



Transfer von Digital-Health- Anwendungen in den Versorgungsalltag

Teil 4: Wirksamkeitsnachweis und Nutzenbewertung – Kontext,
Methoden und Integration in die agile Produktentwicklung

In der Analyse „Transfer von Digital-Health-Anwendungen in den Versorgungsalltag“ fragt die Bertelsmann Stiftung danach, wie der Prozess des Transfers ausgestaltet ist, welche Hürden einem effektiven Transfer entgegenstehen und was getan werden müsste, um diese Hürden zu überwinden. Basis der Analyse ist ein umfassendes, idealtypisches Transfermodell für Digital-Health-Anwendungen – von der Idee bis zum „Betrieb“ eines Angebots im 1. Gesundheitsmarkt – sowie die Beschreibung der Hürden im Prozess. Unter Digital-Health-Anwendungen werden dabei sogenannte Gesundheits-Apps oder Online-Anwendungen verstanden, die für die primäre Nutzung durch Bürger angeboten werden.

Die Analyse besteht aus sechs verschiedenen Bausteinen, die jeweils nach Fertigstellung in Form von Teilberichten veröffentlicht werden. In den Bausteinen der Analyse werden Vorschläge für Verbesserungen der Rahmenbedingungen und Verfahren erarbeitet – immer bezogen auf identifizierte Hürden:

1. Transfermodell, Varianten und Hürden
2. Innovations- und Forschungsförderung
3. Medizinproduktezertifizierung
- 4. Nutznachweis & Nutzenbewertung**
5. Vertrags- und Vergütungsformen
6. Markt- und Qualitätstransparenz

Detaillierte Informationen zum Gesamtvorhaben zu den Arbeitspaketen sowie die zur Verfügung stehenden Teilberichte zum Download finden sich unter www.der-digitale-patient.de/digital-health-transfer

Transfer von Digital-Health-Anwendungen in den Versorgungsalltag

**Teil 4: Wirksamkeitsnachweis und Nutzenbewertung
Kontext, Methoden und Integration in die agile Produktentwicklung**

Karsten Knöppler, Sebastian Hesse, Patricia Ex

März 2018

Zusammenfassung

Bei der Einführung neuer Produkte, Verfahren und Leistungen im Gesundheitswesen sind die Wirksamkeit und der Nutzen essenzielle Kriterien bei Entscheidungen über deren Einsatz. Das gilt sowohl für individuelle Behandlungsentscheidungen in der Patientenberatung und medizinischen Versorgung als auch bei Systementscheidungen hinsichtlich der Zulassung, Zertifizierung und Vergütung. In einem solidarischen Gesundheitswesen sind Wirksamkeit und Nutzen zudem zentrale Grundlagen für die ausgleichende Steuerung der Qualität der Versorgung, der Absicherung der ökonomischen Tragfähigkeit und der Präferenzgerechtigkeit aus Perspektive der Patienten.

Für Leistungserbringer, Anbieter und Hersteller stellt sich damit die Frage, wie für neue Angebote die Wirksamkeit belegt und der Nutzen als Grundlage für die Preisfindung bewertet werden kann und muss. Bei den klassischen Leistungen von ärztlichem oder nicht ärztlichem Fachpersonal, aber auch der Arzneimittel und Medizinprodukte, gibt es hierfür weitgehend standardisierte oder zumindest etablierte Verfahren. Dieser regulatorische Rahmen gilt im Prinzip auch für Digital-Health-Anwendungen, die mindestens aus der Perspektive der Zertifizierung bzw. Zulassung den Medizinprodukten zuzuordnen sind. Bei der Vergütung ergibt sich hier jedoch mehr Interpretationsspielraum für den zugrunde zu legenden regulatorischen Rahmen und somit auch für die Methoden beim Nachweis der Wirksamkeit und der Bewertung des Nutzens.

Digital-Health-Anwendungen unterscheiden sich in verschiedener Hinsicht von Arzneimitteln oder Medizinprodukten. Die bislang etablierten, klassischen Studientypen und Methoden sind aufgrund der besonderen Eigenschaften und Anforderungen dieser Anwendungen nur bedingt geeignet. So sind die Anwendungen häufig im Kern Prozessinnovationen, wie beispielsweise Neue Versorgungsformen, werden aber als Produkte (z. B. Medizinprodukte) zertifiziert und angeboten. Sie können somit auch als hybride Angebote aus Produkten und Dienstleistungen – sogenannte Lösungen – bezeichnet werden. Zudem nehmen die Prozessinnovationen

neben den Versorgungsprozessen der Leistungserbringer vor allem das Gesundheitshandeln der Patienten (in Interaktion mit ihrem Umfeld) in den Blick. Hinzu kommt: Die Produktlebenszyklen und Release-Zyklen sind im Gegensatz zu klassischen Medizinprodukten oder Arzneimitteln kurz, das heißt, es findet eine kontinuierliche, iterative Weiterentwicklung statt. Zentrale Anforderung ist somit die Kompatibilität der Studientypen und Methoden mit einem kurzen Produktlebenszyklus, einem iterativen Vorgehen in der Produktentwicklung und der Evaluation einzelner Komponenten sowie ihrer Kombinationen.

Ziel dieses Teilberichts ist es, den klassischen, etablierten Methoden und Verfahren für den Wirksamkeitsnachweis und die Nutzenbewertung, die auch für Digital Health relevant sind, neue Methoden gegenüberzustellen und Implikationen für die Sicherung der Innovationsfähigkeit am Gesundheitsstandort Deutschland abzuleiten. Die klassischen Methoden sind vor allem mit Prozessinnovationen, die iterativ und in kurzen Release-Zyklen in Verkehr gebracht werden, und mit kurzen Produktlebenszyklen wenig kompatibel. Zudem werden neue Möglichkeiten der Probanden- und Datengewinnung im Real-Life-Setting bislang kaum genutzt.

Kontext

Insbesondere neue Produkte und Verfahren stehen vor der Herausforderung, eine Verbesserung gegenüber den bestehenden nachweisen zu müssen, um sich am sogenannten ersten Gesundheitsmarkt etablieren zu können. Digital-Health-Anwendungen stehen in dem Ruf, die klassischen Versorgungssysteme revolutionieren zu können. Umso wichtiger ist es, diese neuen Ansätze mit einer angemessenen Forschung zu begleiten und ihre Wirksamkeit im Vergleich zu bestehenden Verfahren transparent zu machen.

