017) The other side of the coin: Harm due to the non-use of health-related data. *Int J Med Inform* 97:43–51. <https://doi.org/10.1016/j.ijmedinf.2016.09.010>
- Jungkunz M, Köngeter A, Mehlis K, Winkler EC, Schickhardt C (2021) Secondary use of clinical data in data-gathering, non-interventional research or learning activities: Definition, types, and a framework for risk assessment. *J Med Internet Res* 23(6):e26631. <https://doi.org/10.2196/26631>
- Kaplan B (2016) How should health data be used? *Camb Q Healthc Ethics* 25(2):312–329. <https://doi.org/10.1017/S0963180115000614>
- Kass NE, Faden RR, Goodman SN, Pronovost P, Tunis S, Beauchamp TL (2013) The research-treatment distinction: a problematic approach for determining which activities should have ethical oversight. *Hastings Cent Rep* 43(1):S4–S15. <https://doi.org/10.1002/hast.133>
- Kaye J, Gibbons S, Heeney C, Smart A (2012) *Governing biobanks: understanding the interplay between law and practice*. Bloomsbury Publishing, London
- Kelley M, James C, Alessi Kraft S, Korngiebel D, Wijangco I, Rosenthal E, Joffe S, Cho MK, Wilfond B, Lee SS (2015) Patient perspectives on the learning health system: the importance of trust and shared decision making. *Am J Bioeth* 15(9):4–17. <https://doi.org/10.1080/15265161.2015.1062163>
- King T, Brankovic L, Gillard P (2012) Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. *Int J Med Inform* 81(4):279–289
- Kopcke F, Kraus S, Scholler A, Nau C, Schuttler J, Prokosch HU, Ganslandt T (2013) Secondary use of routinely collected patient data in a clinical trial: an evaluation of the effects on patient recruitment and data acquisition. *Int J Med Inform* 82(3):185–192. <https://doi.org/10.1016/j.ijmedinf.2012.11.008>
- Krause SW, Schildmann J, Lotze C, Winkler EC (2013) Rationing cancer care: a survey among the members of the german society of hematology and oncology. *J Natl Compr Canc Ne* 11(6):658–665

- Kuter DJ, Mehta A, Hollak CE, Giraldo P, Hughes D, Belmatoug N, Brand M, Muller A, Schaaf B, Giorgino R, Zimran A (2013) Miglustat therapy in type 1 Gaucher disease: clinical and safety outcomes in a multicenter retrospective cohort study. *Blood Cells Mol Dis* 51(2):116–124. <https://doi.org/10.1016/j.bcmd.2013.04.005>
- Laurie G, Jones KH, Stevens L, Dobbs C (2014) A review of evidence relating to harm resulting from uses of health and biomedical data. Nuffield Council on Bioethics
- Lee ES, Black A, Harrington RD, Tarczy-Hornoch P (2015) Characterizing secondary use of clinical Data. Vortrag, AMIA Summits on Translational Science Proceedings
- Lee LM (2017) Ethics and subsequent use of electronic health record data. *J Biomed Inform* 71:143–146. <https://doi.org/10.1016/j.jbi.2017.05.022>
- Liu V, Musen MA, Chou T (2015) Data breaches of protected health information in the United States. *JAMA* 313(14):1471–1473
- Lucero RJ, Kearney J, Cortes Y, Arcia A, Appelbaum P, Fernandez RL, Luchsinger J (2015) Benefits and risks in secondary use of digitized clinical data: views of community members living in a predominantly ethnic minority urban neighborhood. *AJOB Empir Bioeth* 6(2):12–22. <https://doi.org/10.1080/23294515.2014.949906>
- Martin-Sanchez FJ, Aguiar-Pulido V, Lopez-Campos GH, Peek N, Sacchi L (2017) Secondary use and analysis of big data collected for patient care. *Yearb Med Inform* 26(1):28–37. <https://doi.org/10.15265/IY-2017-008>
- Mathews JD, Forsythe AV, Brady Z, Butler MW, Goergen SK, Byrnes GB, Giles GG, Wallace AB, Anderson PR, Guiver TA, McGale P, Cain TM, Dowty JG, Bickerstaffe AC, Darby SC (2013) Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *BMJ* 346:f2360. <https://doi.org/10.1136/bmj.f2360>
