

Strategiepapier

„Nutzbarmachung digitaler Daten für KI- und datengetriebene Gesundheitsforschung“

Dieses Dokument wurde von den Mitgliedern der Arbeitsgruppe „Nutzbarmachung digitaler Daten“ des Forums Gesundheitsforschung erarbeitet, unterstützt durch die Geschäftsstelle des Forums. Nach abschließender Beratung und Beschluss hat das Forum Gesundheitsforschung dieses Dokument dem Bundesministerium für Bildung und Forschung am 03.12.2020 als Empfehlung übergeben.

Inhalt

Zusammenfassung	3
I. Ziel	15
II. Ausgangslage	15
1. Voraussetzungen für KI-Entwicklungen und datengetriebene Forschung in der Medizin	15
2. Initiativen zur Nutzung gesundheitsbezogener digitaler Daten für die Forschung	17
III. Handlungsempfehlungen	20
1. Weiterentwicklung und Bündelung von Initiativen zur Verbesserung des Datenzugangs	20
2. Entwicklung technischer Lösungen zur Datenverknüpfung, Datensicherheit und Datenschutz	23
3. Verbesserung der Verfügbarkeit und Qualität von Gesundheitsdaten	25
4. Förderung der Kultur des Datenteilens	26
5. Ethische und rechtliche Aspekte der Datennutzung	28
6. Patienteneinbindung bei der Forschung mit Gesundheitsdaten	32
IV. Exkurs	34
1. Zulassungs- und haftungsrechtliche Aspekte KI-basierter Anwendungen in der Medizin	35
2. Perspektivische Empfehlung: Ein Plädoyer für die Datenüberlassung für die Forschung	38
V. Literatur	41

Zusammenfassung

1. Ausgangslage

Von der Verfügbarkeit umfangreicher und qualitativ hochwertiger Daten aus Versorgung und Forschung hängt der Erfolg datengetriebener Gesundheitsforschung im Allgemeinen sowie der Entwicklung und der Anwendung von Methoden der Künstlichen Intelligenz (KI) im Besonderen maßgeblich ab. Zudem stellt der Zugang zu Datensätzen auch die Grundlage für die Nachvollziehbarkeit von Forschungsergebnissen dar. Viele nationale und internationale Initiativen zielen darauf ab, die Qualität und Verfügbarkeit von Forschungs- und Versorgungsdaten zu verbessern. Die aktuelle Corona-Pandemie zeigt, wie schnell Forschung Lösungsansätze erarbeiten kann, wenn Ergebnisse frühzeitig zur wissenschaftlichen Diskussion gestellt und Daten schnell verfügbar gemacht werden. Das vorliegende Strategiepapier des Forums Gesundheitsforschung greift diese und weitere Erfahrungen auf und gibt Handlungsempfehlungen zur Nutzbarmachung großer Datenmengen für KI-Entwicklungen und andere Formen datengetriebener Gesundheitsforschung. Primärer Fokus des Papiers ist die Identifizierung von, durch die Akteure des Forums Gesundheitsforschung, umsetzbaren Maßnahmen. Für die Verbesserung der Verfügbarkeit von Daten aus Versorgung und Forschung für datengetriebene Gesundheitsforschung sind auch Maßnahmen anderer Akteure unerlässlich. In einem Exkurs werden daher weitere Aspekte betrachtet, die für die Implementierung von KI-Entwicklungen und eine Vereinfachung der Datenüberlassung von hoher Relevanz sind. Die Handlungsempfehlungen hierzu richten sich an Akteure außerhalb des Forums Gesundheitsforschung.

Für KI-Entwicklungen und datenintensive Forschung müssen Versorgungs- und Forschungsdaten über Standort- und Disziplingrenzen hinweg zusammengebracht werden. Technischen und organisatorischen Lösungen zur Datenverknüpfung und Datensicherheit kommt hierbei eine wichtige Rolle zu. Die Herausforderungen sind jedoch nicht nur technischer Natur. Das Vertrauen der Forschenden und der Bevölkerung in die Sicherheit und wissenschaftliche Sinnhaftigkeit der Datennutzung und eine ausgewogene Balance von Kooperation und Wettbewerb in der Forschung sind weitere essentielle Faktoren. Darüber hinaus müssen ethische und rechtliche Rahmenbedingungen der Datennutzung und datengestützter Verfahren in der Medizin geklärt werden.

Daten aus der Versorgung und Forschung

Um Forschung mit Versorgungsdaten zu stärken, hat die Bundesregierung in ihrer Hightech-Strategie 2025 in der Mission „Forschung und Versorgung digital vernetzen – für eine intelligente Medizin“ das Ziel formuliert, bis 2025 an allen Universitätskliniken eine forschungskompatible elektronische Patientenakte zu implementieren. Die Umsetzung dieser elektronischen Patientenakte wird maßgeblichen Einfluss darauf haben, wie Individualdaten von Patienten bzw. Versicherten im Versorgungsalltag zukünftig genutzt und im Weiteren auch für die Forschung zur Verfügung gestellt werden können. Mit der Innovationsinitiative „Daten für Gesundheit“ haben das Bundesministerium für Bildung und Forschung, das Bundesministerium für Gesundheit und das Bundesministerium für Wirtschaft und Energie dazu Handlungsfelder definiert, die die Aktivitäten zur Stärkung der datengestützten Gesundheitsforschung bündeln und aufzeigen, wie es zukünftig weitergehen soll.

Für den Bereich der Forschung besteht die Herausforderung darin, das Wissen, das in verschiedenen Initiativen zur Verbesserung des Datenzugangs gewonnen wurde, zusammenzuführen und die aufgebauten bzw. sich im Aufbau befindlichen Infrastrukturen gut aufeinander abzustimmen. Offen ist bisher, welche konkreten Schritte hierzu erforderlich sind.

Die bisherige Entwicklung technischer Lösungen für Datenverknüpfung, Datensicherheit und Datenschutz hat gezeigt, dass die Ziele von KI und datengetriebener Forschung nicht im Widerspruch zum Schutz der Privatsphäre stehen müssen. Für die (Weiter-)Entwicklung praxistauglicher Modelle fehlt es jedoch an Möglichkeiten zur frühzeitigen Analyse und Bewertung hinsichtlich Leistung und Konformität der vorhandenen technischen Lösungen in definierten Anwendungsszenarien.

In verschiedenen Empfehlungen wurde bereits darauf hingewiesen, dass zur Steigerung der Qualität der Versorgungsdaten technische, infrastrukturelle und organisatorische Voraussetzungen weiter verbessert werden müssen. Darüber hinaus wird es als notwendig erachtet, ärztliches und pflegerisches Personal im Umgang mit digitalen Versorgungsdaten zu stärken und weiteres spezifisches Personal auszubilden und einzusetzen. Unklar ist, durch welche Anreize und Rahmenbedingungen dies befördert werden könnte.

Eine Kultur des Datenteilens ist bisher nicht ausreichend etabliert. Hauptgründe hierfür sind die bestehende Reputationskultur und Konkurrenzsituation, aber auch Einschränkungen aufgrund engefasseter Patienteneinwilligungen (s. u.). Datenteilen wird bisher bei der Drittmittelförderung und innerhalb der Akademia zu wenig bis gar nicht honoriert.

Der rechtliche und ethische Rahmen

Die Nutzung von Forschungs- und Versorgungsdaten für KI-Entwicklungen und datenreiche Forschung wird wesentlich durch rechtliche und ethische Rahmenbedingungen bestimmt. Die EU-Datenschutz-Grundverordnung (DSGVO) bildet den grundlegenden – und grundsätzlich forschungsfreundlichen – rechtlichen Rahmen für die Verarbeitung personenbezogener Daten in der medizinischen Forschung. Die DSGVO gewährt auch den Rahmen für einen Broad Consent, d. h. die Möglichkeit der Einwilligung in die Datenverarbeitung für bestimmte Bereiche wissenschaftlicher Forschung entsprechend Erwägungsgrund 33 zur DSGVO¹. Mit dem Patientendatenschutz-Gesetz (PDSG) wurde die Möglichkeit des Broad Consent in § 363 Abs. 8 SGB V gesetzlich verankert. Daneben eröffnet § 363 Abs. 1 bis 7 SGB V die Möglichkeit der Datenfreigabe an das Forschungsdatenzentrum.

Die Möglichkeit des Broad Consent wird zum Teil von einzelnen Ethikkommissionen und Datenschutzbeauftragten unterschiedlich ausgelegt. Mit der Neuregelung des PDSG sollte hier eine gesetzliche Orientierung geschaffen worden sein, die es nun in der Praxis mit Leben zu füllen gilt.

Als selbstlernende und teilautonome Systeme können KI-Anwendungen zudem in einem rechtlichen Spannungsverhältnis zum Verbot der automatisierten Entscheidungsfindung (Art. 22 Abs. 1 DSGVO) stehen. Es ist derzeit unklar, wann genau von einer automatisierten Entscheidung auszugehen ist und welche Informationen über die verwendeten Algorithmen und zugrundeliegenden Daten im Einzelnen zur Verfügung gestellt werden müssen.

Darüber hinaus wird der Marktzugang KI-basierter Medizinprodukte-Software im gegenwärtigen Recht nicht berücksichtigt und dadurch erschwert.

¹ Erwägungsgrund 33 (Einwilligung zur wissenschaftlichen Forschung) zur EU-Datenschutz-Grundverordnung: Oftmals kann der Zweck der Verarbeitung personenbezogener Daten für Zwecke der wissenschaftlichen Forschung zum Zeitpunkt der Erhebung der personenbezogenen Daten nicht vollständig angegeben werden. Daher sollte es betroffenen Personen erlaubt sein, ihre Einwilligung für bestimmte Bereiche wissenschaftlicher Forschung zu geben, wenn dies unter Einhaltung der anerkannten ethischen Standards der wissenschaftlichen Forschung geschieht. Die betroffenen Personen sollten Gelegenheit erhalten, ihre Einwilligung nur für bestimmte Forschungsbereiche oder Teile von Forschungsprojekten in dem vom verfolgten Zweck zugelassenen Maße zu erteilen.

Patientenautonomie und Patienteneinbindung

Patientenautonomie und Patienteneinbindung spielen eine wichtige Rolle für die Bereitschaft, Gesundheitsdaten für Forschungszwecke freizugeben. Die fortschreitende Digitalisierung vieler Lebensbereiche ermöglicht es Patientinnen und Patienten, auch eigene digitale Gesundheitsdaten in Forschungsprojekte einzubringen. Durch die elektronische Patientenakte werden zukünftig viele für die Forschung relevante Daten bei den Patientinnen und Patienten selber zusammengeführt. Damit stellt sich die Frage, wie Patientinnen und Patienten an datenreicher medizinischer Forschung partizipieren können (und wollen) und welche Partizipationsmöglichkeiten neue Informationstechnologien eröffnen könnten.

2. Handlungsempfehlungen

Basierend auf der Analyse der Ausgangslage hat die Arbeitsgruppe des Forums Gesundheitsforschung die im Folgenden beschriebenen Empfehlungen erarbeitet. Die Empfehlungen sind in der Übersichtstabelle 1 auf S. 8 zusammengefasst. In der Tabelle sind sowohl die für die Umsetzung der Empfehlungen vorgeschlagenen Akteure benannt als auch die für erforderlich gehaltenen Umsetzungsschritte aufgeführt.

Dabei werden zum einen notwendige Meilensteine und Aktivitäten der Akteure des Forums vorgeschlagen, um den Zugang und die Nutzbarkeit von Daten für KI- und datengetriebene Gesundheitsforschung zu verbessern. In ihrer 11. Sitzung am 04.11.2020 haben sich die im Forum Gesundheitsforschung vertretenen Akteure über ein abgestimmtes Vorgehen verständigt und dazu verpflichtet, die in ihrer Zuständigkeit liegenden Empfehlungen umzusetzen.

Zum anderen werden in einem Exkurs ergänzende Aspekte betrachtet, die für die Implementierung von KI-Entwicklungen und eine Vereinfachung der Datenüberlassung von hoher Relevanz sind. Die damit verbundenen Herausforderungen müssen ebenfalls zügig angegangen werden. Sie liegen jedoch nicht im unmittelbaren Aktionsrahmen der Akteure des Forums Gesundheitsforschung, sondern sind Empfehlungen an weitere Akteure.

1. Weiterentwicklung und Bündelung von Initiativen zur Verbesserung des Datenzugangs

Die durch verschiedene Initiativen zur Verbesserung des Datenzugangs für KI- und datengetriebene Gesundheitsforschung angestoßenen Lösungen und aufgebauten Infrastrukturen sollten aufeinander abgestimmt werden und dort, wo dies sinnvoll ist, zusammengeführt werden. Ziel ist ein interoperables, nutzerfreundliches und nachhaltiges System für die Integration und den Austausch von Daten und erforderlichen Metadaten aus allen Bereichen der Gesundheitsforschung und der klinischen Versorgung zum Wohl der Patientinnen und Patienten. Darüber hinaus müssen einheitliche Datenmodelle für präklinische und klinische Forschungsdaten entwickelt, Datenbanken zu relevanten klinischen Fragestellungen aufgebaut und einheitliche Erfassungssysteme für Studiendaten zur Verfügung stehen und angewendet werden. All dies muss sorgfältig mit der Entwicklung und dem Einsatz einer forschungskompatiblen elektronischen Patientenakte abgestimmt werden. Zur Weiterentwicklung der vorhandenen Initiativen zur Verbesserung des Datenzugangs sollten - unterstützt von BMBF und DFG - die wichtigsten Initiativen gemeinsam, den dafür notwendigen Konvergenzprozess einleiten. Hierbei sind auch das BMG und die Gematik einzubinden, um die Abstimmung mit der elektronischen Patientenakte sicherzustellen. Darüber hinaus sollten vom BMBF Pilotprojekte gefördert werden zur Integration von DZG- und DIZ-Datenbeständen der MII in spezifischen Anwendungsfeldern unter Sicherstellung von Konvergenz mit NFDI-Strukturen und zur Nutzung dieser integrierten Datenbestände.

2. Entwicklung technischer Lösungen zur Datenverknüpfung, Datensicherheit und Data Privacy

Neuere KI-Technologien ermöglichen föderiertes maschinelles Lernen, d. h. dezentrales Trainieren von Algorithmen über mehrere Standorte hinweg, bei dem individuelle Patienten-/Probandenrechte möglichst wenig tangiert werden. Eine systematische Evaluierung dieser Lösungsansätze ist bisher nicht erfolgt. Es sollten daher Förderangebote vorhanden sein, in denen regelmäßig systematische Performance-Monitorings („Benchmarking Competitions²“) durchgeführt werden. Diese erlauben es den Stand der Implementierung von föderierten Trainingsansätzen von Vorhersage-Modellen in wohl definierten Szenarien zu eruieren und zu bewerten. Universitäre und außeruniversitäre Einrichtungen und Zuwendungsgeber sollten vermehrt Benchmarking Competitions durchführen bzw. fördern. Diese Fördermaßnahme sollte hierfür geeignete, bestehende Infrastrukturen (z. B. der MII und der DZG) sinnvoll kombinieren und unter Einbezug medizinischer Fachexperten praktisch anwenden. Darüber hinaus sollten in Fördermaßnahmen auch Mittel zur Verfügung gestellt werden, um neue Konzepte, Designs, Methodiken und Standards, z. B. über europäische Codes of Conducts gemäß Art. 40 DSGVO, für datenintensive Forschung im Rahmen des Datenschutzes zu entwickeln oder zu validieren, die über Projekte hinweg Verwendung finden können. Dabei sollten die Schnittstellen und Bezüge zu anderen wissenschaftlichen Disziplinen wie Mathematik, Informatik oder Ingenieurwissenschaften weiter ausgebaut werden.

3. Verbesserung der Verfügbarkeit und Qualität von Gesundheitsdaten

Um Versorgungsdaten für die Forschung nutzbar zu machen, sind mehrere Schritte erforderlich. Hierzu gehören u. a. die bundesweit einheitliche Nutzung international anerkannter und erprobter Standards, technische und semantische Interoperabilität und eine strukturierte Datenerfassung. Die diesbezüglichen Herausforderungen an eine elektronische Patientenakte für die Versorgung sind dabei denen für die Nutzung der Daten für die Forschung sehr ähnlich. Auf die von der MII in Kooperation mit den DZG und Biobanken entwickelten Kerndatensätze muss daher bei der Entwicklung der Interoperabilität der elektronischen Patientenakte zurückgegriffen werden. Am Ende muss ein Prozess stehen, der Patientendaten nur einmal in möglichst niederschwelliger, KI-unterstützter und strukturierter Weise in ein IT- System überträgt und diese dann für die Versorgung wie auch die Forschung im Sinne der Patienten gleichermaßen nutzbar macht. KI-basierte Verfahren werden helfen, niederschwellige Dateneingaben sowie Auswertungen zu ermöglichen.

Auch müssen ärztliches und pflegerisches Personal durch Aus-, Fort- und Weiterbildung und geeignete technische Ausstattung befähigt, motiviert und incentiviert werden, digitale Versorgungsdaten effizienter und umfangreicher zur Verbesserung der individuellen Patientenversorgung sowie der Versorgungssteuerung zu nutzen – nicht zuletzt durch Anwendung von KI-basierten Verfahren. Die Programme für Clinician und Medical Scientists sollten ebenfalls durch geeignete Module ergänzt werden. So wie der Nationale Kompetenzbasierte Lernzielkatalog Medizin (NKLM) zukünftig die digitale Medizin longitudinal im Medizinstudium verankert, sollte dies auch in der Ausbildung der weiteren Gesundheitsberufe integriert werden. Begleitet werden muss dies durch Fort- und Weiterbildungsangebote für bereits im Beruf stehende Arbeitskräfte. Neben der Bereitstellung der dafür notwendigen Technologien muss

² Eine Benchmarking Competition ist ein Instrument, um zu evaluieren, ob und wie, und auch wie gut, sich Vorhersage-Modelle föderiert, also ohne Bewegen der Trainingsdaten, trainieren lassen.

auch die Ausbildung und der Einsatz sogenannter „Data (Curation) Scientists“ in der sektorübergreifenden Versorgung bis hin zur Einrichtung neuer Masterstudiengänge für z.B. (Clinical) Data Scientists verstärkt werden.

Für die Abstimmung zu bereits laufenden und neuen Aktivitäten zur Aus-, Fort- und Weiterbildung sollte durch die Universitätsmedizin unter Einbindung von MII- und DZG-Vertretern, Partnern der Selbstverwaltung, Fachverbänden, Gesundheitsfachschulen, des stationären und ambulanten Sektors, der Technologieentwickler sowie Bund und Ländern ein entsprechender Steuerungskreis eingerichtet werden. Die sinnvolle Einbindung medizinischen Fachpersonals in die Anwendung von KI-Systemen ist jedoch in hohem Ausmaß davon abhängig, dass die Systeme nicht parallel zur Expertise des Personals agieren, sondern in Interaktion mit ihnen. Die hierfür notwendige Entwicklung interaktiver Technologien sollte gemeinsam durch Gesundheitsversorgung, Forschung und Industrie erfolgen.

Das Forum Gesundheitsforschung unterstützt darüber hinaus die bereits von verschiedenen Initiativen geforderte Etablierung von für datengetriebene Forschung und KI-Entwicklungen verfügbaren qualitätsgesicherten Trainings- und Testdatenpools.