Die möglichen Studien können zunächst hinsichtlich ihrer primären Zielsetzung unterschieden werden: Ziel kann hier

typischerweise die Untersuchung der Risiken (Risk), der Wirksamkeit (Efficacy) und/oder des Nutzens (Effectiveness) sein. Um die Risiken und die Wirksamkeit zu untersuchen, werden meist Interventionsstudien durchgeführt, die in der Regel als klinische Studien oder RCT-Studien (Randomized Controlled Trial) bezeichnet bzw. ausgeprägt werden. Die Bewertung des Nutzens hingegen erfolgt gewöhnlich im Rahmen einer Studie zur Technologiefolgeabschätzung (Health Technology Assessment – HTA).

Im deutschen Gesundheitswesen sind hierfür verschiedene Akteure zuständig. Die Zuständigkeiten leiten sich weitgehend vom regulatorischen Rahmen ab. Dieser besteht zum einen aus der Medizinproduktezulassung, zum anderen aus dem Vergütungsrahmen der gesetzlichen Krankenversicherung. Zu beachten sind auf der nicht regulierten Ebene zudem die Fachgesellschaften in ihrer Rolle des Verfassens und Veröffentlichens medizinischer Leitlinien. In der Praxis steht als erster Schritt die Medizinproduktezertifizierung an, dann die Information der Fachgesellschaften im Kontext der Leitlinienerstellung und abschließend die Verhandlung mit Kostenträgern über die Vergütung. Im Allgemeinen werden bislang jedoch beide Methoden – Interventionsstudien und Studien zur Technologiefolgeabschätzung – im Anschluss an die Produktentwicklung angewendet.

Methoden

Anhand der Forschungsdimensionen lässt sich der Unterschied zwischen den einzelnen Methoden verdeutlichen. Während Interventionsstudien auf Basis eigener Primärdaten im Wesentlichen die klinische Evidenz, die Wirksamkeit und die Sicherheit betrachten, haben HTAs meistens einen deutlich erweiterten Fokus, der auch die Wirtschaftlichkeit sowie ethische, soziale, rechtliche und organisatorische Aspekte einschließen kann. Der entscheidende Unterschied ist jedoch die Nutzung von Sekundärdaten, u. a. Studienergebnisse und GKV-Routinedaten.

Bei den Interventionsstudien dominieren heute die RCTs. Sie stehen für das höchste Evidenzniveau. Im Kontext von Digital Health haben sich in den vergangenen Jahren jedoch diverse neue Studiendesigns entwickelt, die den Anforderungen von Digital Health besser gerecht werden:

- Continuous Evaluation of Evolving Behavioral Intervention Technologies (CEEBIT)
- Multiphase Optimization Strategy (MOST)
- Sequential Multiple Assignment Randomized Trial (SMART)

Micro-Randomized Trial

Die wesentlichen Unterschiede dieser neuen Methoden liegen darin, dass die Datenerhebung im „Real-Life-Setting“ erfolgt, der Erhebungszeitpunkt iterativ beispielsweise mit den Zyklen der Produktentwicklung erfolgen kann sowie einzelne Komponenten als auch ihre Kombinationen getestet werden können. Insgesamt fließen die Ergebnisse kontinuierlich in die Produktentwicklung mit ein. Randomisierung und Kontrollgruppen sind hierbei in der Regel vorgesehen.

Auch bei den HTAs wurde in den vergangenen Jahren eine neue Variante entwickelt:

Prospektives Health Technology Assessment

Der wesentliche Faktor des prospektiven HTA ist, dass er nicht nach der Produktentwicklung und der Interventionsstudie angewendet wird, sondern als Ausgangspunkt der Produktentwicklung. In diesem Fall werden hinsichtlich der Wirksamkeit der Intervention zunächst Studienergebnisse zu alternativen Behandlungsoptionen sowie Annahmen herangezogen und in einer Modellrechnung zusammengeführt, die auch Komponenten und Kombinationen berücksichtigen kann. Diese Methode kann zudem schrittweise mit den Ergebnissen aus iterativen Interventionsstudien gefüllt werden.

Agile Arbeitsmodelle der Anbieter

Neben den hybriden Produkteigenschaften von Digital Health liegen weitere Ursachen für deren disruptiven Charakter jedoch tiefer und auch in der Organisation der Anbieter. Diese arbeiten in der Produktentwicklung überwiegend agil und gehen nicht nach klassischen, sequenziellen Modellen vor.

Das klassische Vorgehen in der Produktentwicklung zeichnet sich dadurch aus, dass beispielsweise eine Produktdefinition, eine Marktanalyse, eine ausführliche Konzeption, die Entwicklung, die Testung und das Qualitätsmanagement, die Zulassung bzw. Zertifizierung und die Markteinführung von separaten Teams und sequenziell, also nacheinander, durchgeführt werden.

In agilen Modellen verorten Anbieter diese Prozessschritte in integrierten Teams, die viele der genannten Schritte aus Analyse, Konzeption, Entwicklung, Testung, Qualitätssicherung gleichzeitig umsetzen: Das Gesamtvorhaben wird in Module

gegliedert und in einzelnen Bearbeitungszeiträumen (Sprints) wird jeweils ein Modul bis zur fertigen Software entwickelt.

Diese innovative Arbeitsweise stößt nun bei der Markteinführung auf die überwiegend klassisch arbeitenden Organisationsstrukturen und Verfahren des Gesundheitswesens. Denn die anbieterinterne agile Produktentwicklung wurde bisher noch nicht in die Evaluationsprozesse ausgewählter Forschungsinstitute oder in die Verhandlungsprozesse streng hierarchisch organisierter Krankenkassen integriert.

In der Medizinproduktezertifizierung und vor allem im Bereich der Software sind bereits erste Methoden zur fortwährenden, iterativen Durchführung von Konformitätserklärungen in der Gesetzgebung angelegt: Durch die Etablierung eines Qualitätsmanagementsystems kann der Anbieter die Konformität selbst kontinuierlich und im Einklang mit eigenen agilen Entwicklungs- und Release-Zyklen absichern und erklären. Das ist jedoch bei der Vergütung und insbesondere beim Nachweis der Wirksamkeit sowie der Bewertung des Nutzens noch nicht etabliert – weder bei Anbietern noch bei den zuständigen Spitzenorganisationen der gesetzlichen Krankenversicherung (GKV) oder bei den einzelnen Krankenkassen.