- Meystre SM, Lovis C, Burkle T, Tognola G, Budrionis A, Lehmann CU (2017) Clinical data reuse or secondary use: current status and potential future progress. *Yearb Med Inform* 26(1):38–52. <https://doi.org/10.15265/IY-2017-007>
- Mitchell H, Lewis D, Marsh K, Hughes G (2014) Distribution and risk factors of *Trichomonas vaginalis* infection in England: an epidemiological study using electronic health records from sexually transmitted infection clinics, 2009–2011. *Epidemiol Infect* 142(8):1678–1687
- Mokken R, Kooiman P, Pannekoek J, Willenborg L (1992) Disclosure risks for microdata. *Stat Neerl* 46(1):49–67
- Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, Dittus RS, Rosen AK, Elkin PL, Brown SH (2011) Automated identification of postoperative complications within an electronic medical record using natural language processing. *Jama* 306(8):848–855
- Oshikoya KA, Wharton GT, Avant D, Van Driest SL, Fenn NE, Lardieri A, Doe E, Sood BG, Taketomo C, Lieu P (2019) Serious adverse events associated with off-label use of azithromycin or fentanyl in children in intensive care units: a retrospective chart review. *Pediatr Drugs* 21(1):47–58
- Parker R, Aggleton P (2003) HIV and AIDS-related stigma and discrimination: a conceptual framework and implications for action. *Soc Sci Med* 57(1):13–24. [https://doi.org/10.1016/S0277-9536\(02\)00304-0](https://doi.org/10.1016/S0277-9536(02)00304-0)
- Pittet D, Safran E, Harbarth S, Borst F, Copin P, Rohner P, Scherrer J-R, Auckenthaler R (1996) Automatic alerts for methicillin-resistant *Staphylococcus aureus* surveillance and control: role of a hospital information system. *Infect Cont Hosp Ep* 17(8):496–502
- Platt J, Kardia S (2015) Public trust in health information sharing: implications for biobanking and electronic health record systems. *J Pers Med* 5(1):3–21. <https://doi.org/10.3390/jpm5010003>
- Prada-Ramallal G, Takkouche B, Figueiras A (2019) Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. *BMC Med Res Methodol* 19(1):53. <https://doi.org/10.1186/s12874-019-0695-y>

- Quintel M, Gattinoni L, Weber-Carstens S (2016) The German ECMO inflation: when things other than health and care begin to rule medicine. *Intens Care Med* 42:1264–1266
- Richter G, Krawczak M, Lieb W, Wolff L, Schreiber S, Buyx A (2018) Broad consent for health care–embedded biobanking: understanding and reasons to donate in a large patient sample. *Genet Med* 20(1):76–82
- Robertson AR, Nurmatov U, Sood HS, Cresswell K, Smith P, Sheikh A (2016) A systematic scoping review of the domains and innovations in secondary uses of digitised health-related data. *J Innov Health Inform* 23(3):611–619. <https://doi.org/10.14236/jhi.v23i3.841>
- Robinson KA, Baughman W, Rothrock G, Barrett NL, Pass M, Lexau C, Damaske B, Stefonek K, Barnes B, Patterson J, Zell ER, Schuchat A, Whitney CG (2001) Epidemiology of Invasive *Streptococcus pneumoniae* Infections in the United States, 1995–1998, Opportunities for prevention in the Conjugate Vaccine Era. *JAMA* 285(13):1729–1735
- Rumbold JMM, Pierscionek BK (2018) What are data? A categorization of the data sensitivity spectrum. *Big Data Res* 12:49–59. <https://doi.org/10.1016/j.bdr.2017.11.001>
- Rusanov A, Weiskopf NG, Wang S, Weng C (2014) Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. *BMC Med Inform Decis Mak* 14(1):51–59
- Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Detmer DE, Expert P (2007) Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc* 14(1):1–9. <https://doi.org/10.1197/jamia.M2273>
- Samor M, Lichtenberg D, Saubermann L, Kawachi C, Carmeli Y (1997) A clinical data repository enhances hospital infection control. *Proc AMIA Annu Fall Symp*, S 56–60