4. Förderung der Kultur des Datenteilens

Neben der Schaffung geeigneter technischer und organisatorischer Rahmenbedingungen, wie sie in diesem Papier beschrieben werden, sollten durch gezielte Maßnahmen stärkere Anreize für das Datenteilen geschaffen werden. BMBF und DFG sollten hierzu zeitnah - unter Berücksichtigung der Freiheit der Wissenschaft und Forschung (Art. 5 GG) - in einem gemeinsamen Vorgehen eine stärkere Verbindlichkeit in Bezug auf die Umsetzung des Datenteilens und der FAIR-Data-Prinzipien (Findable, Accessible, Interoperable, Reusable) umsetzen. Das BMBF wird gebeten, auch andere Ressorts, die Forschung im medizinischen Bereich fördern, einzubeziehen. Entsprechende Anforderungen an Antragstellende sollten auch eine Zeitfrist beinhalten, die es den Forschenden erlaubt, zunächst ihre Daten für eine eigene Publikation zu nutzen, bevor sie offengelegt werden. Mittelfristig sollte das akademische Bewertungssystem so weiterentwickelt werden, dass die Integration und Veröffentlichung von Daten analog zu einer wissenschaftlichen Leistung gewürdigt werden. Eine entsprechende Weiterentwicklung des akademischen Bewertungssystems ist innerhalb von Arbeitsgruppen der DFG bereits aufgegriffen worden. Diese Initiative sollte in der Allianz der Wissenschaftsorganisationen besprochen und umgesetzt werden. Zudem sollten Data Challenges³ als Instrument zur Vertrauensbildung und zum Anschub der translationalen Anwendung von KI zu ausgewählten Fragestellungen in aktuell bedeutsamen medizinischen Themengebieten ausgeschrieben und gefördert werden.

5. Ethische und rechtliche Aspekte der Datennutzung

Aufbauend auf existierenden Dokumenten sollten, koordiniert durch die MII unter Einbeziehung der Wissenschaft sowie im engen Austausch mit den Datenschutzaufsichtsbehörden und medizinischen Ethikkommissionen, praxisnahe Empfehlungen und Mustertexte für Datenerhebung und -verarbeitung weiterentwickelt werden. Diese sollten so umgesetzt werden, dass sie Patientinnen und Patienten eine souveräne Entscheidung ermöglichen. Soweit angebracht sollten u. a. auch erprobte Prozesse für Widerruf, Speicherfristen, Zweckbindung

³ Die Data Challenge ist ein Datenanalyse-Wettbewerb. Angewendet im Bereich der Entwicklung von Vorhersage-Modellen geht es um eine oder mehrere vorgegebene Aufgaben, die von verschiedenen, konkurrierenden Teams gelöst werden sollen. Ziel der Challenge ist es, einen dafür bereitgestellten Datensatz zu analysieren und mit dem Wissen über die Vergangenheit aus diesem Datensatz zukünftige Entwicklungen vorherzusagen.

und Anonymisierung einbezogen werden. Hierbei sind europäische Entwicklungen zu berücksichtigen: Die Datenschutzgrundverordnung (DSGVO) sieht in Art. 40 die Erarbeitung von nationalen oder EU-weit geltenden Verhaltensregeln vor. Diese werden aktuell durch die Akteure erarbeitet und den nationalen bzw. europäischen Datenschutzaufsichtsbehörden zur Annahme vorgelegt.

Mit den Regelungen zur Geltung des § 27 Bundesdatenschutzgesetzes (BDSG) und zur Federführenden Datenschutzaufsicht in der Versorgungs- und Gesundheitsforschung ist in dem neuen § 287a SGB V eine wesentliche Erleichterung für (bundes-)länderübergreifende Vorhaben der Versorgungs- und Gesundheitsforschung geschaffen worden. Trotzdem sollte eine weitere Vereinheitlichung der Auslegung des Datenschutzrechts auf Landes-, Bundes- und Unionsebene angestrebt werden.

In einem ersten Schritt gilt es jedoch, die im Jahr 2020 erreichten Verbesserungen und Klarstellungen (MII-Mustertext Consent; Patientendatenschutzgesetz, PDSG; § 287a SGB V) in Zusammenarbeit mit Vertretern der Wissenschaft weiter auszubauen und wo möglich in der Praxis der lebenswissenschaftlichen Forschung sowie der Datenschutzaufsicht zu verankern. Hierzu besteht Gesprächs- und Koordinationsbedarf mit allen betroffenen Akteuren. Hierbei ist auch die Frage des Zugangs zu Daten durch die Industrie bzw. Wirtschaft einzubeziehen.

Als selbstlernende und teilautonome Systeme können KI-Anwendungen in einem rechtlichen Spannungsverhältnis zum Verbot der automatisierten Entscheidungsfindung (Art. 22 Abs. 1, DSGVO) stehen. Es besteht Forschungsbedarf dazu, inwieweit dieses Verbot Auswirkungen bei KI-Anwendungen in der Medizin haben könnte und wie dieses Spannungsverhältnis aufgelöst werden kann. Dieser Bedarf sollte z. B. im Rahmen der Umsetzung der KI-Strategie der Bundesregierung aufgegriffen werden.

Empfehlungen zur rechtlichen Anpassung zur Verbesserung des Marktzugangs KI-basierter Medizinprodukte-Software sind im Exkurs aufgeführt.

6. Patienteneinbindung bei der Forschung mit Gesundheitsdaten

In einer DZG-übergreifenden Arbeitsgruppe sollten technische und organisatorische Instrumente entwickelt werden, die die Partizipationsmöglichkeiten der Patientinnen und Patienten bei datenreicher Forschung stärken. Es sollten konkrete Partizipationsmodelle entwickelt und Best-Practice-Beispiele beschrieben werden. Bei der Stärkung der Patienteneinbindung könnte zukünftig auch interaktiven Technologien eine stärkere Rolle zukommen. Technische Lösungsansätze hierfür entwickelt beispielsweise die Fraunhofer-Gesellschaft in der Medical Data Space-Infrastruktur. Darüber hinaus sollten praxistaugliche und nutzungsfreundliche Instrumente entwickelt und implementiert werden, die ein studien- und standortübergreifendes Consent Management für die Nutzung großer Datenmengen, insbesondere für KI-Anwendungen, erleichtern. Die Entwicklung der Instrumente sollte von Vertretern der DZG, der MII und der AUFGE gemeinsam und unter Einbeziehung von Patientenorganisationen abgestimmt und vorangetrieben werden. Die Instrumente sollten der wissenschaftlichen Gemeinschaft zur Verfügung gestellt werden.

Exkurs

Die folgenden Empfehlungen beziehen sich auf wichtige zulassungs- und haftungsrechtliche Aspekte bei der Implementierung datengestützter Verfahren in der Medizin und auf eine Vereinfachung der Datenüberlassung. Die Empfehlungen gehen über den primären Fokus des Papiers, nämlich die Identifizierung von durch die Akteure des Forums Gesundheitsforschung umsetzbare konkrete Maßnahmen zur Verbesserung der Nutzbarmachung digitaler

Daten für KI-Entwicklungen und datengetriebene Gesundheitsforschung hinaus. Sie richten sich primär an Akteure außerhalb des Forums Gesundheitsforschung. Aufgrund ihrer hohen Bedeutung für die tatsächliche Nutzung von KI-Entwicklungen in der Medizin und für eine Vereinfachung der Datenüberlassung empfiehlt das Forum Gesundheitsforschung den vorgeschlagenen Akteuren die Umsetzung der Maßnahmen.

1. Zulassungs- und haftungsrechtliche Aspekte KI-basierter Anwendungen in der Medizin

KI-basierte Anwendungen können medizinischen Zwecken dienen und sind dann als Medizinprodukte einzustufen. Zulassungs- und haftungsrechtliche Aspekte spielen daher beim Einsatz datengestützter Verfahren in der Medizin eine wichtige Rolle. Kontinuierlich lernende und adaptive KI-Algorithmen können einen anderen als den ursprünglich freigegebenen Output liefern. Es gibt bisher noch keine Medizinproduktegesetzgebung, die eine praxisgerechte kontinuierliche Bewertung der Sicherheit und Leistungsfähigkeit von medizinischen KI-basierten Anwendungen, die sich oder ihre Performance nach der Zulassung selbst verändern, erlaubt. Mit dem „Weißbuch zur Künstlichen Intelligenz - Ein europäisches Konzept für Exzellenz und Vertrauen“ hat die Europäische Kommission im Februar 2020 politische Optionen vorgelegt, die die schnelle und ethisch vertretbare Entwicklung der KI unterstützen und gleichzeitig die Achtung der europäischen Werte und Grundrechte gewährleisten sollen.

Beim Einsatz KI-basierter Anwendungen stellen sich zahlreiche Haftungsfragen. Im Bereich der Medizin greifen dabei verschiedene Haftungsregime mit unterschiedlichen Voraussetzungen ineinander, etwa die Arzthaftung und die Produkt- und Produzentenhaftung, mit entsprechend unterschiedlichen Haftungssubjekten. Bisher ungeklärt ist dabei insbesondere, welche Haftungssubjekte in welcher Rangfolge heranzuziehen sind und wie die Pflichten der Haftungssubjekte mit den Autonomie- bzw. Intelligenzrisiken, die selbstlernenden Systemen inhärent sind, abgeglichen werden können.

Im Hinblick auf zulassungsrechtliche Aspekte muss das für die Marktzulassung von KI-basierten Anwendungen einschlägige Medizinprodukterecht daraufhin überprüft werden, inwieweit es den Besonderheiten der kontinuierlich lernenden und adaptiven KI-basierten Software hinreichend Rechnung tragen kann (innovationsaffine Gestaltung der Regel 11 in Anhang VIII der EU-Medizinprodukteverordnung; Berücksichtigung der Besonderheiten von KI-basierter Software bei der Erstellung des geplanten Leitfadens „Artificial Intelligence under MDR/IVDR⁴ framework“). Änderungen sollten, sofern erforderlich, bei einer Aktualisierung des EU-Medizinproduktegesetzes und der Erstellung des o.g. Leitfadens sowie bei der Erarbeitung des künftigen KI-Rechtsrahmens der EU Berücksichtigung finden.

Im Hinblick auf haftungsrechtliche Aspekte gibt es einen Forschungs- und Prüfungsbedarf zu den verschiedenen haftungsrelevanten Bereichen von KI-basierten Anwendungen (z. B. Arzthaftungsrecht, Produkthaftungsrecht) und den bereichsübergreifenden rechtlichen Zusammenhängen.

2. Perspektivische Empfehlung: Ein Plädoyer für eine Datenüberlassung für die Forschung

Es wird dafür plädiert, unter Berücksichtigung der neuen gesetzgeberischen Entwicklungen zu untersuchen, wie eine dauerhafte, institutionalisierte Überlassung von Patientendaten an die Forschung im Rahmen der datenschutzrechtlichen Bestimmungen realisiert werden kann. Diese Überlassung sollte mit einer möglichst niederschweligen Möglichkeit zum Output verbunden sein. Dieses Vorgehen würde die Vorteile einer Entkopplung der Entscheidungsfindung vom akuten Versorgungskontext mit dem gesamtgesellschaftlichen Mehrwert einer umfangreichen Datenverfügbarkeit für die Forschung verbinden, ohne die Rechte des

⁴ Medical Devices Regulation / In Vitro Diagnostics Regulation

Individuums unverhältnismäßig einzuschränken. Den individuellen und solidargemeinschaftlichen Nutzen einer solchen Form der Datenüberlassung gilt es, für die Bürgerinnen und Bürger verständlich und transparent darzulegen. Akteure der Forschung, Versorgung und Politik sollten tragfähige Konzepte für die Datenüberlassung vorantreiben, d. h. diese umfassend technisch, rechtlich und ethisch bewerten und in Pilotprozessen mit ethischer, rechtlicher und soziologischer Begleitforschung erproben.

Tabelle 1: Empfohlene Aktivitäten der Akteure des Forums Gesundheitsforschung, um den Zugang und die Nutzbarkeit von Daten für KI-Entwicklungen und datengetriebene Gesundheitsforschung zu verbessern.

Über den Stand der Umsetzung der Empfehlungen soll von den jeweiligen Akteuren regelmäßig im Forum Gesundheitsforschung berichtet werden, um ein koordiniertes und aufeinander abgestimmtes Vorgehen zu gewährleisten.

Empfehlung	Akteure	Umsetzungsvorschlag
1. Weiterentwicklung und Bündelung von Initiativen zur Verbesserung des Datenzugangs		
<p>Konvergenzprozess der Infrastrukturen für den Datenzugang</p> <ul style="list-style-type: none"> • Entwicklung einheitlicher Datenmodelle für präklinische und klinische Forschungsdaten • Entwicklung einheitlicher Erfassungssysteme für Studiendaten außerhalb von DZG und MII • Erstellung eines interoperablen, nutzerfreundlichen und nachhaltigen Systems für die Integration und den Austausch von Daten aus allen Bereichen der biomedizinischen Forschung und der klinischen Versorgung 	<p>Universitätskliniken, Med. Fakultäten sowie weitere Fachbereiche, AUFE, MII, DZG, NFDI4Health, GHGA, Biobanken, DFG, BMBF, BMG</p>	<p>Prozess initiieren durch BMBF und DFG mit dem Ziel der Konvergenz von Dateninfrastrukturen. Identifizieren von Bedarf und Lücken für das Erreichen des Ziels. Angesprochen sind Initiativen wie MII, DZG, NFDI4Health, GHGA, Biobanken u. ggf. weiterer Akteure</p> <p>Förderung von Pilotprojekten zur Integration von DZG-Datensätze mit den DIZ-Datenbeständen der MII durch BMBF</p>
2. Entwicklung technischer Lösungen zur Datenverknüpfung, Datensicherheit und Data Privacy		
<p>Forschung an der Schnittstelle von Data Science, Gesundheit und Recht</p> <ul style="list-style-type: none"> • Performance-Monitorings („Benchmarking Competitions“) zur Evaluierung von KI-Lösungen mit förderiertem Lernen incl. der Entwicklung entsprechender Anforderungsstandards • Entwicklung von Konzepten, Designs, Methodiken und Standards für datenintensive Forschung im Rahmen des Datenschutzes 	<p>Forschungsförderer, Forschung</p>	<p>Förderangebot für Workshops, interdisziplinäre Projekte, Konzept- und Methodenentwicklungen durch Forschungsförderer und Forschungseinrichtungen</p>

3. Verbesserung der Verfügbarkeit und Qualität von Gesundheitsdaten		
<p>Qualifizierung des Personals in der medizinischen Versorgung</p> <ul style="list-style-type: none"> • Ärztliches, pflegerisches Personal und Dokumentare durch Aus-, Fort- und Weiterbildung, Schulungen und geeignete technische Mittel befähigen, motivieren und incentivieren, einen Beitrag zur Generierung digitaler Versorgungsdaten zur Verbesserung der individuellen Patientenversorgung und Versorgungsqualität zu leisten. • Verstärkung der Ausbildung und des Einsatzes von Data (Curation) Scientists in der Versorgung • Die Clinician und Medical Scientists Programme durch geeignete Module ergänzen • Neue Masterstudiengänge für z.B. (Clinician) Data Scientists einrichten 	<p>Hochschulen, Universitätskliniken, Akteure der Selbstverwaltung, ambulanter und stationärer Sektor, Technologieentwickler</p>	<p>Einrichtung eines Steuerungskreises „Qualifizierung für Digitalisierung“ durch Hochschulen und Universitätskliniken</p> <p>unter Einbindung von MII- und DZG-Vertretern, Vertretern der SGKF der DFG, Vertretern des ambulanten und stationären Sektors, Akteuren der Selbstverwaltung, Fachverbänden, Gesundheitsfachschulen, Technologieentwicklern, des Fachverbands für Dokumentation und Informationsmanagement in der Medizin (DVMD), der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) sowie Bund und Ländern</p>
<p>Technologieentwicklung</p> <p>Entwicklung von Methodik, Software und interaktiven Technologien zur Nutzung digitalisierter Versorgungsdaten durch ärztliches und pflegerisches Personal</p>	<p>Forschung, Industrie, Forschungsförderer in Zusammenarbeit mit Gesundheitsversorgung</p>	<p>Gemeinsame Technologieentwicklung durch Forschung, Industrie u. Gesundheitsversorgung</p>
4. Förderung der Kultur des Datenteilens		
<p>Einfordern und Fördern von Datenteilen</p>	<p>Forschungsförderer</p>	<p>Gemeinsames Vorgehen von DFG und BMBF (und ggf. weiterer Förderer) zur Finanzierung von Datenteilen zur Herstellung einer stärkeren Verbindlichkeit in Bezug auf Umsetzung des Datenteilens und der FAIR Data Principles in Förderanträgen</p>
<p>Würdigung der Integration und Veröffentlichung von Daten analog zu einer wissenschaftlichen Leistung im akademischen Bewertungssystem</p>	<p>Forschungsförderer, Universitäten, AUFÉ, Journale, Fachgesellschaften</p>	<p>Austausch von Ideen, Ansätzen und Erfahrungen zur Verbreiterung des akademischen Bewertungssystems und gemeinsame Implementierung guter Konzepte. Erste Grundlagen hierfür werden aktuell in AG der DFG-Geschäftsstelle entwickelt:</p> <ul style="list-style-type: none"> • Umgestaltung der Publikationslisten in Anträgen und CVs, damit dort andere wissenschaftliche Ergebnisse, z. B. veröffentlichte Forschungsdaten, in der Zukunft häufiger genannt werden.