Implikationen

Im Folgenden geht es um Handlungsempfehlungen, die es in der Summe ermöglichen, die Begleitforschung für den Nachweis der Wirksamkeit und des Nutzens auf agile Prinzipien umzustellen. Hierfür ist eine enge organisatorische Verzahnung erforderlich, die über die Organisationsstruktur des Anbieters hinausgeht. Sie muss die Organisation der Kunden, das heißt in der Regel der Krankenkassen und begleitenden Forschungseinrichtungen, mit beinhalten. Im Weiteren sind aber auch entsprechende Organe der GKV einzubeziehen, die für die Aufnahme von Leistungen in den GKV-Katalog verantwortlich sind. Sie müssen Ergebnisse aus agilen Studiendesigns für den Nachweis der Wirksamkeit und der Bewertung des Nutzens – hier: Methoden bei der Technologiefolgeabschätzung – akzeptieren. Im Idealfall werden sie bereits zu Beginn der Produktentwicklung in Abstimmung mit Interventionsstudie und Technologiefolgeabschätzung einbezogen; der Freigabeprozess für die Vergütung, wie etwa die Nutzenbewertung durch den G-BA, startet gleichzeitig.

Insgesamt ist im Zusammenspiel aller relevanten Akteure eine grundsätzlich neue Verfahrensweise erforderlich, die sich in bestimmten Punkten an den Werten und Grund-

prinzipien des agilen Vorgehens orientiert. Wichtig ist, die Bewertung des Nutzens und den Nachweis der Wirksamkeit in den Prozess der Produktentwicklung zu integrieren. Zudem sollte die Zusammenarbeit mit Kunden (Krankenkasse) sowie einer Forschungseinrichtung für die Interventionsstudie (auch die Technologiefolgeabschätzung) bereits in dieser Phase erfolgen.

Im Folgenden wird das Vorgehen mittels eines Sets an Handlungsempfehlungen operationalisiert und nach Methoden bzw. relevanten Akteuren geordnet.

Handlungsempfehlung 1 (an Hersteller und Forschungseinrichtungen): Neue, agile Designs für Interventionsstudien als integralen Bestandteil der Produktoptimierung anwenden – anstelle nachgelagerter, klassischer RCTs

Mögliche agile Studiendesigns zur Messung der Wirksamkeit sind u. a. MOST, CEEBIT, SMART oder Micro-Randomized Trials (siehe oben). Sie sind geeignet, entlang der Release-Zyklen eine Ergebnisorientierung auch wissenschaftlich zu fundieren.

Die Methoden sollten mit den kurzen Entwicklungszyklen und den agilen Entwicklungsmethoden von Digital-Health-Anwendungen synchronisiert werden. Dies kann über eine Reflexion der Ergebnisse entlang der Release- und Sprintplanung erfolgen. Geachtet werden sollte auf eine regelmäßige Lieferung von Zwischenergebnissen zu einzelnen Modulen und Modulkombinationen an die Produktentwicklung und zur Aktualisierung der Nutzenbewertung.

Die kontinuierliche Ergebnismessung liefert – im Gegensatz zu klassischen Methoden – bereits während der laufenden Entwicklung Feedback, schützt vor Fehlentwicklungen und hilft trotz anfänglichem Mehraufwand, die begrenzten Ressourcen von Start-ups optimiert einzusetzen.

Handlungsempfehlung 2 (an Hersteller und Forschungseinrichtungen): Neue Kanäle zur Rekrutierung von Studienteilnehmern nutzen und Anwendungen im Real-Life-Setting erproben

Digitale Vertriebskanäle wie App-Stores, Apps und soziale Netzwerke sollten genutzt werden, um – im Gegensatz zur klassischen Ansprache von Studienprobanden über feste Studienzentren – einen ortsungebundenen Einschluss von Patienten zu ermöglichen.

Damit geht auch eine Erprobung der Anwendungen im Real-Life-Setting einher. Die Wirksamkeit unter Laborbedingungen zu erproben, wie in klassischen Studien, wird hier übersprungen, da die Digital-Health-Anwendungen ohnehin über die Beeinflussung des Gesundheitshandeln stark mit dem Real-Life-Setting interagieren. Diese Effekte sollten bereits im Studiendesign mit berücksichtigt und gemessen werden.

Die Studienteilnehmer über digitale Vertriebskanäle anzuwerben, kann Aufwände für die Rekrutierung verringern.

Handlungsempfehlung 3 (an Hersteller und Forschungseinrichtungen): Daten über die digitalen Anwendungen erheben

In klassischen Studiendesigns werden die Daten im Allgemeinen durch medizinisches Fachpersonal erhoben. Üblicherweise sind zwei bis fünf Messpunkte geplant, die über relativ weite Beobachtungszeiträume verteilt sind.

Neue digitale Instrumente wie das Apple ResearchKit sowie Tracking- und Aktenlösungen ermöglichen unter Einbezug der Sensoren der mobilen Endgeräte, die Probanden effizient und engmaschig zu beobachten. Es können sehr viel engere Zusammenhänge zwischen der Intervention und den Effekten hergestellt werden.

Diese Möglichkeiten sollten – auch um den Aufwand bei der Datenerhebung zu reduzieren – systematisch genutzt werden und bei der Ausgestaltung der Studiendesigns mit einfließen.

Handlungsempfehlung 4 (an Hersteller, Forschungseinrichtungen, Krankenkassen, G-BA): Prospektive Technologiefolgeabschätzung als Ausgangspunkt und integralen Bestandteil der Produktentwicklung einsetzen

Bei der klassischen Herangehensweise folgt die Technologiefolgeabschätzung im Anschluss an den Wirksamkeitsnachweis, zum Beispiel durch eine RCT-Interventionsstudie. Diese dient primär dazu, bei der Vergütungsverhandlung in der GKV Preise zu bestimmen.

Es ergibt sich aber erhebliches Potenzial, wenn die Erkenntnisse aus der Technologiefolgeabschätzung als Ausgangspunkt für die Produktentwicklung genutzt werden. Hierzu kann das Grundgerüst eines klassischen HTA schon in der frühen Phase der Produktentwicklung aufgesetzt werden. Die Wirksamkeit von vergleichbaren Inter-

ventionen liefert wichtige Anhaltspunkte für den Zielbereich und für die Wirksamkeit der jeweiligen Lösung.