- Sardar MR, Badri M, Prince CT, Seltzer J, Kowey PR (2014) Underrepresentation of women, elderly patients, and racial minorities in the randomized trials used for cardiovascular guidelines. *JAMA Intern Med* 174(11):1868–1870
- Schilsky RL, Michels DL, Kearbey AH, Yu PP, Hudis CA (2014) Building a rapid learning health care system for oncology: the regulatory framework of CancerLinQ. *J Clin Oncol* 32(22):2373–2379. <https://doi.org/10.1200/JCO.2014.56.2124>
- Schuol S, Schickhardt C, Wiemann S, Bartram CR, Tanner K, Eils R, Meder B, Richter D, Glimm H, von Kalle C (2015) So rare we need to hunt for them: reframing the ethical debate on incidental findings. *Genome Med* 7(1):1–7
- Shabani M, Borry P (2016) “You want the right amount of oversight”: interviews with data access committee members and experts on genomic data access. *Genet Med* 18(9):892–897. <https://doi.org/10.1038/gim.2015.189>
- Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V (2019) Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform* 7(2):e12239
- Smith G, Hippisley-Cox J, Harcourt S, Heaps M, Painter M, Porter A, Pringle M (2007) Developing a national primary care-based early warning system for health protection—a surveillance tool for the future? Analysis of routinely collected data. *J Public Health (Oxf)* 29(1):75–82. <https://doi.org/10.1093/pubmed/fdl078>
- Stattin P, Holmberg E, Johansson JE, Holmberg L, Adolfsson J, Hugosson J, National Prostate Cancer Register of S (2010) Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden follow-up study. *J Natl Cancer Inst* 102(13):950–958. <https://doi.org/10.1093/jnci/djq154>
- Steinbusch PJ, Oostenbrink JB, Zuurbier JJ, Schaepkens FJ (2007) The risk of upcoding in casemix systems: a comparative study. *Health policy* 81(2–3):289–299

- Strech D, Borchers K, Freyer D, Neumann A, Wasem J, Marckmann G (2008) Ärztliches Handeln bei Mittelknappheit. *Ethik Med* 20(2):94–109
- Sturmer T, Jonsson Funk M, Poole C, Brookhart MA (2011) Nonexperimental comparative effectiveness research using linked healthcare databases. *Epidemiology* 22(3):298–301. <https://doi.org/10.1097/EDE.0b013e318212640c>
- Sweeney L (2000) Uniqueness of simple demographics in the US population. LIDAP-WP 4:2000
- Sweeney L (2002) k-anonymity: a model for protecting privacy. *Int J Uncertain Fuzziness Knowl-Based Syst* 10(05):557–570
- Terris DD, Litaker DG, Koroukian SM (2007) Health state information derived from secondary databases is affected by multiple sources of bias. *J Clin Epidemiol* 60(7):734–741. <https://doi.org/10.1016/j.jclinepi.2006.08.012>
- Vitale C, Fini M, Spoletini I, Lainscak M, Seferovic P, Rosano GM (2017) Under-representation of elderly and women in clinical trials. *Int J Cardiol* 232:216–221
- Warren-Gash C (2017) Linking and sharing routine health data for research in England. PHG Foundation, Cambridge
- Weichert T (2018) Big Data im Gesundheitsbereich. <http://www.abida.de/sites/default/files/ABIDA%20Gutachten-Gesundheitsbereich.pdf>. Zugegriffen: 15. Mai 2020
- Weiner M, Xie D, Tannen R (2008) Clinical trials in silico: rigorous assessment of treatment effect using electronic health records. *AMIA Annu Symp Proc*, S 1172
- Xu H, Doan S, Birdwell KA, Cowan JD, Vincz AJ, Haas DW, Basford MA, Denny JC (2010) An Automated Approach to Calculating the Daily Dose of Tacrolimus in Electronic Health Records. *Summit Transl Bioinform* 71–75
- Yeniterzi R, Aberdeen J, Bayer S, Wellner B, Hirschman L, Malin B (2010) Effects of personal identifier resynthesis on clinical text de-identification. *J Am Med Inform Assoc* 17(2):159–168

**Open Access** Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