		Verstärkung der Verbindlichkeit der Ausführungen zum Forschungsdaten-Management in Projektanträgen und deren Begutachtung
Durchführen von Data Challenges zur Vertrauensbildung und Beschleunigung des Einsatzes von KI-Lösungen in der Medizin	Forschungsförderer	Förderangebot für Data Challenges
5. Ethische und rechtliche Aspekte der Datennutzung		
Erhöhung der Rechtssicherheit <ul style="list-style-type: none"> • Weiterentwicklung von Empfehlungen und Mustertexten zur Datenerhebung und -verarbeitung im Forschungskontext • Patientendatenschutzgesetz und § 287a SGB V in der Praxis verankern • Weitere Vereinheitlichung der Auslegung des Datenschutzrechts auf Landes-, Bundes- und Unionsebene prüfen • ggf. weiteren gesetzgeberischen Handlungsbedarf prüfen 	Stakeholder mit Datenschutzaufsichtsbehörden und medizinischen Ethikkommissionen	Befassung der MII unter Einbeziehung der Wissenschaft (u.a. Universitätsmedizin, DZG, AUFE, NFDI4Health, GHGA, TMF) zu folgenden Themen: <ul style="list-style-type: none"> • Handreichung zur Planung von Datenerhebung und -verarbeitung einschl. Mustertexte, Musternutzungsordnung für personenbezogene Daten • generische Geschäftsordnungen für Governance • Leitfaden für Use and Access Committees • Zugang zu Daten durch Industrie bzw. Wirtschaft
Forschung zu ethischen und rechtlichen Vorgaben für Decision Support Systeme und automatisierter Entscheidungsfindung, z. B. zum rechtlichen Spannungsverhältnis bzgl. KI-Anwendungen im Verhältnis zu den Regelungen der automatisierten Entscheidungsfindung	Forschungsförderer, Forschung	Förderangebot im Rahmen der Umsetzung der KI-Strategie der Bundesregierung
6. Patienteneinbindung bei der Forschung mit Gesundheitsdaten		
Stärkung der Partizipationsmöglichkeiten durch Entwicklung von <ul style="list-style-type: none"> • technischen und organisatorischen Instrumenten für Partizipation • praxistauglichen Instrumenten für ein studien- und standortübergreifendes Consent Management 	Forschung, Förderer	Modellhafte Bearbeitung in DZG-übergreifender Arbeitsgruppe und Verbreitung der Lösungsansätze unter Einbeziehung von Vertretern der MII, AUFE und Patientenorganisationen

Tabelle 2 Exkurs: Empfehlungen zu zulassungs- und haftungsrechtlichen Aspekten bei der Implementierung KI-basierter Anwendungen in der Medizin und zu perspektivischen Möglichkeiten einer Datenüberlassung

Empfehlung	Akteure	Umsetzungsvorschlag
1. Zulassungs- und haftungsrechtliche Aspekte datengestützter Verfahren in der Medizin		
Herstellung innovationsaffiner Marktzugangspfade für Medizinprodukte-Software	BMG, EU-Kommission	Überprüfung des derzeitigen EU-Rechtsrahmens zu Medizinprodukten auf notwendige Anpassungen im Hinblick auf Besonderheiten der kontinuierlich lernenden und adaptiven KI-basierten Software
Forschung zu den verschiedenen haftungsrelevanten Bereichen von KI-Anwendungen (z. B. Arzthaftungsrecht, Produkthaftungsrecht) und den bereichsübergreifenden rechtlichen Zusammenhängen	Forschungsförderer, Forschung	Förderangebot im Rahmen der Umsetzung der KI-Strategie der Bundesregierung
2. Perspektivische Empfehlung: Ein Plädoyer für die Datenüberlassung für die Forschung		
Plädoyer für eine dauerhafte, institutionalisierte Überlassung von Patientendaten an die Forschung, verbunden mit einer möglichst niederschweligen Möglichkeit zum Opt-out und transparenter Kommunikation	Forschung, Versorgung, BMG, BMBF, Politik	Konzeptentwicklung, technische, rechtliche und ethische Bewertung, anschließende Pilotierung z.B. in MII, DZG (u. a. integriert in vorgeschlagener Arbeitsgruppe zu Punkt 5)

I. Ziel

Ziel des Strategiepapiers ist es, unter Berücksichtigung existierender Initiativen und Empfehlungen [1, 2, 3] (s. Kap. II.) aufzuzeigen, welche zusätzlichen spezifischen Schritte die im Forum Gesundheitsforschung vertretenen Akteure gemeinsam ergreifen sollten, um den Zugang und die Nutzbarkeit von Daten aus Forschung und Versorgung für KI-Entwicklungen und datengetriebene Gesundheitsforschung im Allgemeinen zu verbessern. Die aktuelle Corona-Pandemie zeigt, wie schnell Forschung Lösungsansätze erarbeiten kann, wenn Ergebnisse frühzeitig zur wissenschaftlichen Diskussion bereitgestellt und Daten schnell (u. a. auch der Industrie) verfügbar gemacht werden. Das Strategiepapier greift diese Erfahrungen auf und entwickelt mit Blick über die Corona-Pandemie hinaus Handlungsempfehlungen zu Problemkomplexen bei der Nutzbarmachung großer Datenmengen für KI-Entwicklungen und datengetriebene Gesundheitsforschung im Allgemeinen, die bisher noch nicht ausreichend angegangen werden. Die Empfehlungen sind für Bereiche der Gesundheitsforschung, die große Datenmengen nutzen, z. B. die Omics-Forschung, von besonderer Bedeutung.

In dem modular aufgebauten Papier werden hierzu:

1. die Ausgangslage und wichtige bestehende Initiativen und Maßnahmen beschrieben und
2. zeitnah umsetzbare Handlungsempfehlungen für einzelne Akteure des Forums Gesundheitsforschung entwickelt.

Für die Verbesserung der Verfügbarkeit von Daten aus Versorgung und Forschung für datengetriebene Gesundheitsforschung sind auch Maßnahmen anderer Akteure unerlässlich. In einem Exkurs werden daher weitere Aspekte betrachtet, deren Umsetzung über die im Forum Gesundheitsforschung vertretenen Akteure hinausgehen. Sie beziehen sich auf zulassungs- und haftungsrechtliche Fragen zur Implementierung von KI-Entwicklungen und perspektivische Ansätze zur Vereinfachung der Datenüberlassung.

II. Ausgangslage

1. Voraussetzungen für KI-Entwicklungen und datengetriebene Forschung in der Medizin

Von der Verfügbarkeit umfangreicher und qualitativ hochwertiger Daten hängt der Erfolg medizinischer KI-Anwendungen und datengetriebener medizinischer Forschung maßgeblich ab. Nur ein Bruchteil der heute vorhandenen Routinedaten im Gesundheitswesen ist überhaupt für datengetriebene Forschung und die Entwicklung und Überprüfung intelligenter Systeme zugänglich. Die Umsetzung der elektronischen Patientenakte wird maßgeblichen Einfluss darauf haben, wie Individualdaten von Patienten bzw. Versicherten im Versorgungsalltag zukünftig genutzt und im Weiteren auch für die Forschung zur Verfügung gestellt werden können.

Bereits bei der Erhebung der Daten gibt es zahlreiche Herausforderungen, denen begegnet werden muss:

- Daten sind häufig unvollständig, unstrukturiert dokumentiert und annotiert, grundsätzlich von variabler Qualität und bisweilen fehlerbehaftet. Der präzisen Annotation relevanter Parameter sowie der sorgfältigen Erfassung der zugehörigen Metadaten kommt eine zentrale Bedeutung zu. Für die Datenaufbereitung (Data Cleaning) müssen in der Zu-

kunft erhebliche Ressourcen aufgewendet werden. Während für das Training von KI-Modulen oftmals auch unvollständige oder sogar fehlerbehaftete Daten ausreichen können, muss die Datenqualität für die Validierung entwickelter KI-Systeme den jeweils aktuellen höchstmöglichen Ansprüchen genügen. Für die Anwendung im Versorgungsalltag wird so sichergestellt, dass Systeme nicht unter falschen Vorzeichen zugelassen werden. Ähnliches gilt für die Entwicklungs- bzw. Validierungsphase datengetriebener medizinischer Forschung.

- Schlussfolgerungen und medizinische Diagnosen, die als Endpunkte für das maschinelle Lernen fungieren, können zu vereinfachend, auf Vermutungen basierend oder schlicht falsch sein.
- Die Definition eines Behandlungserfolges ist ein komplexes Zusammenspiel aus Informationen zu Überleben, Abheilung der Erkrankung, Rückfallfreiheit, Aufenthaltsdauer, Lebensqualität, Toxizität, Kosten und weiteren indikationsspezifischen Faktoren.

Eine einfache Kategorisierung eines Krankheitsverlaufs in "gut" oder "schlecht" für das algorithmische Lernen oder datengetriebene Forschung ist in vielen realen Behandlungsfällen unmöglich. Die Festlegung geeigneter Endpunkte sollte zudem stets den Stand des Wissens repräsentieren und idealerweise in Zusammenarbeit mit den entsprechenden Fachgesellschaften erfolgen.

Wenn mit KI- oder anderen Methoden komplexe Zusammenhänge analysiert werden sollen, muss mit großen Stichprobenzahlen (d. h. einer repräsentativen Patientengruppe) gearbeitet werden. Wenn dies nicht geschieht, besteht die Gefahr, falsche und nicht generalisierbare Schlussfolgerungen aus den Daten zu ziehen. Falsch angewendet könnten sogar diskriminierende bzw. „rassistische“ Ergebnisse entstehen. Zum Beispiel könnte ein KI-Modell, welches überwiegend mit Daten von Patienten gleicher Ethnie trainiert wird, in unbeabsichtigter Weise systematisch falsche Vorhersagen für Patienten anderer Herkunft machen. KI-Methoden müssen gegen solche unbeabsichtigten oder auch potentiell böswilligen Verzerrungen robust sein – dies insbesondere vor dem Hintergrund, dass medizinische Forschungsdaten oftmals unbalanciert sind, d. h. nicht alle Subgruppen und Varianten in identischer Anzahl enthalten sind.

Entscheidungen der KI müssen darüber hinaus nachvollziehbar und dürfen nicht manipulierbar sein. Insbesondere ist für KI-Produkte vollständige Transparenz zu fordern, so dass ein KI-Modul keine Black Box darstellt, sondern vielmehr Auskunft darüber gibt, aufgrund welcher Daten-Zusammenhänge oder etwa welcher genauen Bildregionen eine Handlungsempfehlung oder ein Klassifikationsergebnis entstanden ist. Um diesen Qualitätsstandard zu erreichen, besteht weiterer Forschungs- und Entwicklungsbedarf⁵.

Auch bei ausreichender Datenmenge sind KI-Systeme in ihrer Natur retrospektiv, da sie mit Daten vergangener Fälle trainiert werden. Es müssen klare und transparente Prozesse entwickelt werden, wie medizinische Innovationen (z. B. Markteintritt eines neuen Medikamentes) in prospektive Entscheidungen der KI integriert werden können, bevor eine entsprechende Datenbasis in die fortlaufende Analyse aufgenommen werden kann.

⁵ KI-Systeme, die Behandlungsentscheidungen treffen, können weitreichende Auswirkungen auf das Behandlungsergebnis von Patienten haben. Häufig werden KI-Systeme mit retrospektiv verfügbaren Daten gegen ärztliche Behandlungsentscheidungen verglichen. Dabei wird übersehen, dass letztere im Zeitalter der evidenzbasierten Medizin durch Publikationen und Leitlinien stark eingeeengt sind. Die Performance von KI-Systemen muss daher einerseits gegen medizinische Behandlungsstandards und andererseits gegen medizinische Expertise evaluiert werden.

Für die meisten medizinischen Fragestellungen sind Datenmengen notwendig, die an einem einzelnen Standort nicht in ausreichender Anzahl und Vielfalt erhoben werden können. Darüber hinaus müssen präklinische, klinische und Versorgungsforschungsdaten verknüpft werden, damit der Translationskreislauf befördert wird und die für KI-Anwendungen und datengetriebene medizinische Forschung erforderlichen komplexen und vielschichtigen Zusammenhänge abgebildet werden können. Wenn mit Hilfe von KI-Anwendungen und datengetriebener Forschung das Verständnis von Krankheiten verbessert werden soll – idealerweise auf der Grundlage ihrer molekularen Krankheitsmechanismen – muss in größeren interdisziplinären und standortübergreifenden Verbänden gemeinsam geforscht werden. So erlaubt die Kombination von Daten über Regionen- und Landesgrenzen hinweg, neben einem Gewinn an statistischer Power auch die Einflüsse regionaler Faktoren auf Patienten-Kohorten (z. B. Regionen-spezifische Ernährungsweise oder anatomische Varianten) abzubilden und zu analysieren. Insofern erfordert die Entwicklung von KI-Anwendungen und datengetriebener Forschung in der Medizin die Zurverfügungstellung und gemeinsame Nutzung qualitativ hochwertiger Daten aus Versorgung und Forschung über Standort- und Disziplinergrenzen hinweg. Um dies zu erreichen, wurde bereits eine Reihe von Initiativen angestoßen. Das folgende Kapitel beschreibt sowohl die Vorarbeiten, die hierzu bereits geleistet wurden, als auch die noch bestehenden Herausforderungen.

2. Initiativen zur Nutzung gesundheitsbezogener digitaler Daten für die Forschung

Um medizinische Daten für die Forschung deutschlandweit nutzbar zu machen, wurden und werden verschiedene Initiativen gestartet. Sowohl national als auch international ist die Nutzbarmachung digitaler Gesundheitsdaten ein aktuell hoch dynamisches Feld. Im Folgenden werden einige für die Nutzung digitaler Daten in der Gesundheitsforschung besonders relevante konkrete Initiativen in Deutschland aufgelistet:

- Medizininformatik-Initiative (MII): Patientenbezogene klinische Primärdaten aus dem Behandlungskontext in Unikliniken sowie darüber hinaus werden in Datenintegrationszentren für die Nutzung durch Forschung und Versorgung verfügbar gemacht. In den Datenintegrationszentren sind auch Datensätze aus der klinischen Forschung verfügbar. Die Datenhaltung selbst erfolgt lokal, nah am Versorgungsgeschehen. Die MII hat einen Kerndatensatz von Daten aus der Patientenversorgung abgestimmt und damit begonnen, hierfür Daten anhand internationaler Standards zu harmonisieren und in den Datenintegrationszentren der Universitätskliniken für Forschungsprojekte bereitzustellen. In der AG „Data Sharing“ wurden Prozessbeschreibungen, Datennutzungsordnungen, -anträge und -verträge für ein standortübergreifendes Data Sharing auf Basis der FAIR-Prinzipien abgestimmt. Es wird eine Antrags- und Registerstelle (ZARS) als zentraler Anlaufpunkt für die Forschung aufgebaut. Die AG „Consent“ entwickelte einen Mustertext für einen Broad Consent und stimmte diesen mit dem Arbeitskreis der deutschen Ethikkommissionen und dem Arbeitskreis der Datenschützer ab. In Anwendungsfällen werden beispielhaft IT-Lösungen entwickelt, die im Klinikalltag einen Patientennutzen aufzeigen. Die TMF koordiniert die MII und trägt damit wesentlich zur Harmonisierung der Prozesse bei.
- Deutsche Zentren der Gesundheitsforschung (DZG): In den DZG sind große krankheitsbezogene qualitätsgesicherte digitale Datenbestände etabliert worden. Es handelt sich dabei um experimentelle, epidemiologische und klinische Forschungsdaten. Es wurden Plattformen etabliert, die eine standardisierte Erfassung von Forschungs-

daten befördern und es ermöglichen, Daten für standortübergreifende Forschungsprojekte zur Verfügung zu stellen.

Die sechs DZG haben unterschiedliche Systeme etabliert, die es erlauben, bereits vorhandene Datensätze an den Mitgliedseinrichtungen/Partner Sites innerhalb der betreffenden DZG transparent zu machen und unter definierten Bedingungen auf Daten und Proben zuzugreifen und sie zu teilen. Ziel ist es, wie in einigen DZG bereits umgesetzt, präklinische und klinische Daten sowie Bioproben nach den FAIR Prinzipien zwischen und auch außerhalb der DZG sichtbar und zugänglich zu machen. Dazu wurde eine DZG-Arbeitsgruppe etabliert, die sowohl technische als auch rechtliche Lösungen des Datenteilens erarbeitet und zeitnah implementieren wird. Das Konzept dafür wurde bereits im Rahmen eines gemeinsamen Antrags mit der Medizin-Informatik-Initiative für die Nationale Forschungsdateninfrastruktur (NFDI4MED, 10/2019) erarbeitet. Angesichts der vorläufigen Zurückstellung einer Förderung in diesem Programm soll das Vorhaben in reduziertem Umfang (zunächst DZG-weit) mit DZG-Mitteln aus dem Budget für DZG-übergeordnete Strukturen verfolgt werden.

- NFDI4Health adressiert als gefördertes NFDI-Projekt die Erschließung strukturierter personenbezogener Gesundheitsdaten aus epidemiologischen Kohorten (insb. der NAKO-Gesundheitsstudie), klinischen Studien (insbesondere des Netzwerks der Koordinierungszentren für Klinische Studien, KKS), Registern, administrativen Gesundheitsdatenbanken sowie Public Health-Surveillance. Es wird ein zentraler Such-Hub etabliert, um standardisierte Daten und ihre Metadaten auffindbar und zugänglich zu machen. Hierzu soll ein Datenschutz-konformes Record Linkage von Primärdaten sowie von Primär- mit Sekundärdaten geschaffen werden. Darüber hinaus werden verteilte Datenanalyseumgebungen einschließlich KI-Anwendungen und Strategien erprobt um synthetische Datensätze zu erzeugen.
- Das NFDI-Projekt GHGA wird eine Infrastruktur für die Archivierung, Verteilung und Analyse von bereits erhobenen humanen Hochdurchsatztechnologie-Daten aufbauen. Die Daten werden in verteilten Datenhubs gespeichert, die an großen datengenerierenden Sequenzierzentren (insbesondere den Zentren des NGS-Kompetenznetzwerks der DFG) aufgebaut werden. Die Interoperabilität der Daten und eine verteilte Datenanalyse sind gewährleistet. Die Metadaten der einzelnen Projekte werden in standardisierter Form öffentlich verfügbar gemacht. In einem elektronischen Antragsprozess soll identifizierbaren Nutzern nach Prüfung der rechtlichen und ethischen Voraussetzungen für eine beschränkte Zeit Zugriff auf die gespeicherten Daten gewährt werden.
- „Medical Data Space“ und das Leitprojekt „MED²ICIN der Fraunhofer-Gesellschaft: Der „Medical Data Space“ ist ein Gesamtkonzept für eine vertrauensvolle, selbstbestimmte und nachvollziehbare Datennutzung und bietet eine Infrastruktur für den sicheren Austausch von Gesundheitsdaten zwischen allen Akteuren des Gesundheitssystems. Die Architektur folgt der Grundidee einer informationellen Selbstbestimmung der Patienten. Durch ein Framework für Datennutzungskontrolle wird eine feingranulare und widerrufbare Steuerung der Nutzungsfreigaben ermöglicht und gleichzeitig die Datennutzung nachvollziehbar gemacht, so dass Urheber und Eigentümer der Daten angemessen an realisierter Wertschöpfung beteiligt werden können. Diese Prinzipien ermöglichen - unter Wahrung von Patienteninteressen – die Nutzung medizinischer Daten durch die Gesundheitswirtschaft.

Mit dem Leitprojekt MED²ICIN soll ein digitales Patientenmodell durch Fusionierung bisher verteilt und teils unstrukturiert vorliegender hochdimensionaler Daten von Individuen zu einem digitalen Abbild geschaffen werden.