Die zunächst fehlenden Daten und Informationen zur Wirksamkeit der eigenen Lösung können in einer Modellrechnung zum Nutzen ersetzt werden durch Annahmen. Diese sind so weit wie möglich an Vergleichswerten aus Studiergebnissen orientiert, die die Wirksamkeit vergleichbarer Produkte und/oder in dem Indikationskontext üblicher Produkte und Verfahren nachweisen. Diese „Grundbefüllung“ mit Wahrscheinlichkeiten und anderen Werten kann schrittweise mit jeder Iteration der eigenen Wirksamkeitsstudien ersetzt werden. So kann bei jedem Turnus die stringente Optimierung auf den Nutzen umgesetzt werden.

Handlungsempfehlung 5 (an Hersteller, Forschungseinrichtungen, Fachgesellschaften und Krankenkassen): Entwicklungspartnerschaften eingehen und neue Verfahren für Wirksamkeit und Technologiefolgeabschätzung akzeptieren

Bei der Produktentwicklung sollten Krankenkassen, Forschungseinrichtungen und Fachgesellschaften stärker als bisher mit einbezogen werden. Ihre Rolle kann u. a. sein:

- Beratung bei der Konzeption und Entwicklung hinsichtlich der Produkthanforderungen
- Beratung zur Methodik der Technologiefolgeabschätzung und Bereitstellung von Datengrundlagen (GKV-Routinedaten) für die prospektive Technologiefolgeabschätzung hinsichtlich der Optimierung des Nutzens
- Beratung zum Studiendesign für agile Interventionsstudien und zu Zwischenergebnissen hinsichtlich der Optimierung der Wirksamkeit und Präferenzgerechtigkeit

Handlungsempfehlung 6 (an den G-BA): Erprobungsrichtlinie § 137e und § 137h SGB V für den ambulanten und stationären Sektor weiterentwickeln und ein frühes, iteratives Verfahren zur Nutzenbewertung ermöglichen

Bei den Verfahren zur Methodenbewertung beim Gemeinsamen Bundesausschuss (G-BA) im Rahmen des § 137e SGB V und des § 137h SGB V sollten neue, agile Studiendesigns berücksichtigt werden. Dies gilt insbesondere bei der Bewertung des Potenzials einer neuen Untersuchungs- und Behandlungsmethode nach § 137e Abs. 1 SGB V und § 137h Abs. 1 SGB V, wenn diese Methode auf einer Digital-Health-Anwendung beruht. Die innovativen Studiendesigns sollten auch bei der Durchführung von Erprobungsstudien berücksichtigt werden. Damit wird die Basis für den Zugang

von Digital-Health-Anwendungen zum ersten Gesundheitsmarkt erweitert. Hersteller von neuen, agilen Anwendungen haben so bessere Marktzugangschancen, ohne dass beim Nachweis des Nutzens für die Versicherten und Patienten Abstriche gemacht werden müssen.

Handlungsempfehlung 7 (an Politik und Selbstverwaltung):
Forschungseinrichtungen für die Anwendung und Weiterentwicklung von agilen Interventionsstudien und HTAs in Kombination fördern

In Teil 2 dieser Reihe – „Bedarfsgerechte Innovations- und Forschungsförderung“ – wurde bereits der Förderbedarf im Bereich Digital Health ausführlich beleuchtet; der Fokus lag auf einem Programm für eine bedarfsgerechte Innovations- und Forschungsförderung. Das Programm sieht eine Förderung der Begleitforschung für Digital Health vor. Die Methodik für die Begleitforschung kann diesem Teilbericht 4 entnommen werden. Sie umfasst die oben genannten Methoden zu agilen Interventionsstudien und prospektiver Technologiefolgeabschätzung. Hierzu sollten Anbieter und Forschungseinrichtungen gezielt gefördert werden. Eine Etablierung dieser Methoden unterstützt die Innovations- und Wettbewerbsfähigkeit am Standort Deutschland im Bereich Digital Health.

Inhalt

1	Einleitung	10
1.1	Ausgangslage	10
1.2	Zielsetzung und Vorgehen	11
2	Kontext: Wirksamkeitsnachweis und Nutzenbewertungen	12
2.1	Abgrenzung: Risiken, Wirksamkeit und Nutzen	12
2.2	Institutioneller Kontext Medizinproduktregulierung und Vergütung	14
2.3	Evidenzklassen für Studien zur Wirksamkeit	18
2.4	Bedarf an Nachweis der Wirksamkeit und Nutzenbewertung bei Digital-Health-Anwendungen	19
3	Wirksamkeitsnachweis	21
3.1	Klassische Studiendesigns	21
3.2	Limitationen klassischer Methoden	23
3.3	Neue Methoden	24
3.4	Gegenüberstellung der Methoden	27
4	Nutzenbewertung	30
4.1	Klassisches Health Technology Assessment	30
4.2	Limitationen klassischer HTA	32
4.3	Prospective Health Technology Assessment	32
4.4	Gegenüberstellung der Methoden	33
5	Datentypen, Datengewinnung und Setting	35
5.1	Datentypen	35
5.2	Datenerhebung	36
5.3	Setting	37
6	Implikationen	39
7	Literatur	45
8	Anhang	48
	Autoren und Autorin	49
	Impressum	51

1 Einleitung

1.1 Ausgangslage

Bei der Einführung neuer Produkte, Verfahren und Leistungen im Gesundheitswesen sind die Wirksamkeit und der Nutzen essenzielle Kriterien bei Entscheidungen über deren Einsatz. Das gilt sowohl für individuelle Behandlungsentscheidungen im Kontext der Patientenberatung und medizinischen Versorgung als auch bei Systementscheidungen hinsichtlich der Zulassung, Zertifizierung und Vergütung. Für ein solidarisches Gesundheitswesen sind Wirksamkeit und Nutzen auch zentrale Grundlagen für die ausgleichende Steuerung der Qualität der Versorgung, der Absicherung der ökonomischen Tragfähigkeit und u. a. der Präferenzgerechtigkeit aus Perspektive der Patienten.

Für Leistungserbringer, Anbieter und Hersteller stellt sich damit die Frage, wie für neue Angebote die Wirksamkeit belegt und der Nutzen neuer Angebote als Grundlage für die Preisfindung bewertet werden kann und muss. Bei den klassischen Leistungen von ärztlichem oder nicht ärztlichem Fachpersonal, aber auch der Arzneimittel und Medizinprodukte, gibt es hierfür weitgehend standardisierte oder zumindest etablierte Verfahren. Dieser regulatorische Rahmen gilt im Prinzip auch für Digital-Health-Anwendungen, die mindestens aus der Perspektive der Zertifizierung bzw. Zulassung den Medizinprodukten zuzuordnen sind (siehe auch Teil 3: „Medizinproduktezertifizierung“, Bertelsmann Stiftung 2017b). Bei der Vergütung ergibt sich hier jedoch mehr Interpretationsspielraum für den zugrunde zu legenden regulatorischen Rahmen und somit auch für die Methoden beim Nachweis der Wirksamkeit und der Bewertung des Nutzens.