- Helmholtz Data Federation (HDF): Die HDF ist eine föderierte Forschungsdateninfrastruktur der Helmholtz Gemeinschaft. Die HDF basiert auf den drei Kernelementen Software für das Forschungsdatenmanagement, Benutzerunterstützung und Speicher- und Analyse-Hardware. Partner in der ersten Phase sind sechs Helmholtz-Zentren aus fünf Forschungsbereichen. Die HDF ist offen für Nutzer im deutschen Wissenschaftssystem und wird durch internationale Verbindungen kompatibel mit der zukünftigen European Open Science Cloud (EOSC) sein.
- Netzwerk Universitätsmedizin (NUM) und LEOSS: Das zur Bewältigung der aktuellen Corona-Pandemie etablierte Netzwerk wird eine Forschungsdatenplattform zur Erfassung von Behandlungsdaten von COVID-19-Patienten in Deutschland aufsetzen. Dabei sollen Strukturen der DZG und MII genutzt und mit dem europäischen Fallregister für COVID-19-Patienten (LEOSS) zusammengearbeitet werden.
- de.NBI-Netzwerk: Das Netzwerk bietet ein breites Portfolio an Ressourcen für die deutsche und internationale Gemeinschaft der Lebenswissenschaften. Dazu gehören hauptsächlich Datenbanken, bioinformatische Software und die Bereitstellung von Rechenkapazität durch ein Verbund-Cloud-System. de.NBI entwickelt und pflegt bereits fünf große Datenbanken (SILVA, PANGEA, BacDive, ProteinPlus und BRENDA).
- Weiterentwicklung des Datentransparenzverfahrens: Mit dem Digitale-Versorgung-Gesetz (DVG) wurden 2019 bestehende gesetzliche Regelungen zur Bereitstellung von Sozialdaten der Krankenkassen zu Forschungszwecken erweitert und die Datenaufbereitungsstelle zu einem Forschungsdatenzentrum weiterentwickelt. Das Forschungsdatenzentrum soll sowohl die Aufbereitung und Bereitstellung von aggregierten, anonymisierten Abrechnungsdaten der Krankenkassen, als auch – sofern die Einwilligung zur Datenfreigabe der Versicherten vorliegt - die Bereitstellung von Gesundheitsdaten aus elektronischen Patientenakten zu Forschungszwecken übernehmen. Forschenden soll auf Antrag die Auswertung anonymisierter Ergebnismengen ermöglicht oder, sofern für die Bearbeitung des Forschungsvorhabens erforderlich, in einem geschützten Raum Zugang zu aufbereiteten pseudonymisierten Einzeldatensätzen gewährt werden.
- Genominitiative genomDE: Diese vom Bundesministerium für Gesundheit (BMG) geförderte Initiative plant die Einführung der medizinischen Genomsequenzierung in die Regelversorgung. Die Daten sollen in einem gesicherten, noch aufzubauenden Datenbanksystem zwischen Versorgung und Forschung vernetzt und zugänglich gemacht werden. Erklärtermaßen werden MII und NFDI-GHGA dabei eine maßgebliche Rolle spielen.
- German Biobank Alliance: In mehreren aufeinander abgestimmten BMBF-Fördermaßnahmen der letzten zehn Jahre wurden deutsche Biobanken aufgebaut und miteinander vernetzt, um qualitativ hochwertige und gut annotierte Bioproben für die Forschung sichtbar und verfügbar zu machen. Der entstandene „Sample Locator“ erlaubt eine föderierte Suche über die individuellen Probenbestände von aktuell 14 Bioban-

ken. Das daran angeschlossene „Negotiator“-Vermittlungssystem bearbeitet in Kooperation mit dem europäischen BBMRI-ERIC durchschnittlich etwa 120 Projekte pro Jahr.

- Verschiedene Initiativen der Bundesländer: Als beispielhafte länderseitige Initiativen sollen hier DigiMed Bayern und die Zentren der personalisierten Medizin in Baden-Württemberg genannt werden.

Neben technologischen Infrastrukturkomponenten (ID-Management, Pseudonymisierungsdienste, Record Linkage, Einwilligungsmanagement, Datenintegration und Datenharmonisierung, Forschungsdatenrepositorien) werden in den o. g. Initiativen vor allem auch organisatorische Rahmenbedingungen geschaffen (Mustertext Einwilligungserklärung, einheitliche Datennutzungsordnung, Data Use & Access Committees, einheitliche Datenfreigabeprozesse, Feedbackprozesse in die Kliniken hinein zur Verbesserung der Datenqualität, deutschlandweite Antragsportale etc.). All dies befindet sich im Aufbau (und teilweise auch schon jahrelanger Nutzung) und stellt damit eine exzellente Grundlage für den Zugang zu möglichst breiten und qualitätsgesicherten Daten für KI-Anwendungen und datengetriebene Forschung dar.

Die o. g. Initiativen leisten zentrale Beiträge zur Nutzung medizinischer digitaler Daten und stehen vor gemeinsamen Herausforderungen; sie tun dies jedoch bisher unabhängig voneinander. Die bisher fehlende Vernetzung der verschiedenen Initiativen zur Nutzbarmachung digitaler Gesundheitsdaten schränkt zum einen eine nachhaltige Wiederverwendung bereits entwickelter Methoden ein. Zum anderen bleiben der Forschung Anwendungsfälle verschlossen, die Daten der verschiedenen Welten zusammenbringen – also etwa Studiendaten eines Patienten (aus einer DZG-Datenhaltung) mit seinen Daten aus der Routineversorgung (aus dem jeweiligen MII-Datenintegrationszentrum). Genau dort – nämlich in der gemeinsamen Analyse nicht nur möglichst vieler Datensätze, sondern solcher aus verschiedenen Quellen, Disziplinen und Typen – liegt das Potenzial von „Big Data“.

III. Handlungsempfehlungen

1. Weiterentwicklung und Bündelung von Initiativen zur Verbesserung des Datenzugangs

Wissenschaftsinitiierte Forschungsprojekte mit digitalen Daten entstehen weitgehend in einem von zwei Szenarien: (i) Prospektive Projekte mit ausreichender Budgetierung für eine zumindest teilweise Erfassung strukturierter Daten, typischerweise in nicht-öffentlichen Datenhaltungen und (ii) Projekte auf Basis von Bestandsdaten (retrospektiv), häufig aus klinischen Datenbanken mit ebenfalls teilweise manuell strukturierten Daten, aber zunehmend auch mittels Integration von Daten aus den Klinikinformationssystemen. Beide spielen eine wichtige Rolle für die wissenschaftliche Landschaft, haben jedoch unterschiedliche Struktur- und Förderbedürfnisse.

Prospektive Studien

Gut finanzierte prospektive Studien, Register, Kohorten und Biobanken profitieren aktuell wenig von den bisher geschaffenen Infrastrukturen (siehe Kapitel II.2). Es fehlt an einheitlichen Strukturen für die strukturierte Erfassung von Daten, die nicht Teil der Patientenakte sind, als *Electronic Data Capture* (EDC). Umfassende technische Lösungen hierfür sind nur kommerziell oder unter ungünstigen Lizenzbedingungen erhältlich und verursachen einen erhebli-

chen finanziellen Mehraufwand. Dazu verfügen viele gängige EDC über keine etablierte Methodik der semantischen und syntaktischen Interoperabilität. Die Einbindung von und Verknüpfung mit klinischen Datensystemen ist nur mit erheblichem Entwicklungsaufwand möglich. Hieraus leitet sich als Konsequenz wiederum die Schwierigkeit der Verknüpfung der in dieser Art dokumentierten wertvollen Daten mit anderen Studien und Datenbanken ab. Da ein solcher Abgleich mit anderen Studien häufig ohnehin nicht geplant ist, führt dies dazu, dass Datenelemente typischerweise in einem proprietären, oft arbiträren Format abgelegt werden, was die Auffindbarkeit in Metaregistern erschwert.

Aus diesen Gründen ist für prospektive Studien zu fordern, dass hochwertige EDC Systeme mit Unterstützung für Ontologien, Anbindung an Metadatenregister und semantischer und syntaktischer Interoperabilität mit den DIZ der MII zur Verfügung stehen und deren Implementierung unterstützt wird. Voraussetzung hierfür ist die Entwicklung einheitlicher Datenmodelle für präklinische und klinische Forschungsdaten, die auf international einheitlichen syntaktischen und semantischen Standards basieren.

Retrospektive Studien

Die Gruppe der retrospektiven Studien, insbesondere solchen mit geringem oder keinem Budget, hat besonders hohes Potenzial, von der fortschreitenden Digitalisierung zu profitieren. Multizentrische Datenabfragen bieten dafür ein großes Potential. Die DIZ der MII können hier Brücken bauen, um Studien multizentrisch und mit großen Datensätzen durchzuführen, was erlaubt, mehr Hypothesen zu prüfen und dabei zu robusteren Ergebnissen zu kommen.

Abzuwarten bleibt, inwieweit die Verfahren der DIZ automatisiert sein werden und welcher Personal- und sonstiger Kostenaufwand mit der Durchführung von Datenabfragen im weiteren Verlauf verbunden sein werden. Da für retrospektive Analysen Projektmittel in der aktuellen akademischen Förderlandschaft schwierig zu erhalten sind, steht und fällt der Nutzen der DIZ für die Breite der klinischen Wissenschaftler und Wissenschaftlerinnen mit der Niederschwelligkeit der Zugangsmöglichkeiten. Aus diesem Grunde sind insbesondere einfache Zugangswege mit klaren und niederschweligen Zugangsbedingungen zu fordern.

Weiterentwicklung bestehender Initiativen (DZG, MII)

Es sollten z. B. vorhandene Einwilligungserklärungen der in den DZG bisher erhobenen Datenbestände im Hinblick auf die Möglichkeit, diese Daten auch über die bisherigen Erhebungsziele hinaus verwenden zu dürfen, analysiert werden. Übergreifend über alle DZG ist ein einheitliches Datenbank-gestütztes Einwilligungsmanagement zu etablieren. Dieses ist die Grundlage für eine Datenlandkarte inkl. gegebener Erlaubnistatbestände für die Datennutzung durch die Forschung. Der DZG übergreifende Budgetanteil sollte erhöht werden, um präklinische und klinische Daten sowie Bioproben nach den FAIR-Kriterien zwischen und auch außerhalb der DZG sichtbar und zugänglich zu machen.

Der Aufbau der DIZ in Deutschland wird langfristig die Methoden und Anwendungen der KI im Hinblick auf die Gesundheitsversorgung voranbringen. Allerdings sind die Datenbestände, die bisher im Rahmen der MII in den DIZ bereitgestellt werden, begrenzt auf die MII Kerndatensatzbasismodule sowie einzelne krankheitsspezifische Bereiche gemäß der in den Konsortien gewählten Use Cases. Je breiter der Gesamtdatenbestand in den DIZ, umso besser lassen sich darauf KI-basierte Anwendungen entwickeln und datengetriebene Forschung durchführen. Deshalb ist der kontinuierliche Ausbau der DIZ-Datenbestände und deren Sichtbarkeit durch ein geeignetes Zugangportal (ZARS) sicherzustellen. Forscher sollten dazu motiviert werden, die in einem Projektrahmen generierten Daten nicht nur einmalig für

das eigene Projekt zu nutzen, sondern diese darüber hinaus auch in die etablierten DIZ einfließen zu lassen.

Outcome-Daten für das Entwickeln von KI-Modellen/Algorithmen fallen oft auch im niedergelassenen Bereich an. Das Einbeziehen der Datenbestände aus diesem Bereich und die Integration dieser Daten in die DIZ muss daher ebenfalls gefördert werden. Dieses Vorgehen wird aktuell durch die BMBF-Bekanntmachung "Digitale FortschrittsHubs Gesundheit" im Förderkonzept Medizininformatik unterstützt.

Konvergenzprozess zu einer integrierten Gesamtlösung

Die Konzepte und Technologien zum Auffinden von und zum Zugang zu vorhandenen Daten, die durch die DZG und die MII bisher unabhängig voneinander entstanden sind, müssen technisch aneinander angeglichen werden. Darüber hinaus ist zu klären, wie diese mit den zukünftigen Systemen der NFDI-Projekte NFDI4Health und GHGA interagieren. Diese Aktivitäten müssen auch sorgfältig mit der Entwicklung und dem Einsatz einer forschungskompatiblen elektronischen Patientenakte abgestimmt werden.

Das in der MII entwickelte Konzept eines deutschlandweiten Zugangsportals zu den an den Universitätskliniken in DIZ verfügbar gemachten Datenbeständen, die in den DZG bereits funktionierende Implementierung für onkologische Datenbestände und die in den Biobanken bereits breit genutzten Suchdienste für Bioproben müssen zu einem von Förderlinien unabhängigen Portal zusammengeführt werden. Dadurch sollen beispielsweise die in den DZG erhobenen und sekundär verwendbaren Datenbestände gemeinsam mit Bioproben und Routedaten auffindbar und nutzbar gemacht werden. Darüber hinaus ist zu klären, wie ein solches übergreifendes Zugangsportal mit den bei NFDI4Health und GHGA zu entwickelnden Such-Hubs und Zugangsportalen zu einem synergistischen Gesamtsystem zusammengebracht werden kann. Auch nach Auslaufen der genannten Fördermaßnahmen muss durch eine ausreichende Strukturfinanzierung die Verwendung der geschaffenen Plattformen durch die Breite der Wissenschaft sichergestellt werden, wozu ein möglichst kostenfreier Zugriff und kurze Bearbeitungszeiten zählen.

Für die in den DZG Krankheitsbild-bezogen, vorhandenen großen Datenbestände sind Anreize zu schaffen, um diese Datenbestände (soweit es die Einwilligungen erlauben) mit den in den DIZ entstehenden Daten zu integrieren. Die Datenbestände von DIZ und DZG sollten auch mit den Strukturen von NFDI4Health und GHGA kompatibel und über diese auffindbar sein, sofern es sich um klinische bzw. epidemiologische Studien oder Omics-Daten handelt. Zusätzlich sollten noch gezielt Förderprogramme angestoßen werden, die mit spezifischen Anwendungsbeispielen auf die Integration von DZG- und DIZ-Datenbeständen fokussieren. Hierbei sollte insbesondere an entsprechenden Analyseszenarien der Mehrwert der damit neu generierten Datenbestände demonstriert werden. Es sollte auch die Integration mit NFDI4Health und GHGA berücksichtigt werden, sofern die jeweiligen Anwendungsbeispiele für diese Forschungsinfrastrukturen relevant sind. Schritt für Schritt sollte auch auf die Integration von Forschungs- und Behandlungsdaten aus der Psychologie, Daten aus weiteren Fachgebieten wie z. B. Nano-Wissenschaften, synthetische Biologie sowie weiteren einschlägigen Forschungsinitiativen wie z. B. Forschungspraxen hingewirkt werden.

Da sich einige der im Kapitel II.2 genannten Initiativen noch im Aufbau befinden, ist derzeit in Teilen unklar, wie die verschiedenen Initiativen synergistisch miteinander interagieren, aufeinander aufbauen und sich gegenseitig ergänzen können. Sowohl auf forschungspolitischer als auch auf technischer Ebene gibt es offene Fragen. Unstrittig ist, dass die gemeinsame

Entwicklung von Standards und die Integration von Datenbeständen zwischen verschiedenen Initiativen und für die Medizin relevanten Fachgebieten weiter vorangebracht werden sollte. Auch um im internationalen Wettbewerb bestehen zu können, ist es entscheidend, die verschiedenen Akteure in Deutschland zusammenzubringen.

Es sollte eine Strategie dazu entwickelt werden, wie das in den verschiedenen Initiativen generierte Wissen und die Infrastrukturen zusammengeführt werden können. Ziel sollte der Aufbau einer interoperablen, nutzerfreundlichen und nachhaltigen Struktur für die Integration und den Austausch von Daten aus den Bereichen der biomedizinischen Forschung und klinischen Versorgung sein. Hierzu gehören u. a. übergreifende interdisziplinäre und fachspezifische Dienste, Schnittstellen und abgestimmte Prozesse und Standards für die biomedizinische Forschungsgemeinschaft. Erforderlich ist eine Stärkung existierender bzw. Einrichtung neuer Abstimmungsstrukturen. Hierzu sollten unterstützt von BMBF und DFG unter Berücksichtigung der Funktionen und Aufgaben der NFDI die relevanten Stakeholder zusammengebracht werden, um den notwendigen Konvergenzprozess zu einer integrierten Gesamtlösung zu entwickeln, die auf Zusammenarbeit und Kompatibilität ausgerichtet ist. Das BMG und die Gematik sollte anlassbezogen eingebunden werden, um die Abstimmung mit der elektronischen Patientenakte sicherzustellen.

Darüber hinaus sollten von BMBF Pilotprojekte zur Konvergenz von Dateninfrastrukturen gefördert werden. Ein wichtiges Projekt ist dabei die Integration von DZG- und DIZ-Datenbeständen der MII in spezifischen Anwendungsfeldern.

Die o. g. Empfehlungen thematisieren die Nutzbarmachung digitaler Daten für KI-Entwicklungen und datengetriebene Forschung im nationalen Kontext. Da für die Entwicklung von KI-Anwendungen interdisziplinäre Forschung an großen Patientenkohorten erforderlich ist, scheint perspektivisch der durch die Datenethikkommission (DEK) unterstützte Aufbau eines europäischen Forschungsdatenraumes für den Bereich Datennutzung für KI-Anwendungen zielführend zu sein [11]. Gleiches gilt für die durch die Bundesregierung vorangetriebene Weichenstellung für den Aufbau des Europäischen Gesundheitsdatenraums, einschließlich der bereits angestoßenen Unterstützung für EU-weite Codes of Conduct für die Sekundärnutzung von Gesundheitsdaten und die in Vorbereitung befindliche Joint Action „Towards the European Health Data Space“⁶. Hierbei ist auch zu diskutieren, welche Bedeutung der European Open Science Cloud in diesem Zusammenhang zukommen kann, die verschiedene europäische Datendienste und Informationsstrukturen verbinden soll.

2. Entwicklung technischer Lösungen zur Datenverknüpfung, Datensicherheit und Datenschutz

Generell stehen die Ziele der KI und datengetriebener Forschung nicht im Widerspruch zu einem Schutz der Privatsphäre. Um die nötigen Verknüpfungen auch über verschiedene Datenquellen hinweg herzustellen, sind technische Möglichkeiten zur datenschutzgerechten Verknüpfung personenbezogener Daten über Einrichtungsgrenzen hinweg (Privacy-Preserving Record Linkage) erforderlich, die von einer breiten Einwilligung (Broad Consent) als Rechtsgrundlage gedeckt sein sollten. Es sollten technisch-organisatorische Sicherungsmaßnahmen/Governance-Strukturen entwickelt werden, um die Transparenz der Datenverarbeitung und damit das Selbstbestimmungsrecht des Einzelnen abzusichern und zugleich

⁶ https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/E/EU2020/Presidency_Summary_Informal_Meeting_of_Health_Ministers-16_July_2020.pdf

Risiken zu minimieren. Bei all diesen Schritten sollte auf nationale bzw. internationale Standards zurückgegriffen werden. Wo es solche (noch) nicht gibt – was aufgrund der sich stetig weiterentwickelnden Fachdomänen durchaus denkbar ist – sollten Mittel zur nachhaltigen Entwicklung und Verstetigung neuer Standards, die auch über Projekte und Initiativen hinweg Verwendung finden können, bereitgestellt werden.

Aktuelle und etablierte Methoden der Datenanonymisierung, wie Aggregation und Verrauschen, oder auch neuere Ansätze, wie die „differential privacy“, eliminieren gezielt Informationen über einzelne Individuen und bieten so statistisch fundierte Garantien für einen Schutz der Privatsphäre. Die neueren Entwicklungen der KI sind dabei von Vorteil: die Erweiterung vieler Methoden der KI und des maschinellen Lernens durch Formalismen für Privacy-Definitionen kann inzwischen besagte Garantien bieten. Diese Methoden sind beim Schutz der Privatsphäre den etablierten Anonymisierungsmethoden überlegen und sollten daher bevorzugt Anwendung finden.

Die Wissenschaft muss – auch wegen ihrer Verantwortung gegenüber der sie finanzierenden Gesellschaft – Alternativen zum Datenaustausch der klassischen Art (weiter-)entwickeln. Hier sollten Methoden des föderierten Lernens Anwendung finden, bei dem KI-Methoden dezentral über mehrere Standorte hinweg lernen können. Vorarbeiten zur dezentralen Datenerhaltung und der föderierten Auswertung von Daten über die dezentralen Standorte hinweg wurden bereits in verschiedenen DZG sowie der MII umgesetzt. Auf solchen dezentralen Datenbeständen lassen sich ebenfalls KI-Modelle unter Datenschutzwahrung trainieren, etwa indem vortrainierte Modelle statt Daten einzelner Patienten getauscht werden oder synthetische statt echter Daten zum Einsatz kommen. Forschung auf „Big Data“ erzwingt also in den wenigsten Fällen eine Zusammenführung der Daten, noch muss sie mit Datenschutz in Konflikt stehen.

Es ist daher an der Zeit, nach technischen Lösungen zu suchen, die relevante Informationen austauschen, ohne dabei individuelle Patienten-/Probandenrechte zu tangieren. Die KI-Technologien der jüngsten Jahre machen dies erstmals möglich und fordern es andererseits ein. Wir sehen diese Entwicklung als Chance, die es zu nutzen gilt.

Die Betrachtung der Privatsphäre muss allerdings über die Ebene der Rohdaten hinausgehen. Auch die Produkte und Derivate, wie z. B. erlernte KI-Modelle, tragen noch Informationen über die Daten in sich. Daher muss, um Privatsphäre zu gewährleisten, auch Sorge getragen werden, dass über solche Derivate kein Zugriff oder Inferenzmöglichkeiten für die Originaldaten entstehen. Es müssen daher Konzepte, Designs, Methodiken und Standards für datenintensive Forschung entwickelt werden, die explorative Forschung und Anwendung im Rahmen des Datenschutzes unterstützen. Dies erfordert interdisziplinäre Forschung an der Schnittmenge von Data Science, Gesundheit und Recht. Hierbei sind u. a. auch Nachbardisziplinen wie Mathematik, Informatik oder Ingenieurwissenschaften angesprochen.

Technische Problemlösungen für die gemeinsame Nutzung von Daten sind aufwändig und teuer. Sie dürfen nicht „beim Kunden“ scheitern, sondern müssen während ihrer Entwicklung bereits auf ihre Performance hin überprüft werden. Benötigt wird ein ganzheitlicher Ansatz, der Methoden der Systemsicherheit, Privatsphäre und Kryptographie anwendungsspezifisch zusammenführt. Forschungsförderern und den Forschungseinrichtungen wird daher empfohlen, in naher Zukunft die Organisation und Durchführung so genannter „Benchmarking Competitions“ zu fördern, die es erlauben, den Stand der Implementierung von föderierten Trainingsansätzen von Vorhersage-Modellen in wohl definierten Szenarien zu eruieren und zu bewerten. Derartige „Benchmarking Competitions“ sollten beispielsweise in der MII durchgeführt werden.

3. Verbesserung der Verfügbarkeit und Qualität von Gesundheitsdaten

Um Versorgungsdaten für die Forschung nutzbar zu machen, sind mehrere Schritte erforderlich. Zunächst müssen diese Daten auffindbar gemacht werden, indem ihre heterogenen Quellsysteme über geeignete Schnittstellen zugreifbar gemacht werden (technische Interoperabilität). Um die dann verknüpfbaren Daten auch wissenschaftlich nutzbar zu machen, ist eine aufwendige Datenaufbereitung (Data Cleaning) und zudem eine möglichst genaue Kenntnis des Datenkontextes notwendig: Wie wurde die Information erhoben? Warum wurde die Information erhoben? Dabei ist auch ein gemeinsames Verständnis über die Inhalte der gespeicherten Daten herzustellen (semantische Interoperabilität). Dazu werden i. d. R. Ordnungssysteme und Nomenklaturen eingesetzt. Auf dieser Basis kann ein Such- und Vermittlungsdienst geschaffen werden, der den Forschenden das Auffinden und Beantragen geeigneter Daten auch über Institutionsgrenzen hinweg ermöglicht.

Da der verwendete Standard die Qualität der Datenerfassung erheblich beeinflusst und „elektronische Datenerfassung“ nicht gleichbedeutend ist mit „strukturierter Datenerfassung“, sind bundesweit einheitliche, international anerkannte und erprobte Standards notwendig, um die o.g. Interoperabilitäten sicherzustellen [5, 6]. Auf die von der MII in Kooperation mit den DZG und Biobanken entwickelten Kerndatensätze muss bei der Entwicklung der Interoperabilität der elektronischen Patientenakte zurückgegriffen werden [5]. Es sollte eine geeignete Organisation (unter Einbeziehung der MII und der DZG) mit der Definition und technischen Umsetzung einer repräsentativen, strukturierten und kontrollierten Gesundheitsdatenplattform beauftragt werden. Am Ende muss ein Prozess stehen, der Patientendaten nur einmal in möglichst niederschwelliger, KI-unterstützter und strukturierter Weise in ein IT-System überträgt und diese dann für die Versorgung wie auch die Forschung im Sinne der Patienten gleichermaßen nutzbar macht. KI-basierte Verfahren werden helfen, niederschwellige Dateneingaben sowie Auswertungen zu ermöglichen. Dies muss begleitet werden durch intensive und frühe Öffentlichkeitsarbeit, um Versicherte und Patienten vom Nutzen einer forschungskompatiblen elektronischen Patientenakte zu überzeugen. Weiterhin müssen auch bei den Leistungserbringern entsprechende Anreize zur strukturierten Datenerfassung in den KIS-Systemen und der elektronischen Patientenakte gegeben werden. Hierzu gehören auch adäquate Vergütungsstrukturen für eine umfassende und strukturierte Datenerfassung, Qualitätssicherung und Bereitstellung.

Die folgenden, von der Technologie- und Methodenplattform für die vernetzte medizinische Forschung (TMF e.V.) bereits empfohlenen, Maßnahmen zur Verbesserung der Verwertbarkeit von Gesundheitsdaten für KI-Entwicklungen und datengetriebene Forschung [4] sollten umgesetzt werden: Die methodische und anwendungsorientierte Weiterentwicklung des Monitorings der Datenqualität sollte durch Forschungsförderung unterstützt werden. Es sollten ein qualitätsgesicherter Lerndatenkörper und für die Forschung verfügbare Public Use Files aufgebaut werden. Die jeweiligen Fachgesellschaften sollten bei der Sicherstellung der Qualität einer solchen Datenbank und bei der Validierung von Daten und darauf trainierter Algorithmen einbezogen werden und repräsentative Datensammlungen zu relevanten klinischen Fragestellungen autorisieren. Des Weiteren sollte die Kommunikation der Medizininformatiker mit dem Ärzte- und Pflegepersonal und Dokumentaren weiter gefördert und gestärkt werden. Das kann durch gemeinsame Workshops, gemeinsame Anträge sowie am besten durch den täglichen Austausch zu, und die Mitarbeit in Projekten umgesetzt werden.

Es wird eine Stärkung des Umgangs des ärztlichen und pflegerischen Personals mit Datenerfassung und Datennutzung empfohlen. Die standardisierte Datenerfassung in den Primärsystemen sollte zusätzlich durch Interventionen (z. B. Programmierung von Kontrollfunktionen) kontrolliert und standardisiert werden. Gleichzeitig sollte das Empowerment von Ärzte-

und Pflegepersonal verstärkt werden. Ärztliches und pflegerisches Personal sollte durch Aus-, Fort- und Weiterbildung und geeignete technische Ausstattung befähigt, motiviert und incentiviert werden, digitale Versorgungsdaten effizienter und umfangreicher zur Verbesserung der individuellen Patientenversorgung sowie der Versorgungssteuerung zu nutzen – nicht zuletzt durch Anwendung von KI-basierten Verfahren. Hierfür ist es wichtig, dass ihnen das richtige Werkzeug bereitgestellt wird, um mit den von ihnen selbst eingepflegten Daten zu arbeiten und diese zu analysieren, die schlussendlich zu einer Verbesserung der Patientenversorgung und Patientenforschung führen. Dies bedeutet, dass sowohl der Zugang zu den unterschiedlichsten Daten als auch die notwendigen Tools und Schulungen bereitgestellt und unterstützt werden müssen. Die hierfür notwendige Entwicklung interaktiver Technologien sollte gemeinsam durch Gesundheitsversorgung, Forschung und Industrie erfolgen.

So wie der Nationale Kompetenzbasierte Lernzielkatalog Medizin (NKLM) zukünftig die digitale Medizin durchgängig im Medizinstudium verankert, sollte dieser Aspekt auch in der Ausbildung der weiteren Gesundheitsberufe integriert werden. Begleitet werden muss dies durch Fort- und Weiterbildungsangebote für bereits im Beruf stehende Arbeitskräfte. Darüber hinaus ist auch eine Verstärkung der Ausbildung und des Einsatzes sogenannter „Data (Curation) Scientists“ in der Versorgung erforderlich. Auch die Programme für Medical und Clinician Scientists sollten durch geeignete Module ergänzt werden. Für die verstärkte Ausbildung von Data Scientists sollten neue Masterstudiengänge etabliert werden. Nur wenn qualifiziertes Personal zur Dokumentation und Datenaufbereitung von Versorgungsdaten im klinischen Alltag ausreichend verfügbar ist, kann eine für die Forschungsnutzung angemessene Datenqualität der Versorgungsdaten sichergestellt werden.

Für die Abstimmung zu bereits laufenden und neuen Aktivitäten zur Aus-, Fort- und Weiterbildung sollte durch die Universitätsmedizin ein entsprechender Steuerungskreis eingerichtet werden. In den Steuerungskreis sollten Vertreter von MII, DZG, des ambulanten und stationären Sektors, Partnern der Selbstverwaltung, Fachverbänden, Gesundheitsfachschulen, Technologieentwicklern sowie Bund und Ländern einbezogen werden.

4. Förderung der Kultur des Datenteilens

In den letzten Jahren haben sich mehrere Publikationen mit der gelebten Praxis des Datenteilens in der medizinischen Forschung und mit den hierfür hemmenden bzw. unterstützenden Faktoren beschäftigt [7-17]. Sie kommen zu dem Schluss, dass bisher nur aus wenigen Studien heraus Daten öffentlich verfügbar gemacht werden. Das Datenteilen findet vor allem innerhalb der jeweiligen Forschergruppe statt, die diese Daten erhoben hat, oder zwischen Forschenden, die persönlich miteinander bekannt sind. Zudem werden öffentlich verfügbare Daten nur wenig nachgefragt, obwohl die Vorteile des Datenteilens den meisten Forschern bekannt sein dürften und diesen z. B. durch die Steigerung von Zitationsraten erwiesenermaßen zugutekommt.

Hauptgründe für den geringen Stellenwert des Datenteilens in der medizinischen Forschung sind die dort bestehende Reputationskultur und Konkurrenzsituation. Datenteilen wird bisher wissenschaftlich zu wenig bis gar nicht honoriert. Es spielt bei der Bewertung von Promotionen, Habilitationen, Stellenbesetzungen oder Berufungen ebenso wenig eine besondere Rolle wie für die Erfolgsaussichten von Drittmittelanträgen. Darüber hinaus befürchten Forschende, dass ein zu frühes Teilen von Daten kompetitiven Publikationen Vorschub leisten könnte. Sie haben Sorge, dass Dritte unfairerweise Vorteile aus ihrer Arbeit ziehen, noch nicht publizierte Erkenntnisse vor ihnen publizieren, ihre Daten falsch interpretieren oder Kooperationspartner den privilegierten Zugang zu Daten verlieren. Forschenden fehlen häufig

auch einfach die notwendigen Kenntnisse und Ressourcen zum Datenteilen. Ihre Daten liegen nicht in einer transferierbaren Form vor, so dass der Aufwand der Datenbereitstellung hoch ist. Und schließlich können viele Daten schon allein aus ethischen Gründen nicht geteilt werden.

Es werden derzeit - auch in der Literatur - unterschiedliche Maßnahmen diskutiert, um das Datenteilen zu befördern: Das Wissen über die Prinzipien des Datenteilens, über die Grundlagen des Datenmanagements und über die aktuell verfügbaren Datenressourcen muss bei den Forschenden deutlich gestärkt werden. Vielfach sind jedoch Sinn und Nutzen des Datenteilens noch zu wenig verinnerlicht. Wichtig scheint dabei auch eine Betonung der Notwendigkeit der Datenkuration in einer strukturierten, annotierten Form, um so die effiziente wissenschaftliche Nutzbarkeit der Daten sicherzustellen. Generell erscheint auch die umfassende Aufklärung und der aktive Einbezug von Datenhaltern und Datenbesitzern einschließlich der Patienten selbst angeraten (siehe hierzu auch Ausführungen in Kapitel 6 Patienteneinbindung bei der Forschung mit Gesundheitsdaten). Ein breiter öffentlicher Diskurs zu den ethischen und sozialen Aspekten des Datenteilens wäre ebenfalls hilfreich. Durch beides ließe sich ein breites Bewusstsein für die Chancen und Risiken der umfassenden Sekundärnutzung medizinischer Daten schaffen und ein gewisser „Druck“ aufbauen, diese auch zu ermöglichen. Dabei soll es nicht darum gehen, alle Daten frei verfügbar zu machen, sondern deren breite Zugänglichkeit durch einheitliche, transparente und kontrollierte Prozesse (oftmals als „Governance“ bezeichnet) zu sichern.

Grundlegende Prozesse des Datenteilens wurden bereits im Rahmen der MII (für die Universitätsklinik), dem Fraunhofer Medical Data Space und innerhalb der DZG entwickelt und etabliert. Diese teils unabhängig voneinander entstandenen Lösungen müssen nun analysiert, harmonisiert und konsolidiert werden, nicht zuletzt um damit der Sorge zu begegnen, dass das Datenteilen in intransparenten oder unzulänglichen Infrastrukturen geschehen könnte. Außerdem muss der Aufwand des Datenteilens weiter reduziert werden, z. B. durch ein größeres Angebot an adäquaten Datenrepositorien.

Forschungsförderer können ein zeitnahes und möglichst einfaches Datenteilen einfordern, z. B. durch die Verankerung von Klauseln in den Zuwendungsbestimmungen zum Datenteilen, zur Anwendung der FAIR-Prinzipien und durch die Überprüfung der Umsetzung entsprechender Klauseln. Darüber hinaus sollten Personalkosten von Dokumentationskräften (Datenkuratoren) als zuwendungsfähige Kosten gefördert werden, so dass die Daten und Metadaten eines Projekts in hinreichender Qualität und Umfang für das anschließende Teilen erfasst werden. So könnte die weitere Finanzierung eines Projekts an die Erfüllung entsprechender Meilensteine geknüpft werden, wie dies auf EU-Ebene bereits häufig praktiziert wird. Darüber hinaus sollten durch die Förderung von Projekten, die nur über den Weg des Datenteilens durchführbar sind und zum Erfolg geführt werden können, von Drittmittelgebern der wissenschaftliche Mehrwert des Datenteilens sichtbar gemacht werden. Weiterhin sollten Forschungsförderer und die Wissenschaft durch gezielte Kommunikationsstrategien die Erfolge des Datenteilens in Form von Use Cases und Best Practice Lösungen als überzeugende Anwendungsbeispiele einer gelungenen gemeinsamen Datennutzung stärker herausstellen. Es wird empfohlen, dass DFG und BMBF zu den o. g. Punkten ein gemeinsames Vorgehen entwickeln. Das BMBF wird gebeten, auch andere Ressorts, die Forschung im medizinischen Bereich fördern, einzubeziehen.

Eine essentielle Voraussetzung zur Veränderung der Kultur des Datenteilens ist, das Datenteilen akademisch besser zu belohnen. Zu den hierfür sinnvollen Maßnahmen gehören z. B. die Sicherstellung von Erstpublikationsrechten, die Nutzung von DOIs als Grundlage für die

Datenzitation, faire Regelungen zur Ko-Autorenschaft und die Wertschätzung von Publikationen, die die ihnen zugrundeliegenden Daten öffentlich machen. Zudem muss die Integration und Veröffentlichung von Daten im akademischen Bewertungssystem analog zu einer wissenschaftlichen Leistung gewürdigt werden können. Arbeitsgruppen innerhalb der DFG (AG Publikationswesen in der Wissenschaft, AG zum Forschungsdatenmanagement) haben die notwendigen Prozesse zur Weiterentwicklung des akademischen Bewertungssystems insgesamt bereits angestoßen. Nach Veröffentlichung entsprechender Handreichungen sollten die entsprechenden Maßnahmen möglichst auch im Kreis der Allianz der Wissenschaftsorganisationen breiter zur Anwendung kommen.

Neben der Förderung der Kultur des Datenteilens ist auch eine strukturelle, nachhaltige Förderung für klinische KI-Forschung in Forschung und Lehre sowie als Karrierepfad erforderlich. Hier ist dringender Handlungsbedarf zur Schaffung attraktiver Karrierewege zum Clinical Data Scientist über alle Ausbildungsebenen geboten. Dies betrifft Studienfächer, Postgraduiertenprogramme für die Medizin, Informatik und Bioinformatik. Darüber hinaus werden gut ausgestattete Nachwuchsgruppen und Lehrstühle benötigt, um exzellente Wissenschaftler und Wissenschaftlerinnen nicht nur im akademischen Kontext, sondern überhaupt in der Gesundheitsforschung zu halten. Zudem sollte frühestmöglich der wissenschaftliche Austausch und die Kommunikation an den Schnittflächen zur IT, Medizininformatik und Bioinformatik gefördert und gestärkt werden, zum Beispiel durch geförderte Projekte, aber auch schon innerhalb der Ausbildung und des Studiums. Die Arbeitsgruppe Medical Scientists des Forums Gesundheitsforschung hat Empfehlungen erarbeitet, die u. a. eine Analyse der wissenschaftlichen Entwicklung und Karrierewege von Clinical Data Scientists vorsehen, um Hinweise dazu zu erlangen, wie die Nachwuchsförderung in diesem Bereich gestaltet werden sollte. Sogenannte „Data Challenges“ können ein wesentlicher Schritt sein, um Vertrauen zwischen KI-Methodikern und Datenhaltern zu etablieren. Beide Seiten sollen mit dem Instrument der Data Challenge schnell den Sinn und die Vorteile des Zusammenarbeitens erkennen. Zudem müssen die Daten nur für einen begrenzten Zeitraum, nämlich für die Dauer der Challenge, und einen begrenzten Personenkreis, nämlich die Teilnehmer, freigegeben werden. Es macht Sinn, solche Data Challenges zu wechselnden Themen, also mit unterschiedlichen Datensätzen und Fragestellungen, regelmäßig wiederkehrend zu fördern.

5. Ethische und rechtliche Aspekte der Datennutzung

Die Verarbeitung personenbezogener Daten in der medizinischen Forschung steht in einem Spannungsfeld zwischen der Forschungsfreiheit und dem Recht auf Datenschutz. Die DSGVO ist im Ausgangspunkt deutlich forschungsfreundlicher als das frühere Recht und enthält diverse Privilegierungen zugunsten der Forschung.

Als Legitimationsgrundlage für die Verarbeitung personenbezogener Daten kommen die Einwilligung der betroffenen Person oder andere gesetzliche Erlaubnistatbestände in Betracht. Eine wirksame Einwilligung setzt – neben ihrer Freiwilligkeit – voraus, dass der Betroffene im Vorfeld hinreichend präzise Informationen darüber erhält, welche Stellen die Daten zu welchem Zweck verarbeiten. Die Grenzen, die durch die informierte Zustimmung der Studienteilnehmenden verschiedener Studien und Forschungsprojekte im Sinne eines „kleinsten gemeinsamen Nenners“ festgelegt werden, schränken die Wiederverwendung von Daten ein.