Die Wirksamkeit und der Nutzen einer Intervention beziehen sich darauf, inwieweit Betroffene von einer zugelassenen Intervention gesundheitlich profitieren. Im Gesundheitswesen haben sich hinsichtlich klassischer Arzneimittel oder Medizinprodukte eine Reihe von Studientypen etabliert, die die Wirksamkeit und den Nutzen messen bzw. bewerten. Dies sind beispielsweise klinische Studien, die bei der Zulassung oder Zertifizierung neue Produkte mit einem Fokus auf Sicherheit und Wirksamkeit (im Sinne der Funktionsweise) für Patienten in der Regel unter Laborbedingungen untersuchen. Zudem sind es Studien zur Nutzenbewertung, die vor allem den Gesundheitssystemeffekt unter Alltagsbedingungen bewerten und auf den oben genannten Studien aufbauen. So soll die Nutzenbewertung auch Entscheidungshilfe für Erstattungsfragen sein und differiert methodisch je nach Vergütungsform.

Digital-Health-Anwendungen unterscheiden sich in verschiedener Hinsicht von Arzneimitteln oder Medizinprodukten (siehe auch Teil 1: „Transfermodell, Varianten und Hürden“, Bertelsmann Stiftung 2016). Die etablierten, klassischen Studientypen und Methoden sind aufgrund der besonderen Eigenschaften und Anforderungen von Digital-Health-Anwendungen nur bedingt für diese geeignet. Die besonderen Eigenschaften dieser Anwen-

dungen bestehen darin, dass sie im Kern oft Prozessinnovationen sind, wie beispielsweise Neue Versorgungsformen, aber als Produkte zertifiziert und angeboten werden, wie etwa Medizinprodukte. Sie können somit auch als hybride Angebote aus Produkten und Dienstleistungen – sogenannte Lösungen – bezeichnet werden. Zudem haben die Prozessinnovationen neben den Versorgungsprozessen der Leistungserbringer vor allem das Gesundheitshandeln der Patienten im Blick. Die Produktlebenszyklen und Release-Zyklen sind im Gegensatz zu klassischen Medizinprodukten oder Arzneimitteln kurz, das heißt, es findet eine kontinuierliche, iterative Weiterentwicklung statt (Bertelsmann Stiftung 2016).

1.2 Zielsetzung und Vorgehen

Ziel dieses Teilberichts ist es, den klassischen, etablierten Methoden und Verfahren für den Wirksamkeitsnachweis und die Nutzenbewertung, die auch für Digital Health relevant sind, neue Methoden gegenüberzustellen und Implikationen für die Sicherung der Innovationsfähigkeit am Gesundheitsstandort Deutschland abzuleiten. Die klassischen Methoden sind vor allem mit Prozessinnovationen, die iterativ und in kurzen Release-Zyklen in Verkehr gebracht werden, und mit kurzen Produktlebenszyklen wenig kompatibel. Zudem werden neue Möglichkeiten der Probanden- und Datengewinnung im Real-Life-Setting bislang wenig genutzt.

In dieser Analyse wird zunächst der Kontext für die klassischen Methoden untersucht. Dies betrifft die Methoden zum Nachweis der Wirksamkeit und zur Bewertung des Nutzens von Medizinprodukten sowie deren institutionelle Einordnung im deutschen Gesundheitswesen. Genannt seien hier u. a. der Gemeinsame Bundesausschuss (G-BA) einschließlich des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), der Medizinische Dienst des Spitzenverbandes Bund der Krankenkassen (MDS) sowie einzelne Krankenkassen. Auch werden Grundlagen wie die Evidenzklassen herausgearbeitet sowie der Bedarf von Digital Health an Wirksamkeitsnachweis und Nutzenbewertung nach Anwendungstypen.

Im methodischen Teil werden die klassischen den neuen Methoden gegenübergestellt, die in den vergangenen Jahren u. a. im Kontext von Digital Health entwickelt wurden. Die Limitationen insbesondere klassischer Methoden bei der Anwendung im Bereich Digital Health werden herausgestellt; gleichzeitig geht es um die wesentlichen methodischen Innovationen neuer Ansätze und deren Kompatibilität mit den spezifischen Digital-Health-Anforderungen, wie etwa schnellen Innovationszyklen.

In der Synthese wird ein neues, konsolidiertes Verfahren für den Nachweis und die Optimierung der Wirksamkeit sowie die Bewertung des Nutzens skizziert. Diese ist mit den agilen Entwicklungsmethoden von Digital Health kompatibel. Implikationen in Form von Handlungsempfehlungen für die Etablierung der Methode im deutschen Gesundheitswesen bilden den Abschluss.

2 Kontext: Wirksamkeitsnachweis und Nutzenbewertungen

2.1 Abgrenzung: Risiken, Wirksamkeit und Nutzen

Ein Grundsatz des deutschen Gesundheitswesens lautet, medizinische Entscheidungen – soweit möglich – auf der Basis von empirisch nachgewiesener Wirksamkeit zu treffen (Evidence based Medicine – EbM). Dieser Grundsatz wird sowohl bei der Erarbeitung von evidenzbasierten Leitlinien, bei der Zulassung und Zertifizierung von Produkten und Leistungen als auch bei deren Vergütung berücksichtigt. Hierzu werden Erkenntnisse der wissenschaftlichen sowie leistungs- und produktbezogenen Forschung als Grundlage für Therapieentscheidungen in die Behandlung der Patienten integriert (vgl. Baethge 2014).

Dafür sind wissenschaftliche, empirische Untersuchungen bzw. Studien zu relevanten Therapien notwendig, die Aufschluss über deren Wirksamkeit geben. Sie ermöglichen es, soweit sie nach standardisierten bzw. vergleichbaren Methoden durchgeführt wurden, bei Therapieentscheidungen verschiedene Optionen zu vergleichen.