Für die Datensammlungen, die im Forschungskontext Voraussetzung für den Einsatz und das Training von KI-Auswerteargorithmen sind, und die dann in der Diagnostik und Therapieplanung als Assistenzsysteme eingesetzt werden, kann der Zweck der Verarbeitung zum Zeitpunkt der Datenerhebung aber nicht immer vollständig angegeben werden. Daher schafft

die DSGVO nunmehr den Rahmen für eine breite Einwilligung (Broad Consent), d. h. die Möglichkeit, in die Datenverarbeitung für bestimmte, zum Zeitpunkt der Einwilligung nicht vollständig vorhersehbare Zwecke der wissenschaftlichen Forschung einzuwilligen, solange dies unter Einhaltung anerkannter ethischer Standards geschieht. Auch die World Medical Association und die CIOMS/WHO erkennen den Broad Consent für Forschung an (24). Mittlerweile hat auch die Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder am 15. April 2020 beschlossen [18], dass im Rahmen der MII die Einwilligung in die Nutzung pseudonymisierter Daten aus der Patientenversorgung für medizinische Forschungszwecke unter bestimmten Voraussetzungen als Broad Consent ausgestaltet sein kann, d.h. ohne Begrenzung auf bestimmte Erkrankungsarten oder konkrete Forschungsvorhaben. Für internationale Kooperationen bestehen nach wie vor hohe Hürden in Bezug auf das Datenteilen. Darüber hinaus strebt die MII eine stärkere Integration des niedergelassenen Bereichs an.

Die rechtlichen und ethischen Vorgaben wird von einzelnen Ethikkommissionen und Datenschützern unterschiedlich ausgelegt, so dass es derzeit keinen verlässlichen, deutschlandweiten Standard für Patienteneinwilligungen, Widerruf, Speicherfristen, Zweckbindung, Anonymisierung etc. gibt. Die Neuregelungen des Patientendatenschutz-Gesetzes (PDSG) zum Broad Consent und zum Datenfreigabeverfahren kann hierbei eine verlässliche gesetzliche Orientierung bieten.

Ein Broad Consent verlagert die Verantwortung für Entscheidung und Kontrolle, ob eine Datennutzung im Einklang mit der Einwilligung sowie mit datenschutzrechtlichen und ethischen Vorgaben steht, auf die Forschungsinstitution und deren Forschungsethik- sowie Datennutzungs- und Zugangskomitees, und auf die Forschenden selbst. Entsprechend sieht das in der MII entwickelte Konzept für eine umfassende Einwilligung zu medizinischen Forschungszwecken, einen stringenten Mechanismus für die Datennutzungs- und Zugangsprüfung sowie die Begutachtung durch Forschungsethikkommissionen vor.

Es besteht dennoch aus ethischer und rechtlicher Sicht Klärungsbedarf, durch welche Governance-Strukturen (Ausgestaltung der Entscheidungsverfahren über Nutzung personenbezogener Daten, z. B. durch Einrichtung von unabhängigen Use and Access Committees, Transparenz gegenüber Probanden und Sicherstellung von Informiertheit, technisch-organisatorische Maßnahmen zur Gewährleistung der Datensicherheit (siehe Kapitel III 3.), verteilte Infrastruktur zur Datenanalyse) ein Broad Consent zu ergänzen ist. Es ist die Aufgabe der Forschungseinrichtungen, der nationalen Datenschutzaufsichtsbehörden und Ethikkommissionen, hierzu Standards zu entwickeln, die sowohl den Forschungsinteressen als auch dem Recht der betroffenen Personen auf Schutz ihrer personenbezogenen Daten gerecht werden. Ein Open Data-Szenario, das auf prozedurale Vorkehrungen verzichtet, wäre hingegen – insbesondere mit Blick auf Gesundheitsdaten als besonders sensible personenbezogene Daten – mit den Standards des europäischen Datenschutzes nicht vereinbar.

Die von der Bundesregierung eingesetzte Datenethikkommission verweist alternativ zum Broad Consent auf Einwilligungsmodelle wie den Dynamic Consent oder den Meta Consent. Diese sollen den einmaligen Einwilligungsakt durch ein engmaschiges Rekontaktierungs- und Einwilligungsprozedere mit entsprechenden Privacy Management Tools und Personal Information Management Systemen ersetzen [19, 23]. Allerdings ist aufgrund z. T. geringer Responseraten und evtl. Selektionsverzerrungen zweifelhaft, ob diese Einwilligungsmodelle mit hinreichender Verlässlichkeit zu methodisch befriedigenden Ergebnissen führen würden. Sie geben den Studienteilnehmenden zwar potenziell mehr Kontrolle über die Verwendung ihrer Daten, bergen aber eigene ethische und datenschutzrechtliche Unwägbarkeiten. Hierzu gehören beispielsweise die notwendige De-Pseudonymisierung für die Rekontaktierung, die

Änderung der Einwilligungsfähigkeit im Krankheitsverlauf (z. B. bei Demenz), die Förderung eines therapeutischen Missverständnisses (also des Eindrucks, man hätte als Studienteilnehmerinnen und Studienteilnehmer auch einen therapeutischen Nutzen durch die Studienteilnahme). Dies gilt insbesondere, wenn über diesen Weg auch Ergebnisse zurückgemeldet werden sollen [24].

Neben der Einwilligung der Betroffenen gibt es auch weitere gesetzliche Erlaubnistatbestände, auf deren Grundlage eine Datenverarbeitung zulässig ist und denen nach der Konzeption der DSGVO die gleiche datenschutzrechtliche Legitimierungswirkung wie der Einwilligung zukommt. Die DSGVO eröffnet den nationalen Gesetzgebern Gestaltungsspielräume hinsichtlich der Ausgestaltung der Erlaubnistatbestände (z. B. Art. 9 Abs. 2 j) DSGVO). Elementare Voraussetzung für die wissenschaftliche Nutzung von Daten ohne Einwilligung des Betroffenen ist stets, dass die Forschungsinteressen das Interesse des Betroffenen wesentlich überwiegen, was nur hinsichtlich konkreter Forschungsprojekte, nicht abstrakt und pauschal, im Vorfeld beurteilt werden kann. Allerdings sind die Erlaubnistatbestände im deutschen Recht – auch als Folge der unterschiedlichen Gesetzgebungskompetenzen des Bundes und der Länder – stark fragmentiert und heterogen. Hier besteht Klärungsbedarf [21], inwieweit eine über den Primärzweck der Datenerhebung hinausgehende Weiterverarbeitung von Daten für im öffentlichen Interesse liegende Forschungszwecke zulässig ist. Hierbei sollte die forschungsaffine Haltung des deutschen Gesetzgebers, die er bereits bei der Ausarbeitung der EU-Gesetzgebung eingenommen hat, berücksichtigt werden. Auch in diesem Zusammenhang gilt, dass eine Privilegierung der Forschung durch datenschutzrechtliche Erlaubnistatbestände, durch flankierende Governance-Strukturen und technisch-organisatorische Sicherungsmaßnahmen zu ergänzen ist.

Mit dem im April 2020 verabschiedeten § 287a SGB V – der vorsieht, dass bei länderübergreifenden Vorhaben der Versorgungs- und Gesundheitsforschung eine federführende Aufsichtsbehörde zu benennen und Bundesdatenschutzrecht (§ 27 BDSG) anzuwenden ist – wollte der Gesetzgeber die vorgenannten regulatorischen Inkohärenzen teilweise beseitigen. Die Regelung, die sowohl für die einwilligungsbasierte Datenverarbeitung als auch für die Datenverarbeitung auf der Basis gesetzlicher Erlaubnistatbestände gilt, muss sich in den kommenden Jahren im Hinblick auf die vom Gesetzgeber angestrebte Verfahrensvereinfachung bewähren.

Eine weitere wesentliche Neuerung betrifft die Forschung mit den Daten der elektronischen Patientenakte (ePA) der Gesetzlichen Krankenversicherungen. Das im Juli 2020 verabschiedete Patientendaten-Schutzgesetz (PDSG) ermöglicht es den Versicherten, künftig die Daten ihrer ePA freiwillig für Forschungszwecke zur Verfügung zu stellen. Dies kann als sog. „Datenfreigabe“ an das Forschungsdatenzentrum im BfArM (§ 363 Abs. 1 - 7 SGB V) oder unmittelbar an den Forschenden auf alleiniger Basis einer informierten Einwilligung erfolgen (§ 363 Abs. 8 SGB V). Die genaue Ausgestaltung dieser Optionen befindet sich derzeit in der Umsetzungsphase.

Gleichzeitig hat der EuGH mit dem Urteil „Schrems II“ vom 16. Juli 2020 erhebliche Herausforderungen für den Datenaustausch mit Nicht-EU-Staaten geschaffen. In dieser Entscheidung hat er den Angemessenheitsbeschluss der Europäischen Kommission für die Datenübermittlung in die USA (das sog. „Privacy Shield“) für unwirksam erklärt und für die Datenübermittlung auf der Grundlage von Standarddatenschutzklauseln hohe Anforderungen an die rechtlichen Rahmenbedingungen im Zielland definiert. Es besteht ein erheblicher Klärungsbedarf in Hinblick auf die Frage, wie ein Austausch von Forschungsdaten mit den USA und weiteren Nicht-EU-Staaten, für die kein Angemessenheitsbeschluss existiert, in der Zu-

kunft rechtssicher erfolgen kann. Ferner ist klärungsbedürftig, welche Formen der Datenübermittlungen als Datenübertragungen in einen Drittstaat i. S. d. DSGVO anzusehen sind. Neben direkten Datenübermittlungen kommt hierbei insbesondere auch der Zugriff auf technische Infrastrukturen, insb. Cloud-Computing-Dienste, in Betracht, die in den USA gehostet werden.

Bei der späteren klinischen Anwendung von KI, z. B. in der Diagnostik, ist insbesondere zu berücksichtigen, dass Patienten das Recht haben, nicht einer ausschließlich auf einer automatisierten Verarbeitung beruhenden Entscheidung unterworfen zu werden [Art. 22 DSGVO, 20]. Eine vollautomatisierte Verarbeitung von Gesundheitsdaten, aus denen z. B. Diagnose und Therapieentscheidungen abgeleitet werden, wäre nur mit ausdrücklicher Einwilligung der betroffenen Person oder aufgrund eines gesetzlichen Erlaubnistatbestands zulässig. Liegen diese Voraussetzungen nicht vor, so ist aufgrund der datenschutzrechtlichen Vorgaben eine spezifische Richtigkeits- und Plausibilitätskontrolle der von der KI-Anwendung erzielten Ergebnisse durch einen Menschen erforderlich. Zudem bestehen im Falle einer automatisierten Entscheidungsfindung erweiterte Informationspflichten, die auch „aussagekräftige Informationen über die involvierte Logik“ beinhalten. Es ist derzeit unklar, wann genau von einer automatisierten Entscheidung auszugehen ist und welche Informationen über die verwendeten Algorithmen und zugrundeliegenden Daten im Einzelnen zur Verfügung gestellt werden müssen. Des Weiteren ist zu untersuchen, welche Schlussfolgerungen sich aus der Verpflichtung zur Gewährleistung von Datenqualität (Art. 5 Abs. 1 d) DSGVO) ergeben. Schon allein aus ethischer Sicht ist es wichtig, bei der Nutzung von Decision Support Systemen in der Medizin sicherzustellen, dass die letzte Entscheidungsverantwortung beim Menschen bleibt und dieser sich dessen auch bewusst ist. Mit Blick auf KI-informierte Entscheidungssysteme können die Empfehlungen der EU High Level Expert Group for trustworthy AI als Ausgangspunkt dienen [22].

Zur Wahrung der Daten- und Zugangsgerechtigkeit ist in jedem Fall sicherzustellen, dass KI-Algorithmen mit repräsentativen Datensätzen trainiert werden, um eine Diskriminierung Einzelner oder bestimmter Bevölkerungsgruppen zu vermeiden.

Mit Blick auf die Bereitstellung notwendiger Daten für datengetriebene Gesundheitsforschung im Allgemeinen, sowie der Entwicklung und Anwendung von Methoden der KI im Besonderen besteht Klärungsbedarf zu folgenden ethischen und rechtlichen Fragen:

- (1) Die breite Einwilligung (Broad Consent) ist Voraussetzung für eine Vielzahl von Forschungsvorhaben zur Optimierung der Patientenversorgung, insbesondere durch die Entwicklung einer personalisierten Medizin unter der Nutzung KI-basierter Entscheidungshilfen. Sie ist durch technisch-organisatorische Sicherungsmaßnahmen zu flankieren (z. B. die Einrichtung von unabhängigen Use and Access Committees, verteilte Infrastrukturen zur Datenanalyse und technische sowie organisatorische Maßnahmen zur Gewährleistung der IT-Sicherheit). Es sollten Empfehlungen und Best Practice Beispiele für gute Governance hinsichtlich Patientenaufklärung, Datenhandling und Use and Access Verfahren erstellt und umgesetzt werden. Dazu beitragen können auch Verhaltenskodizes für alle Beteiligten entlang der Daten-Wertschöpfungskette, wie sie von der DSGVO nahegelegt werden [25]. Hierbei sind auch Fragen des Zugangs zu Daten für die Industrie einzubeziehen. Über allem sollte die Forderung stehen, personenbezogene Gesundheits- und Patientendaten nicht in frei zugänglichen Datenbanken zu lagern, sondern nur in solchen mit kontrolliertem Zugang und transparenter Zugangsregelung. Bei klinischen Studien und Forschungsvorhaben, bei denen eine spezifische Einwilligung

ausreichend ist, sollte die Zustimmung zu einer breiteren Verwendung zusätzlich eingeholt werden, wenn daran ein Forschungsinteresse besteht (Add-on Broad Consent separat zur spezifischen Einwilligung).

- (2) Im engen Austausch mit den verschiedenen nationalen Datenschutzbehörden und dem Europäischen Datenschutzausschuss sollten Leitlinien für die Wirksamkeit und den Einsatzbereich von Einwilligungen (insbesondere Broad Consent) sowie Standards für die in diesem Zusammenhang erforderlichen Prozesse zur Anonymisierung und Pseudonymisierung entwickelt werden.
- (3) Das Jahr 2020 ist/war im Hinblick auf die datenschutzrechtlichen Rahmenbedingungen für die Forschung ein Schlüsseljahr. Die mit dem PDSG, § 287a SGB V und dem Mustertext Consent der MII erreichten Verbesserungen und Klarstellungen gilt es nun mit Leben zu füllen und in der Praxis der lebenswissenschaftlichen Forschung sowie der Datenschutzaufsicht zu verankern. Hierfür sollte ein koordinierter Dialogprozess mit allen betroffenen Akteuren gestartet werden.
- (4) Der Nutzen, die Ausgestaltung und Umsetzbarkeit von abgestuften Einwilligungskonzepten (Dynamic Consent) sollte vor einem weiteren In-Betracht-Ziehen empirisch beforscht werden. Untersucht werden sollte dabei insbesondere, in welchem Maß eine abgestufte Entscheidungsbefugnis tatsächlich von den Betroffenen erwünscht ist und genutzt wird.
- (5) Es ist – unter Berücksichtigung der neuen gesetzgeberischen Entwicklungen - eine Vereinheitlichung der Auslegung des Datenschutzrechts sowohl national auf Landes- und Bundesebene als auch EU-weit anzustreben. [21].
- (6) Es besteht Forschungsbedarf hinsichtlich der Auswirkungen des Verbots der automatisierten Entscheidungsfindung bei KI-Anwendungen. Hierbei ist zu untersuchen, welche Bedeutung Systemen mit automatisierter Entscheidungsfindung im Medizinbereich tatsächlich zukommt bzw. perspektivisch zukommen wird, welche Auswirkungen entsprechende Verbote auf die Ausgestaltung und Einsatzfähigkeit der Systeme hätten und wie den besonderen Transparenzvorgaben Rechnung getragen werden kann.
- (7) Es besteht zudem Forschungsbedarf hinsichtlich der Auswirkungen des EuGH-Urteils vom 16. Juli 2020 („Schrems II“) auf internationale Kooperationen und die Nutzung länderübergreifender technischer Infrastrukturen in der KI-gestützten Gesundheitsforschung.

Es wird empfohlen, dass zu den Punkten 1-3 eine Arbeitsgruppe eingerichtet wird, in der im engen Austausch mit den Datenschutzaufsichtsbehörden und den medizinischen Ethikkommissionen prioritär Empfehlungen zur Planung von Datenerhebung einschließlich Mustertexten, Musternutzungsordnung für personenbezogene Daten, generischen Geschäftsordnungen für Governance, und Leitfäden für Use and Access Committees weiterentwickelt werden. Der Punkt 6 sollte beispielsweise im Rahmen der KI-Strategie der Bundesregierung aufgegriffen werden.

6. Patienteneinbindung bei der Forschung mit Gesundheitsdaten

Befragungen von Patientinnen und Patienten sowie der Allgemeinbevölkerung zeigen durchgängig eine positive Einstellung gegenüber der Verwendung persönlicher digitaler Gesundheitsdaten für die medizinische Forschung, solange diese Nutzung der akademischen Forschung vorbehalten bleibt und der Verbesserung von Diagnose, Therapie und Prävention

dient (siehe Deloitte Global Health Care Consumer Survey 2019 [26], FORSA TMF-Befragung [27] etc.). So waren von 1.000 repräsentativ ausgewählten Deutschen über 18 Jahren, die 2019 im Auftrag der TMF durch das Forsa-Institut telefonisch befragt wurden, mehr als drei Viertel (79 %) bereit, ihre persönlichen Gesundheitsdaten für die medizinische Forschung zur Verfügung zu stellen. Über 70 % der Befragten sprachen sich für eine langfristige Nutzung von mindestens fünf Jahren aus, und 56 % wären damit einverstanden gewesen, dass ihre Gesundheitsdaten zeitlich unbegrenzt für die medizinische Forschung genutzt werden. Zwischen dem Antwortverhalten von chronisch Kranken und Gesunden gab es keine statistisch signifikanten Unterschiede. Allerdings waren die Befragten skeptisch gegenüber einer Herausgabe ihrer Daten an die Industrie (z. B. für die Arzneimittelforschung); hiermit waren nur 17 % von jenen Befragten einverstanden, die eine Nutzung ansonsten grundsätzlich bejahten. Wesentliche Voraussetzungen für die Bereitschaft zum Teilen persönlicher Gesundheitsdaten sind offensichtlich (i) Transparenz hinsichtlich der Verwendung der Daten, (ii) die Gewährleistung von Datenschutz und Datensicherheit (iii) der Ausschluss einer Kommerzialisierung und einer Weiterleitung an Dritte wie z. B. Arbeitgeber oder Versicherer und (iv) die Verhinderung von Diskriminierung und Stigmatisierung [28].

Neben diesen eher organisatorisch-technischen Anforderungen spielt die Wahrung der Patientenautonomie eine entscheidende Rolle für die Neigung, sensitive medizinische Daten für Forschungszwecke preiszugeben. Die fortschreitende Digitalisierung der meisten Lebensbereiche, die breite Verfügbarkeit medizinischer Fachinformation sowie die Möglichkeiten zur Selbstdiagnose (auch mithilfe von Wearables mit eingebauter Sensorik) sowie die direkte Verfügbarkeit (quasi-)diagnostischer Tests („direct-to-consumer“) bedeuten, dass Patienten immer mehr zu aktiven Verwaltern und Gestaltern ihrer gesundheitlichen Angelegenheiten werden. Nicht zuletzt mit Einführung der digitalen Patientenakte werden daher die meisten für die Forschung relevanten Daten zukünftig bei den Patientinnen und Patienten selbst zusammengeführt. Damit rückt der Patient ins Zentrum der Bemühungen um eine möglichst umfassende Erschließung medizinischer Daten für die Forschung. Zudem bietet die Einbeziehung der Betroffenen in die Datennutzung auch den Forschenden viele Vorteile: Ihre wissenschaftlichen Fragen können gemeinsam präzisiert und an die Bedürfnisse der Versorgung angepasst werden, Studienteilnehmerinnen und -teilnehmer lassen sich gezielter rekrutieren, und auch die Datenqualität wird sich durch die größere Nähe zum Datensubjekt verbessern lassen.