Das gilt vor allem auch für Arzneimittel, Medizinprodukte und Verhaltensinterventionen. Insbesondere neue Produkte und Verfahren stehen vor der Herausforderung, eine Verbesserung gegenüber den bestehenden nachweisen zu müssen, um sich am Markt etablieren zu können. Digital-Health-Anwendungen stehen in dem Ruf, das Potenzial zu haben, die klassischen Versorgungssysteme zu revolutionieren (vgl. Pham, Wiljer und Cafazzo 2016). Umso wichtiger ist es, diese neuen Ansätze mit einer angemessenen Forschung zu begleiten und deren Wirksamkeit in Relation zu bestehenden Verfahren transparent zu machen.

Die möglichen Studien können zunächst hinsichtlich der primären Zielsetzung unterschieden werden: Ziel kann hier typischerweise die Untersuchung der Risiken (Risk), der Wirksamkeit (Efficacy) und/oder des Nutzens (Effectiveness) sein. Um die Risiken und die Wirksamkeit zu untersuchen, werden meist Interventionsstudien durchgeführt, die in der Regel als klinische Studien oder RCT-Studien (Randomized Controlled Trial) bezeichnet bzw. ausgeprägt werden. Die Bewertung des Nutzens hingegen erfolgt gewöhnlich im Rahmen einer Studie zur Technologiefolgeabschätzung (Health Technology Assessment – HTA). Die verschiedenen Studientypen werden im Folgenden definiert.

2.2.1 Studien zum Nachweis von Risiken und Wirksamkeit (Interventionsstudien)

Interventionsstudien, die die Wirksamkeit neuer Produkte belegen, werden u. a. bei der Zulassung von Arzneimitteln und teils auch bei der Zertifizierung von Medizinprodukten gesetzlich vorgeschrieben. Der regulatorische Rahmen wird hierbei auf EU-Ebene gesetzt.

Während er für Arzneimittel strenge Vorgaben macht, sind diese für Medizinprodukte verhältnismäßig unkonkret.

Digital-Health-Anwendungen fallen unter die Medizinprodukteregulierung. Hier gilt die Besonderheit, dass die Anforderungen nicht für alle Medizinprodukte in gleicher Weise gelten, sondern bestimmte Produkte je nach Risikoklasse von einer Anforderung befreit sein können.

Mit Interventionsstudien kann und muss (bei bestimmten Produkttypen bzw. Risikoklassen) im Kontext der Medizinproduktezertifizierung untersucht werden, ob eine neue Technologie sicher ist (Risiken) und ob sie unter normalen Bedingungen wirkt wie angenommen (Wirksamkeit) (SVR-Gutachten 2014, Punkt 59). Die Studien erfolgen in der Regel unter Laborbedingungen, um vor dem Inverkehrbringen Risiken auszuschließen und die grundsätzliche Wirksamkeit nachzuweisen.

Die Zertifizierung von Medizinprodukten mit Vergabe eines CE-Kennzeichens ist auf europäischer Ebene gesetzlich geregelt, wobei die Notwendigkeit und die konkreten Anforderungen an die Studien vom Risiko eines Produktes abhängen. Die Logik der Einordnung von Digital-Health-Anwendungen in die Risikoklassen der Medizinprodukteregulierung ist in Teil 3 dieser Reihe, „Medizinproduktezertifizierung“, genauer ausgeführt (vgl. Bertelsmann Stiftung 2017b). Laut Sauerland (2017) lässt sich insgesamt feststellen, dass im Zuge der Reform der Medizinprodukterichtlinie die klinischen Anforderungen an den Marktzugang neuer Produkte mit einer hohen Risikoklasse verbessert wurden und künftig aussagekräftige vergleichende klinische Studiendaten vorgelegt werden müssen.

2.1.2 Studien zur Bewertung des Nutzens (Technologiefolgeabschätzung)

Der Nutzen einer Technologie ist im Anschluss an die Zertifizierung insbesondere für nationale Entscheidungen über die Erstattung durch die gesetzliche Krankenversicherung im Rahmen des Kollektivvertrages relevant und teilweise auch für Vergütungsregelungen einzelner Krankenkassen. In der Patientenversorgung bilden Nutzen und Nutznachweise zudem für die Therapieempfehlung und -wahl eine wichtige Grundlage.

Bei der Nutzenbewertung wird geprüft, inwieweit Betroffene von einer Intervention gesundheitlich profitieren (vgl. Windeler 2006). Die Methode zur Bewertung des Nutzens einer Technologie beispielsweise durch Gesundheitsbehörden heißt Technologiefolgeabschätzung (Health Technology Assessment – HTA).

Dabei wird das Ziel verfolgt, den tatsächlichen Nutzen einer Gesundheitstechnologie unter Alltagsbedingungen zu bewerten. Der Begriff des Nutzens kann in diesem Kontext verstanden werden als „Effekte einer Intervention, die in mehr als geringfügigem Ausmaß eine Verbesserung der Prognose und/oder der Symptomatik/Lebensqualität von Patienten“ (ebd.: 12) zur Folge haben. Dabei ist nicht nur die Frage entscheidend, wie eine Technologie wirkt, sondern vor allem, ob und in welchem Ausmaß Betroffene einen gesundheitlichen Vorteil durch die Intervention im Vergleich zu einer anderen oder keiner Intervention haben (vgl. Ex, Nüsse und Henschke 2016). Damit entspricht der Nutzen einer bestimmten Form der Wirksamkeit, nämlich der unter Alltagsbedingungen (Effectiveness), im Gegensatz zur Wirksamkeit unter Studienbedingungen (Efficacy), und mit direkter Relevanz für die Patienten (vgl. Windeler 2006).

Die Technologiefolgeabschätzung kann mit vielen verschiedenen Schwerpunkten durchgeführt werden, u. a. der Bewertung von Sicherheit, Wirksamkeit bzw. Nutzen, der Kostenwirksamkeit sowie ethischen, organisatorischen, sozialen und legalen Implikationen. De facto werden jedoch vor allem die medizinische Effektivität und Wirtschaftlichkeit untersucht (vgl. Ex, Busse und Henschke 2016). Im Wesentlichen werden die Studienergebnisse aus verschiedenen Interventionsstudien für die betrachtete Indikation relevanter Therapieformen genutzt und verglichen. Die Anwendung findet meist im Kontext der Vergütung statt. Diese ist auf nationaler Ebene geregelt und unterscheidet sich in Deutschland nach der jeweiligen Vertrags- bzw. Vergütungsform.

2.2 Institutioneller Kontext Medizinprodukte regulierung und Vergütung

Im deutschen Gesundheitswesen sind bezüglich der Studien zur Wirksamkeit und zur Technologiefolgeabschätzung verschiedene Akteure zuständig. Deren Zuständigkeiten leiten sich weitgehend vom regulatorischen Rahmen ab. Dieser besteht zum einen aus der Medizinprodukte regulierung (siehe Teil 3 dieser Reihe: „Medizinproduktezertifizierung“, Bertelsmann Stiftung 2017b) und dem Vergütungsrahmen der gesetzlichen Krankenversicherung (siehe Teil 5 dieser Reihe: „Vertrags- und Vergütungsformen in der GKV“, Bertelsmann Stiftung 2017c). Zu beachten sind zum anderen auf der nicht regulierten Ebene auch die Fachgesellschaften in ihrer Rolle beim Verfassen und Veröffentlichen medizinischer Leitlinien. In der Praxis steht als erster Schritt die Medizinproduktezertifizierung an, dann die Platzierung der Angebote bei den Fachgesellschaften im Kontext der Leitlinienerstellung und abschließend die Verhandlung mit Kostenträgern über die Vergütung. An dieser Stelle sei nochmals auf Teil 3 dieser Reihe verwiesen, in dem der institutionelle Rahmen der Medizinproduktezertifizierung detaillierter dargelegt ist.

Dieser Teilbericht fokussiert im Folgenden weiter den institutionellen Kontext im Zusammenhang mit der Vergütung, weil hier die höheren Anforderungen an die methodische Qualität gestellt werden. Die Rolle der Fachgesellschaften in Bezug auf die Leitlinienerstellung sei nur cursorisch erwähnt. Da die methodischen Anforderungen mit denen der Vergütung weitgehend gleichzusetzen sind und hier kein eigener regulatorischer Rahmen definiert ist, wird nicht weiter differenzierend darauf eingegangen.

Die wesentlichen Akteure bei der Medizinproduktezertifizierung sind die Hersteller, eine Forschungseinrichtung und die Benannte Stelle. Interventionsstudien oder Technologiefolgeabschätzungen zu einzelnen Produkten werden hier in der Regel zunächst von den Herstellern selbst initiiert. Anlass sind das Eigeninteresse und/oder die regulatorischen Anforderungen. Unabhängige Institute und Forschungseinrichtungen werden dabei gewöhnlich beauftragt, um Interventionsstudien und/oder Studien zur Technologiefolgeabschätzung durchzuführen. Die Benannte Stelle prüft bei der Medizinproduktezertifizierung in Abhängigkeit von der Risikoklasse die Ergebnisse der Interventionsstudien. Die Anforderungen an die methodische Qualität der Studien sind hier relativ gering. Bei der Leitlinienerstellung und der Vergütung sind diese deutlich höher und durch übergreifende Evidenzklassen definiert (siehe folgende Kapitel).

In der Praxis werden im Anschluss an die Medizinproduktezertifizierung – soweit relevant – die Fachgesellschaften über die Studienergebnisse informiert. Eine Aufnahme in Leitlinien kann die Argumentation eines Vergütungsanspruchs durch die GKV unterstützen. Dies wäre gegeben, wenn Angebote hier genannt bzw. empfohlen werden.

Bei den Vergütungsverhandlungen agieren über die Institutionen der Medizinproduktezertifizierung hinaus solche, die neben dem Hersteller eigenständig Gesundheitstechnologien bewerten. Dies können u. a. der G-BA, das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), der Medizinische Dienst der Krankenkassen (MDK) sowie Abteilungen bei der Bundesärztekammer und der Kassenärztlichen Bundesvereinigung sein.

Die konkrete methodische und verfahrenstechnische Ausgestaltung der Technologiefolgeabschätzung ist abhängig von der angestrebten Vergütungsform. Für die folgende Analyse wird unterschieden zwischen (1) der offiziellen Bewertung des G-BA unter Einbeziehung des IQWiG, (2) einzelnen Nutzenbewertungen durch den MDK als spezialisierte Institution im Interesse der Krankenkassen und (3) den Erstattungsentscheidungen einzelner Krankenkassen.

Der Zugang zur Vergütung durch Krankenkassen erfolgt in der Regel zunächst über selektivvertragliche bzw. Regelungen mit einzelnen Krankenkassen oder Krankenversicherungen und im Weiteren dann ggf. über eine Vergütungsregelung für die gesamte gesetzliche Krankenversicherung im Rahmen der kollektivvertraglichen Regelungen.

2.2.1 Krankenkassen

Einzelne Krankenkassen können individuell über die Vergütung einer neuen Technologie für ihre Versicherten entscheiden. Hier sind verschiedene Vergütungsformen möglich. Die wesentlichen sind in Teil 5 dieser Reihe, „Vertrags- und Vergütungsformen“, genauer beschrieben (vgl. Bertelsmann Stiftung 2017c). Grundlage einer solchen Entscheidung sind sowohl Ergebnisse aus Interventionsstudien zur Wirksamkeit als auch zur Technologiefolgeabschätzung.

Die produktbezogene Forschung zur Wirksamkeit erbringt in der Regel der Hersteller. In ausgewählten Fällen wie etwa bei der Entwicklung von Behandlungsprogrammen werden Krankenkassen auch bei der Begleitforschung bzw. Evaluation selbst aktiv. Während klassische produktbezogene Forschung gewöhnlich nicht von Krankenkassen finanziert wird, stellt Digital Health hier eine Mischform aus Produkt- und Prozessinnovation dar. In der Praxis ist daher zu beobachten, dass Krankenkassen die in Digital Health oft enthaltenen Prozessinnovationen durch eigene Evaluationen und Forschungsaktivitäten beispielsweise über die Rechtsgrundlage der Selektivverträge begleiten.

Den Nutzen bewertet hingegen meistens die einzelne Krankenkasse selbst und sie entscheidet auch über die Vergütung. In beiden Fällen werden jedoch üblicherweise Ergebnisse aus Interventionsstudien zur relevanten Indikation systematisch gesichtet und bei angemessener methodischer Qualität in die Entscheidung einbezogen. In ausgewählten Fällen beauftragen die Krankenkassen ihren Medizinischen Dienst oder andere Institute mit einer Technologiefolgeabschätzung. Hierbei gibt es keine gesetzlichen oder untergesetzlichen Vorgaben bezüglich der Unterlagen, die zum Nachweis des Nutzens bei der Krankenkasse vorzulegen sind. Allerdings erhöht der Hersteller seine Erfolgsaussichten, wenn er mit wissenschaftlich fundierten Ergebnissen und Unterlagen an eine Krankenkasse herantritt.