Somit stellt sich die Frage, wie Patienten und Datenfreigeber in geeigneter Weise in die Gestaltung der medizinischen Forschungslandschaft einbezogen werden können. Wie können und wollen sie an datenreicher medizinischer Forschung partizipieren, und wie werden neue Informationstechnologien ihre Partizipationsmöglichkeiten bestimmen? Auf dem Weg zu einer durchaus berechtigten und wünschenswerten Stärkung der Patientenbeteiligung ist allerdings Sorge zu tragen, dass der Einzelne intellektuell nicht überfordert wird und über ausreichende technische Ressourcen und ein gewisses Maß an „Health Literacy“ verfügt. Nur so werden Patientinnen und Patienten ihre Datenhoheit auch tatsächlich fair ausüben und ihre Beteiligungsmöglichkeiten nutzen können. Wo es Defizite in dieser Hinsicht gibt, müssen den Betroffenen von Seiten der Forschung entsprechende Unterstützungsangebote gemacht werden.

Es wird empfohlen, in einer DZG übergreifenden Arbeitsgruppe unter Einbeziehung von Vertretern aus MII, AUFGE sowie Vertretern aus Patientenorganisationen technische und organisatorische Instrumente zu entwickeln und anschließend der wissenschaftlichen Gemeinschaft zur Verfügung zu stellen, die die Partizipationsmöglichkeiten der Patientinnen und Pa-

tienten bei datenreicher Forschung stärken. Es sollten konkrete Partizipationsmodelle entwickelt und Best-Practice-Beispiele beschrieben werden. Patientinnen und Patienten sollten bei der Planung und Entwicklung dieser Modelle einbezogen werden. Bei der Stärkung der Patienteneinbindung könnte zukünftig auch interaktiven Technologien eine stärkere Rolle zukommen. Bei der Entwicklung der Instrumente sollten Vertreter der MII, der DZG und der AUFE einbezogen werden.

Gerade mit Blick auf KI-Anwendungen auf der Grundlage großer Datenmengen scheint ein studien- und standortübergreifendes Patienten Consent Management dringend erforderlich, das die gemeinsame Nutzung solcher Daten rechtlich legitimiert und praktisch ermöglicht. Mittelfristiges Ziel muss dabei die strukturierte, digitalisierte Dokumentation aller relevanten Einwilligungskomponenten sein, die neben den Vorteilen einer einfachen Abrufbarkeit durch die Forscher auch die Möglichkeit zur Nutzung individualisierter Kommunikationswege zum Patienten (und damit Datenfreigebenden) eröffnet. Diese wiederum würden eine deutlich stärkere Beteiligung am Datennutzungsprozess in der Forschung realisieren, als bisher bei der Verwendung analoger Einwilligungsmedien möglich. Es wird daher empfohlen in der DZG/MII/AUFE Arbeitsgruppe auch praxistaugliche Instrumente zu entwickeln und zu implementieren, die ein studien- und standortübergreifendes Consent Management (CM) für die Nutzung großer Datenmengen, insbesondere für KI-Anwendungen, erleichtern. Aufbauend auf einer Bestandsaufnahme existierender CM-Lösungen sollten Pflichtenhefte und Umsetzungsvorschläge erarbeitet werden. Diese sollten im Dialog mit potentiellen Entwicklern von Lösungen und Anbietern von Services umgesetzt werden.

An die Unterstützung der Forschung durch die Überlassung von Daten sind auf Seiten der Patientinnen und Patienten in der Regel bestimmte Erwartungen geknüpft: Die Forschung an den geteilten Daten soll dem Interesse der Allgemeinheit dienen und die Werte der Datenfreigebenden widerspiegeln. Risiken müssen minimiert und Transparenz über die Verwendung der Daten aufrechterhalten werden. Um darüber im Interesse der Patientinnen und Patienten wachen zu können, sind wie in Kapitel 5 gefordert geeignete Governance-Strukturen erforderlich, an denen auch Patientenorganisationen und ggf. andere relevante Gruppierungen wie Kirchen und Verbände zu beteiligen sind. Die Werte, Bedürfnisse und Interessen der Patientinnen und Patienten müssen Grundlage dieser Governance sein. Nur so wird sich langfristig eine breite gesellschaftliche Akzeptanz für die Forschung an großen Mengen sensibler, weil medizinischer, personenbezogener Daten sichern lassen.

Da gerade KI-basierte Projekte und Verfahren zur Datennutzung oftmals sehr weit vom Individuum und vom primären Erhebungszweck verortet sind, stellt sich hier die Frage nach der Datenverantwortung in besonderer Weise. Ein Weg, den Patienten in dieser Hinsicht mehr Beteiligung zu ermöglichen, wäre die Schaffung technischer Voraussetzung für eine individuelle, durch den Patienten selbst anpassbare Einstellung der Zugriffsrechte und -möglichkeiten für seine Daten (9). Der Medical Data Space der Fraunhofer-Gesellschaft könnte hierzu einen ersten Lösungsansatz bieten. So attraktiv diese Option aus ethischer und datenschutzrechtlicher Sicht auch scheinen mag, bei entsprechenden Überlegungen müsste stets berücksichtigt werden, dass die Extraktion von Einzeldaten bei komplexen KI-Methoden eine große technische Herausforderung darstellt.

IV. Exkurs

Die folgenden Empfehlungen beziehen sich auf wichtige zulassungs- und haftungsrechtliche Aspekte bei der Implementierung datengestützter Verfahren in der Medizin und auf perspektivische Möglichkeiten einer Datenüberlassung. Die Empfehlungen gehen über den primären

Fokus des Papiers, nämlich die Identifizierung von durch die Akteure des Forums Gesundheitsforschung umsetzbare konkrete Maßnahmen zur Verbesserung der Nutzbarmachung digitaler Daten für KI-Entwicklungen und datengetriebene Gesundheitsforschung hinaus. Sie richten sich primär an Akteure außerhalb des Forums Gesundheitsforschung. Aufgrund ihrer hohen Bedeutung für die tatsächliche Nutzung von KI-Entwicklungen in der Medizin und für eine Vereinfachung der Datenüberlassung empfiehlt das Forum Gesundheitsforschung den vorgeschlagenen Akteuren die Umsetzung der Maßnahmen.

1. Zulassungs- und haftungsrechtliche Aspekte KI-basierter Anwendungen in der Medizin

1.1 Zulassungsrechtliche Aspekte

Zulassungsrechtliche Rahmenbedingungen

KI-basierte Anwendungen können medizinischen Zwecken dienen und sind dann als Medizinprodukte-Software einzustufen. Insofern ist die ab 26. Mai 2021 geltende EU-Medizinprodukteverordnung („MPVO“) zu beachten. Medizinprodukte dürfen nur in Verkehr gebracht werden, wenn sie ein Konformitätsbewertungsverfahren erfolgreich durchlaufen und eine CE-Kennzeichnung erhalten haben. Die bisherigen Rechtsunsicherheiten hinsichtlich der Qualifizierung und Klassifizierung von Software als Medizinprodukt bleiben auch unter dem Regime der neuen MPVO im Wesentlichen bestehen.

Die Qualifizierung von Software als Medizinprodukt wird untergesetzlich durch eine von der Koordinationsgruppe Medizinprodukte (Medical Device Coordination Group, MDCG) entwickelte „Software-Leitlinie“ konkretisiert. Die „Software-Leitlinie“ schöpft mit Medical Device Software (MDSW) einen neuen Begriff und definiert diese als Software, die dazu bestimmt ist, allein oder in Kombination einen spezifizierten Zweck zu erfüllen – unabhängig davon, ob es sich um *stand-alone software* handelt oder die Software das Medizinprodukt steuert bzw. beeinflusst. Die Risikoklassifizierung der Software und mithin der für den Marktzugang erforderliche Aufwand richtet sich nach Anhang VIII Ziff. 6.3 (Regel 11) MPVO (Klassifizierungsregeln). Regel 11 kann in bestimmten Konstellationen zu Klassifizierungen führen, die dem basalen regulatorischen Konzept eines *risk based approach* nicht gerecht werden und damit innovationshemmende Wirkung entfalten, ohne gleichzeitig ein Mehr an Sicherheit zu gewährleisten. Insofern liegt ein Fall der Überregulierung vor. Deutschland sollte als EU-Ratsmitglied darauf hinwirken, dass bei der nächsten Reform des EU-Medizinprodukterechts die Regel 11 innovationsaffiner gestaltet und den Besonderheiten KI-basierter Software Rechnung getragen wird.

Der Marktzugang (speziell) KI-basierter Medizinprodukte-Software wird im gegenwärtigen Recht nicht berücksichtigt und dadurch erschwert. Denn die Konformitätsbewertung eines Medizinprodukts stellt notwendig auf einen bestimmten Zeitpunkt ab, an dem die Anforderungen der MPVO erfüllt sein müssen. Kontinuierlich lernende und adaptive KI-Algorithmen können jedoch einen anderen Output als den ursprünglich für einen Input freigegebenen Output liefern. Eine derartige Medizinprodukte-Software verliert ihren Konformitätsstatus – und damit die Marktzugangsvoraussetzung – zwangsläufig schon mit ihrer ersten Änderung. Zudem bezieht sich die Konformität auf die Zweckbestimmung, die der Hersteller dem Medizinprodukt gegeben hat. Insofern ist zumindest denkbar, dass eine KI-gestützte Weiterentwicklung

der Software dazu führt, dass sie dem eigentlich intendierten Zweck nicht mehr genügt bzw. diesen nicht mehr verfolgt oder sogar über ihn hinausgeht.

Bis zu einer regulatorischen Lösung bzw. Änderung des Status quo bleibt damit als einzig gangbarer Weg die Durchführung eines erneuten Konformitätsbewertungsverfahrens bei relevanten KI-induzierten Änderungen der Software. Ein paralleles Problem ergibt sich im Hinblick auf die einmalige Produktkennung (UDI), mit der Hersteller ihr Medizinprodukt versehen müssen (Art. 27 MPVO). Diese muss erneuert werden, wenn Änderungen an der Software vorgenommen werden (Anhang VI, Ziff. 6.5.2, 6.5.3 MPVO). Bei einer sich selbst ändernden Software müsste die UDI also ständig erneuert werden. Auch dies zeigt, dass die derzeitige Regulierung nicht auf KI-basierte Anwendungen ausgelegt ist. Entsprechendes gilt für ggf. gegenüber der Benannten Stelle zu erfüllende Anzeigepflichten bei Änderungen (vgl. Anhang IX Abschnitt 4.10, Anhang X Abschnitt 5).

Der Problemkomplex einer sich fortentwickelnden Software wird also bis dato nicht hinreichend adressiert. Einen dem FDA-Vorhaben entsprechenden Plan, dass zu erwartende Modifikationen bzw. Erweiterungen der Software bereits im ursprünglichen Zertifizierungsverfahren angegeben werden können, gibt es auf EU-Ebene bislang nicht. Eine regulatorische Lösung müsste wohl dahingehen, dass nicht (nur) der Status quo einer Software zertifiziert wird, sondern ähnlich wie in den USA auch ihre Lernprozesse bzw. ihr Algorithmus.

Deutschland sollte über seine Vertreter in der MDCG, d. h. durch die hierfür benannten Beauftragten des Bundesministeriums für Gesundheit (BMG) bzw. der Zentralstelle der Länder für Gesundheitsschutz bei Arzneimitteln und Medizinprodukten (ZLG), bei der Erstellung des in der Untergruppe „New Technologies“ geplanten Leitfadens „Artificial Intelligence under MDR/IVDR Framework“ in der Weise mitwirken, dass eine konsequent am risikobasierten Ansatz des EU-Medizinprodukterechts orientierte Anwendung der jeweils einschlägigen Bestimmungen auf KI-basierte Software unter Berücksichtigung ihrer Besonderheiten empfohlen wird [29-32].

1.2 Haftungsrechtliche Aspekte

Beim Einsatz KI-basierter Anwendungen stellen sich zahlreiche Haftungsfragen. Im Bereich der Medizin greifen dabei verschiedene Haftungsregime mit unterschiedlichen Voraussetzungen ineinander, etwa die Arzthaftung und die Produkt- und Produzentenhaftung, mit entsprechend unterschiedlichen Haftungssubjekten, etwa den Behandelnden (Ärzte, Krankenhäuser) oder den Medizinprodukteherstellern. Einige Fragen seien im Folgenden skizziert:

Haftung der Hersteller der Software/Hardware

Für den Hersteller kommt eine Haftung insbesondere auf Grundlage der deliktsrechtlichen Produzentenhaftung (nach BGB) sowie auf Grundlage der – unionsrechtlich harmonisierten – Produkthaftung (nach dem Produkthaftungsgesetz) in Betracht [33]. Hinsichtlich des Anwendungsbereichs der Produkthaftung (nach Produkthaftungsgesetz) ist dabei schon streitig, ob dieser Software (Computerprogramme, sonstige Daten) umfasst oder auf körperliche Produkte beschränkt ist. Offen sind sodann die Herstellerpflichten. Wie ist mit dem für KI kennzeichnenden Umstand der Lernfähigkeit und dem daraus resultierenden „Intelligenzrisiko“ bzw. „Autonomierisiko“ umzugehen? Kann sich der Hersteller durch bloßen Verweis auf die „Lernfähigkeit“ des Produkts von einer Haftung für „Lernfehler“ befreien? Müssen hier ggf. prozedurale Maßnahmen zur Beherrschung dieser Risiken ergriffen werden und wie weit reichen diese? Welche Pflichten bestehen nach Inverkehrbringen des Systems (Produktbeobachtungspflichten), wie weit reichen diese (Produktrückruf, Produktdeaktivierung, Pflicht zur Bereitstellung von Updates etc.) und wie lange bestehen diese?

Haftung der Behandelnden (insb. Ärztinnen und Ärzte)

Hier bestehen zahlreiche ungeklärte Fragen, unter anderem: Welche Anforderungen muss eine KI erfüllen, damit ihr Einsatz dem Facharztstandard (§ 630a Abs. 2 BGB) genügt? Ist das bei einem selbstlernenden System mit dem ihm inhärenten Autonomie- bzw. Intelligenzrisiko überhaupt möglich? Darf sich der Arzt auf die Ergebnisse einer KI verlassen? Wie weit reicht seine Überprüfungspflicht? Wer trägt das Risiko von auf Autonomie bzw. künstlicher Intelligenz beruhenden Fehlfunktionen? Sind intelligente Systeme vollbeherrschbare Risiken (§ 630h Abs. 1 BGB), die zur Umkehr der Beweislast im Haftungsprozess führen? Worüber ist beim Einsatz von KI aufzuklären (§ 630e Abs. 1 BGB)? Könnten Fortschritte bei der KI-Forschung dazu führen, dass haftungsrechtlich der Einsatz von KI aus Gründen „besserer“ Ergebnisse (beispielsweise im radiologischen Bereich) sogar geboten ist? Ist die vollständige Übertragung medizinischer Behandlung auf eine KI mit dem Arztberufsbild zu vereinbaren und berufsrechtlich (sowie datenschutzrechtlich s.o.) zulässig?

Haftung weiterer Akteure

Jenseits der Hersteller und der Behandelnden können weitere Akteure beteiligt sein, hinsichtlich derer die Haftung zu klären ist. Zu denken ist hier insbesondere an die Zulieferer von Daten. Deren Haftung steht beispielsweise in Frage, wenn die Fehlfunktion des Systems darauf beruht, dass fehlerhafte Daten zugeliefert wurden oder notwendige Daten nicht verfügbar waren.

Versicherungsrechtliche Fragen

Eng verknüpft mit der Haftung ist die Frage nach der Versicherung der Haftungsrisiken. Auch insoweit besteht Forschungsbedarf. Über rein versicherungsrechtliche Fragen dürfte sich dieser auch auf versicherungsmathematische und versicherungsbetriebswirtschaftliche Fragen erstrecken.

Das haftungsrechtliche Umfeld der Nutzung digitaler Daten in der Gesundheitsforschung und – erst recht – der Anwendung KI-basierter Entwicklungen in der Praxis ist derzeit nicht umfassend vermessen. Entsprechendes gilt für die Versicherung solcher Risiken. Ungeklärte Haftungsrisiken können sich nachteilig auf die Innovationskraft der Hersteller sowie der Nutzer als potentiell Haftpflichtige auswirken. Bei den Patientinnen und Patienten können Unklarheiten und Fehlvorstellungen zu Vorbehalten, Ängsten und Ablehnung gegenüber KI führen. Es ist deshalb für beide Seiten – Geschädigte und potentiell Haftpflichtige – von zentraler Bedeutung, dass – im ersten Schritt – Klarheit über die Haftung und die Handlungsspielräume gefunden wird. Dabei sind vor allem die Besonderheiten unter „KI“ erfassten vielfältigen Verfahren und deren Relevanz für die Haftung herauszuarbeiten. Davon ausgehend kann dann im zweiten Schritt bestimmt werden, wie die Haftung und gegebenenfalls Versicherung ausgestaltet werden sollten, um den berechtigten Interessen beider Seiten bestmöglich zu gerecht zu werden. Beispielsweise kann das Entschädigungsinteresse nicht nur durch eine individuell zu begründende Haftung, sondern auch durch eine davon unabhängige, direkt an die Schädigung anknüpfende Entschädigung aus einem Pool (wie bei der Kranken- oder Unfallversicherung) gewährt werden. Andererseits können über das Haftungsrecht auch Anreize zur Entwicklung sicherer Produkte und damit Innovationsanreize gesetzt werden.

Handlungsempfehlungen

Mangels verlässlicher Grundlagen sind im Moment noch keine dezidierten Aussagen zu möglichen gesetzgeberischen Handlungsbedarfen formulierbar. Deshalb muss zunächst die Forschung in diesem Bereich forciert werden. Erforderlich ist eine umfassende Grundlegung

zu den haftungsrechtlichen Fragen (z. B. Arzthaftungsrecht, Produkthaftungsrecht etc.) von KI-basierten Anwendungen sowie den damit verbundenen versicherungsrechtlichen sowie weiteren Rechtsfragen (z. B. ärztliches Berufsrecht, Sozialversicherungsrecht etc.). Eine koordinierte und strukturierte Forschung muss sowohl haftungs- und versicherungsrechtliche Einzelfragen als auch Fragen zum Gesamtsystem umfassen.

Dies sollte zum einen in Form von Einzelstudien zu den verschiedenen Aspekten (etwa des Arzthaftungsrechts, des Medizinprodukterechts etc.), zum anderen aber auch in Form bereichsübergreifender Studien erfolgen, um die vielfältigen Querverbindungen und Interaktionen mit den jeweils anderen Rechtsbereichen sowie angrenzenden Rechtsbereichen (z. B. dem Sozialversicherungsrecht, dem ärztlichen Berufsrecht etc.) in den Blick zu nehmen.

Es wird empfohlen, dass dies im Rahmen der Umsetzung der KI-Strategie der Bundesregierung aufgegriffen wird. Zu berücksichtigen ist, dass das EU-KI-Weißbuch vorsieht zu überprüfen, ob die bisherigen Regelungen ausreichend sind und wie diese ggf. angepasst werden sollten. Eine Anpassung des EU-Rechtsrahmens zu KI-Entwicklungen ist für Q1 2021 geplant [34-42].

2. Perspektivische Empfehlung: Ein Plädoyer für die Datenüberlassung für die Forschung

Im Jahr 2020 gab es hinsichtlich der datenschutzrechtlichen Rahmenbedingungen für die lebenswissenschaftliche Forschung in Deutschland eine Reihe wichtiger und hilfreicher Entwicklungen. Mit dem PDSG, dem § 287a SGB V und der Zustimmung der Datenschutzkonferenz zum Mustertext zur Patienteneinwilligung der MII konnten wesentliche Verbesserungen und Klarstellungen erzielt werden. Kurz- und mittelfristig gilt es nun, diese neuen Möglichkeiten mit Leben zu füllen und in der Praxis der lebenswissenschaftlichen Forschung sowie der Datenschutzaufsicht zu verankern. Gleichwohl möchte das vorliegende Strategiepapier das Erreichte weiterdenken und in diesem Kapitel eine perspektivische Empfehlung zum Thema Patienteneinwilligung formulieren.

Wie jede andere Form der Nutzung personenbezogener medizinischer Daten bedarf auch die datengetriebene Gesundheitsforschung bzw. die Entwicklung und Validierung KI-basierter Verfahren einer Rechtsgrundlage in Form einer gesetzlichen Erlaubnis oder der informierten Einwilligung der Betroffenen. Bei allen aktuellen Initiativen zur Erschließung von Patientendaten für die medizinische Forschung in Deutschland steht dabei die Einwilligungslösung im Fokus. Die Einholung der Einwilligungen erfolgt zumeist in konkreten Versorgungssituationen, was mit einer Reihe von Nachteilen verbunden sein kann.

- Zum einen werden die entsprechenden Aufklärungs- und Einwilligungsdokumente den Patientinnen und Patienten meistens während der Aufnahme und zusammen mit anderen, behandlungsrelevanten Dokumenten wie z. B. Behandlungsvertrag und Schweigepflichtentbindung ausgehändigt. Eine Aufklärung zur späteren Forschungsnutzung von Daten kann aus Kapazitäts- und Praktikabilitätsgründen nur schriftlich bzw. unter Zuhilfenahme anderer Medien oder mündlich durch das Aufnahmepersonal erfolgen. Letzteres bedeutet aber im Versorgungsalltag einen erheblichen zusätzlichen Ressourcenaufwand.
- Während der Aufklärung und Einwilligung befinden sich die Patientinnen und Patienten in der Regel in Erwartung einer einschneidenden Diagnose oder einer unangenehmen bzw. gefährlichen Therapie. Die Abwägung des Für und Wider einer Sekundärnutzung ihrer Daten für die Forschung bedeutet dann eine zusätzliche Belastung der Patienten.

Zudem kann die zeitliche und räumliche Verknüpfung der Einwilligung mit konkreten Versorgungsmaßnahmen zu Missverständnissen hinsichtlich der Bedeutung der Entscheidung des Patienten für den Behandlungsforgang führen.

- Es besteht darüber hinaus das grundsätzliche Problem, bei Patientinnen und Patienten ein hinreichendes Verständnis bezüglich des Nutzens und der Risiken einer Sekundärnutzung medizinischer Daten für Forschungszwecke herzustellen. Zumindest in schriftlicher Form gelingt dies in der zur Verfügung stehenden Zeit und trotz intensiver Bemühungen um Lesbarkeit und einfache Darstellung nicht in vollem Umfang.

Erste empirische Untersuchungen signalisieren, dass unter Patientinnen und Patienten in Deutschland ungeachtet der o. a. Einschränkungen eine große Bereitschaft besteht, persönliche medizinische Daten für die Forschung zur Verfügung zu stellen - ggf. auch ohne explizite Einwilligung. Kürzlich durchgeführte Umfragen ergaben diesbezügliche Zustimmungswerte von weit über 70 % [43, 44]. Wenngleich diese Studien noch weiterer unabhängiger Bestätigung bedürfen, verdienen ihre Ergebnisse insbesondere deshalb Aufmerksamkeit, weil spätestens mit der Stellungnahme „Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung“ [45] des Deutschen Ethikrats vom November 2017 die Möglichkeit einer „Datenspende“ thematisiert wird.

Die Anwendung des Begriffs „Spende“ auf Daten ist semantisch nicht unproblematisch, da Daten in Gegensatz zu Geld und stofflichen Gütern unbegrenzt teilbar sind und sich daher durch ihre Nutzung nicht verbrauchen. Zudem impliziert der Begriff „Spende“ eine fehlende Rückholbarkeit des übermittelten Gutes, die angesichts des in Art. 7 Abs. 3 und Art. 21 DSGVO verankerten Widerrufs- und Widerspruchsrecht problematisch wäre. Für die nachfolgenden Überlegungen haben die Autoren des vorliegenden Strategiepapiers daher den Begriff der „Datenüberlassung für die Forschung“ gewählt und befassen sich mit der Frage, wie diese ausgestaltet sein könnte.

Zunächst spricht eine Reihe praktischer Überlegungen dafür, die rechtliche Zulässigkeit einer Implementierung der Datenüberlassung an die Forschung in Form einer Opt-out Lösung zu prüfen. Die Art und Weise, wie die Entscheidung zur Datenüberlassung getroffen wird (d. h. Opt-in oder Opt-out), hat nämlich einen großen Einfluss auf den wissenschaftlichen Wert der Daten. Zum einen gingen bei einer Opt-in-Lösung im akuten Behandlungszusammenhang einige wichtige Versorgungsfälle für die medizinische Forschung verloren, z. B. Nottfälle oder primär nicht einwilligungsfähige Patienten. Verlegt man den Einwilligungsvorgang aus dem Behandlungskontext in den privaten Bereich, so ergeben sich Zweifel an der praktischen Umsetzbarkeit und an einer ausreichenden Beteiligung. Wie im Fall der Organspende wäre wohl - trotz grundsätzlich großer Bereitschaft - mit einer geringen und letztlich nicht-repräsentativen Teilnahme zu rechnen. Durch mangelnde Repräsentativität können medizinische Daten aber nicht nur wertlos werden, sondern gerade bei KI-basierter Nutzung zu potentiell gefährlichen Fehlschlüssen führen. Daher scheint eine Untersuchung sinnvoll, ob eine Verbindung der Datenüberlassung für die Forschung mit einem Opt-out-Mechanismus im Rahmen der DSGVO zulässig wäre.

Obwohl denkbar ist, dass Patienten die Nutzungsrechte an ihren Daten gezielt einer bestimmten Person oder Einrichtung einräumen möchten, z. B. aus Dankbarkeit für eine erfolgreiche Behandlung, werden die allermeisten Patienten medizinische Forschung im Allgemeinen unterstützen wollen [43]). Es spricht daher einiges dafür, die Datenüberlassung für die Forschung so auszugestalten, dass die darunterfallenden Daten grundsätzlich allen wissenschaftlich ausgewiesenen öffentlichen und privaten Forschungseinrichtungen zugänglich sind. Die Entscheidung über die Nutzung der Daten sollte - nach Prüfung der gesetzlichen

Voraussetzungen und ggf. begleitet durch ein positives Ethikvotum - allein aus wissenschaftlichen Erwägungen getroffen werden. In der Bevölkerung lässt sich bei Befragungen eine erhebliche Zurückhaltung gegenüber der Forschungsnutzung medizinischer Daten durch private Einrichtungen (die „Industrie“) feststellen [45]. Solche Ressentiments gegen die industrielle medizinische Forschung und deren Nutzung von verschlüsselten Patientendaten verkennen aber deren große Bedeutung für den medizinischen Fortschritt. Die Zusammenarbeit zwischen öffentlich und privat finanzierter Forschung ist eine notwendige Gegebenheit, die von vielen Förderorganisationen sogar aktiv unterstützt wird. Technische Lösungen zur Nachvollziehbarkeit der Datennutzung unter Wahrung der digitalen Souveränität der Patientinnen und Patienten könnten entsprechende Hürden abbauen.

Angesichts der breiten und ungeteilten Zustimmung der Bürgerinnen und Bürger zur Nutzung ihrer Daten (siehe Kapitel III 6.) sowie der Herausforderungen der klassischen Einwilligungslösung scheint die hier skizzierte freiwillige, dauerhafte Überlassung von Patientendaten für die Forschung, verbunden mit einer möglichst niederschweligen Möglichkeit zum Opt-out, eine geeignete Form ihrer Legitimierung zu sein, sofern sie mit der DSGVO vereinbar ist. Die Industrie sollte für die Forschung an und Entwicklung von Therapeutika, Diagnostikverfahren sowie Impfstoffen in die Datennutzung einbezogen werden können. Eine Herausforderung auf der Ebene des nationalen Rechts könnten allerdings die unterschiedlichen Fassungen der Forschungsklauseln der Landesgesetze mit z. T. restriktiven Voraussetzungen bedeuten, die von der DSGVO in dieser Form nicht zwingend geboten sind. Inwiefern die Neuregelung des § 287a SGB V hier bereits eine Verbesserung bewirken kann, sollte in einem ersten Schritt ergebnisoffen geprüft werden. Sollte eine bundeseinheitliche Opt-out-Lösung für die Datenüberlassung eines darüberhinausgehenden gesetzlichen Erlaubnistatbestandes bedürfen, sollte diese Regelung auf die restriktiven Tatbestandsmerkmale verzichten, wie sie sich in manchen Forschungsklauseln auf Landesebene finden. Voraussetzung für eine europa- und grundrechtskonforme Lösung wäre in jedem Fall aber, dass technische und organisatorische Rahmenbedingungen definiert werden, die im Rahmen der Datenüberlassung für die Forschung die Datensouveränität der Patienten sicherstellen.

Zur Stärkung der patientenbasierten medizinischen Forschung – auch und gerade zu innovativen KI-Verfahren – empfiehlt das Forum daher zusammenfassend, für den medizinischen Fortschritt eine bundeseinheitliche zweckgebundene Datenüberlassung für die Forschung zu konzeptionieren und die hierfür erforderlichen gesetzlichen Voraussetzungen zu identifizieren und zu prüfen. Konzeptentwicklung und ethisch-rechtliche Bewertung sollten ggf. iterativ im Dialog mit verschiedenen Stakeholdern und unter Berücksichtigung der sich entwickelnden gesetzlichen Rahmenbedingungen erfolgen. Diese Aktivitäten könnten durch eine ethische, rechtliche und soziologische Begleitforschung flankiert und unterstützt werden.

V. Literatur

Kapitel I-III 1.-3.

1. Wissenschaftsrat: Zum Wandel in den Wissenschaften durch datenintensive Forschung | Positionspapier (Drs. 8667-20), Köln Oktober 2020, <https://www.wissenschaftsrat.de/download/2020/8667-20.html>
2. Deutsche Forschungsgemeinschaft: Digitaler Wandel in den Wissenschaften I Impulspapier, Oktober 2020, <http://digitalerwandel.dfg.de>
3. Hightech-Forum: Bio-IT-Innovationen - Konvergenz von Biowissenschaften und Informationstechnologien I Impulspapier, Oktober 2020 https://www.hightech-forum.de/wp-content/uploads/hft_discussion_paper_bio_it_innovations.pdf
4. Handlungsempfehlungen der TMF e. V. für die Arbeitsgruppe Gesundheit der Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“; 09.09.2019
5. Grundsatzpapier der Deutschen Hochschulmedizin: Mehrwert digitaler Daten für Versorgung und Forschung sicherstellen – Politik muss jetzt handeln; 06/2020
6. *Griffon N, Charlet J, Darmoni SJ.* Managing free text for secondary use of health data. Yearbook of medical informatics 2014; 9: 167-169 10.15265/iy-2014-0037.

Kapitel III 4. Förderung der Kultur des Datenteilens

7. *Borgman, Christine L.* (2012): The conundrum of sharing research data. In: Journal of the American Society for Information Science and Technology 63 (6), S. 1059–1078. DOI: 10.1002/asi.22634.
8. *Lo, Bernard; DeMets, David L.* (2016): Incentives for Clinical Trialists to Share Data. In: The New England journal of medicine 375 (12), S. 1112–1115. DOI: 10.1056/NEJMp1608351
9. *Fecher, Benedikt; Friesike, Sascha; Hebing, Marcel* (2015): What drives academic data sharing? In: PloS one 10 (2), e0118053. DOI: 10.1371/journal.pone.0118053.
10. *van Horn, John Darrell; Gazzaniga, Michael S.* (2013): Why share data? Lessons learned from the fMRIDC. In: NeuroImage 82. DOI: 10.1016/j.neuroimage.2012.11.010.
11. *Piwovar, Heather A.; Day, Roger S.; Fridsma, Douglas B.* (2007): Sharing Detailed Research Data Is Associated with Increased Citation Rate. In: PLOS ONE 2 (3), e308. DOI: 10.1371/journal.pone.0000308.
12. *Rowhani-Farid, Anisa; Allen, Michelle; Barnett, Adrian G.* (2017): What incentives increase data sharing in health and medical research? A systematic review. In: Res Integr Peer Rev 2 (1), S. 1–10. DOI: 10.1186/s41073-017-0028-9.
13. *Wallis, Jillian C.; Rolando, Elizabeth; Borgman, Christine L.* (2013): If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. In: PloS one 8 (7), e67332. DOI: 10.1371/journal.pone.0067332.
14. *Tenopir, Carol; Allard, Suzie; Douglass, Kimberly; Aydinoglu, Arsev Umur; Wu, Lei; Read, Eleanor et al.* (2011): Data sharing by scientists: practices and perceptions. In: PloS one 6 (6), e21101. DOI: 10.1371/journal.pone.0021101.

15. *Tenopir, Carol; Dalton, Elizabeth D.; Allard, Suzie; Frame, Mike; Pjesivac, Ivanka; Birch, Ben et al.* (2015): *Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide*. In: PloS one 10 (8), e0134826. DOI: 10.1371/journal.pone.0134826.
16. *Krahe, Michelle; Wolski, Malcolm; Toohey, Julie; Mickan, Sharon; Scuffham, Paul; Reilly, Sheena:* *Identifying Barriers and Enablers to Sharing Data Among Researchers: Designing Effective Behaviour Change Interventions*. eResearch Australia.
17. *Kellmeyer P., Mueller O., Feingold-Polak R. and Levy-Tzedek S.* "Social robots in rehabilitation: A question of trust". *Science Robotics*. 2018 Aug 15 3(21). <https://www.doi.org/10.1126/scirobotics.aat1587>

Kapitel III 5. Ethische und rechtliche Aspekte der Datennutzung

18. Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder (DSK), Beschluss zu den Einwilligungsdokumenten der Medizininformatik-Initiative des Bundesministeriums für Bildung und Forschung, 15.4.2020, https://www.datenschutzkonferenz-online.de/media/dskb/20200427_Beschluss_MII.pdf
19. Datenethikkommission, Gutachten, <https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.html>
20. Deutscher Ethikrat, Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung, Stellungnahme, 30.11.2017, <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf>.
21. Weichert/Krawczak, Vorschlag einer modernen Dateninfrastruktur für die medizinische Forschung in Deutschland, *GMS Med Inform Biom Epidemiol* 2019, Vol. 15(1), <https://dx.doi.org/10.3205/mibe000197>.
22. European Highlevel Expert Group on Artificial Intelligence; <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>
23. Budin-Ljøsne, H.J.A. Teare, J. Kaye, S. Beck, H.B. Bentzen, L. Caenazzo, C. Collett, F. D’Abramo, H. Felzmann, T. Finlay, M.K. Javaid, E. Jones, V. Katić, A. Simpson, D. Mascalonzi, Dynamic Consent: a potential solution to some of the challenges of modern biomedical research, *BMC Medical Ethics* 18(1) (2017) 4.
24. K.S. Steinsbekk, B. Kåre Myskja, B. Solberg, Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? *European journal of human genetics: EJHG* 21(9) (2013) 897-902
25. [66] Article 40, GDPR; Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679

Kapitel III 6. Patienteneinbindung bei der Forschung mit Gesundheitsdaten

26. *Betts, D, Gorenda, L,* A consumer-centered future of health. Deloitte Insights. Deloitte Center for Health Solutions. 2019
27. FORSA, TMF e. V., Deutliche Mehrheit der Deutschen bereit zur Datenspende für die medizinische Forschung, <http://www.tmf-ev.de/News/articleType/ArticleView/articleId/4456.aspx>, abgerufen am 17.02.2020
28. *ANDREJEVIC, Mark.* Big Data, Big Questions| The Big Data Divide. *International Journal of Communication*, [S.l.], v. 8, p. 17, jun. 2014. ISSN 1932-8036. Available at: <<https://ijoc.org/index.php/ijoc/article/view/2161>>.

Kapitel IV: Exkurs

1. Zulassungs- und haftungsrechtliche Aspekte KI-basierter Anwendungen in der Medizin

29. *Detting*: Künstliche Intelligenz und digitale Unterstützung ärztlicher Entscheidungen in Diagnostik und Therapie, PharmR 2019, 633
30. *Eichelberger*, Zivilrechtliche Haftung für KI und smarte Robotik, in: Ebers u.a. (Hrsg.), Rechts- handbuch Künstliche Intelligenz und Robotik, 2020, § 5 (S. 172-199)
31. *Eichelberger*, Arzthaftung bei automatisierten und autonomen Systemen, in: Chibanguza/Kuss/Steege (Hrsg.), Handbuch Künstliche Intelligenz - Recht und Praxis automati- sierter und autonomer Systeme (in Vorbereitung für 2021)
32. *Katzenmeier*, Big Data, E-Health, M-Health, KI und Robotik in der Medizin, MedR 2019, 259
33. *Wagner*, Produkthaftung für autonome Systeme, AcP 217 (2017), 707
34. *Frost*, Künstliche Intelligenz in Medizinprodukten und damit verbundene medizintechnik- und da- tenschutzrechtliche Herausforderungen, MPR 2019, 117.
35. *Gassner*, MedTech meets M-Health, MPR 2015, 73.
36. *Gassner*: Software als Medizinprodukt – zwischen Regulierung und Selbstregulierung, MPR 2016, 109.
37. *Gassner/Modi*: Anmerkung zu EuGH, Urt. v. 07.12.2017 – C-329/16 (Syndicat national de l'indust- rie des technologies médicales (Snitem)/Premier ministre u. a.), EuZW 2018, 168.
38. *Gassner/Strobl*: E-Health, Health-Apps & Co. – rechtliche Aspekte, e-beratungsjournal.net 2019, 97.
39. *Gassner/Juknat*: Regulatory Approaches to AI in Medical Practise, ELPR 2019, 176.
40. *Gassner/Schreiegg*: Software als Medizinprodukt – Brüsseler Spitzen, MPR 2019, 198.
41. IMDRF: "Software as a Medical Device": Possible Framework for Risk Categorization and Corre- sponding Considerations vom 14.9.2014.
42. MDCG: Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MPVO and Regulation (EU) 2017/746 – IVDR vom 11.10.2019.

2. Perspektivische Empfehlung: Ein Plädoyer für die Datenüberlassung für die Forschung

43. *Richter G, Krawczak M, Lieb W, Wolff L, Schreiber S, Buyx A* (2018) Broad consent for healthcare-embedded biobanking: understanding and reasons to donate in a large patient sam- ple. Genet Med 20: 76-82.
44. *Richter G, Borzikowsky C, Lieb W, Schreiber S, Krawczak M, Buyx A* (2019) Patient view on re- search use of clinical data without consent: Legal, but also acceptable? Eur J Hum Genet 27: 841–847.
45. <https://www.tmf-ev.de/News/articleType/ArticleView/articleId/4456.aspx>