Folgende fünf Eckpunkte haben sich in der Praxis als wesentliche Bestandteile erwiesen. Die Punkte eins bis vier sollte der Anbieter jeweils vorbereiten. Den letzten Punkt übernimmt die Krankenkasse:

1. Problem: Beschreibung der Bedeutung und der Versorgungssituation des relevanten Indikationsbereiches mit Therapieoptionen, Defiziten und Lösungsoptionen (u. a. Differenzierung nach Prozess-, Struktur- und Ergebnisqualität).

2. Lösung: Definition der auf Digital Health basierten Intervention anhand einer verständlichen und nachvollziehbaren Beschreibung des Produktes und der Funktionsweise; zentrale Punkte sind hierbei, welche Personen- bzw. Fallgruppen mit welcher Kombination von Maßnahmen durch die Anwendung angesprochen werden.

3. Wirksamkeit: Der Hersteller muss nachweisen, welche Verbesserung der Versorgung (u. a. über Struktur-, Prozess- und/oder Ergebnisqualität) für die Zielgruppe der Patienten erreicht werden kann. Hierfür sind Ergebnisse einer eigenen Interventionsstudie gefordert bzw. hilfreich; diese sollten eine angemessene methodische Qualität haben und sinnvollerweise durch ein unabhängiges Institut bzw. eine Forschungseinrichtung begleitet werden.

4a. Nutzen Beispielpopulation: Die Ergebnisse zur Wirksamkeit werden anhand einer Beispielpopulation hochgerechnet und mit der Wirksamkeit anderer Therapieoptionen bzw. Behandlungsverfahren verglichen. Mit betrachtet wird hier auch der weitere Kontext wie Versorgungsstrukturen, Vergütungssysteme, gesetzliche Rahmenbedingungen etc. Diese Punkte sollte der Anbieter vorbereiten, ggf. auch in Begleitung eines externen Instituts.

4b. Nutzen kassenspezifische Population: Die Zahlen zu einer kassenspezifischen Population und den spezifischen Rahmenbedingungen (z. B. regionale Spezifika und Morbidität der Population) können meist nur von der Kasse selbst beigebracht werden. Somit wird eine spezifische gesundheitsökonomische Bewertung gewöhnlich von der jeweiligen Krankenkasse mit den jeweils eigenen und intern üblichen Methoden modellhaft gerechnet. Einbezogen werden dabei über die reine Intervention hinaus auch weitere übergreifende Effekte, z. B. in den Bereichen Marketing, Qualität (Prozess, Struktur, Ergebnis), Ökonomie und Innovation. Diese Punkte werden im Kontext der eigenen Strategie des Hauses bewertet.

2.2.2 Medizinischer Dienst des Spitzenverbandes (MDS)

Im Gegensatz zu Arzneimitteln werden nicht grundsätzlich alle Medizinprodukte nach der Marktzulassung bzw. CE-Kennzeichnung einer Nutzenbewertung unterzogen. Ob dies geschieht, hängt bei den Produkten wesentlich von der Vergütungsform ab. Bei bestimmten Vergütungsformen von Medizinprodukten beauftragt der Medizinische Dienst des Spitzenverbandes Bund der Krankenkassen (MDS) die Technologiefolgeabschätzung – etwa wenn Medizinprodukte, angewendet von Patienten, als Hilfsmittel über einen Eintrag ins Hilfsmittelverzeichnis vergütet bzw. erstattet werden sollen. Auch beauftragen Krankenkassen den MDS regelmäßig, wenn zwischen einem Krankenhaus und Krankenkassen ein NUB-Entgelt (für neue Untersuchungen und Behandlungen) verhandelt werden soll oder wenn eine Krankenkasse über einen integrierten Versorgungsvertrag verhandelt. Für eine detaillierte Übersicht der Vergütungsformen siehe Bertelsmann Stiftung (2017a).

Der MDS hat dafür ein definiertes methodisches Vorgehen, das insgesamt dem Prozedere eines Health Technology Assessment (HTA) entspricht (vgl. Hermann-Frank und Lelgemann 2013). Zunächst wird eine Recherche über das Produkt einschließlich CE-Zertifizierung, Produktinformation und Gebrauchsanweisung durchgeführt. Es folgt eine bibliographische Literaturrecherche in den gängigen Datenbanken, u. a. in der Cochrane Library und PubMed. In einem ersten analytischen Schritt werden die Abstracts der Studien überprüft, um anschließend die Volltexte zu analysieren. Bei der Analyse der Studien wird insbesondere das Verzerrungspotenzial bewertet, also inwieweit die Studienergebnisse durch zufällige oder systematische Verzerrungen (Bias) die Aussagen verfälschen (vgl. Statista). Dann folgt eine sozialmedizinische Empfehlung.

Das Verfahren ist deutlich formaler als die Bewertung durch eine einzelne Krankenkasse. Die Anforderungen an den Anbieter ähneln jedoch denen des vorab skizzierten Vorgehens: Das entsprechende Material zum Nachweis der Wirksamkeit und des Nutzens sollte er vorbereitet haben.

2.2.3 Gemeinsamer Bundesausschuss (G-BA)

Im Kontext der kollektivvertraglichen Versorgung der gesetzlichen Krankenversicherung (GKV) ist der G-BA zuständig für die Durchführung von Nutzenbewertungen und Anpassungen am Regelleistungskatalog der GKV. Das hier angewendete Verfahren der Methodenbewertung erfolgt in der Regel unter Einschluss (Beauftragung) des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), das 2004 mit der Maßgabe gegründet wurde, als wissenschaftlich unabhängiges Institut neue Technologien für Entscheidungen des G-BA zu bewerten.

Die Rahmenbedingungen zu Nutzenbewertungen sind in der Verfahrensordnung des G-BA (2017) sowie im Methodenpapier des IQWiG (2014) festgelegt. Bei der Prüfung bittet der G-BA die Fachöffentlichkeit um Stellungnahme und beauftragt in der Regel das IQWiG, eine Bewertung der Technologie durchzuführen. Nach dem Health Technology Assessment gibt das IQWiG eine Empfehlung an den G-BA, woraufhin dieser entscheidet. Die Methodenbewertung überprüft den Nutzen, die medizinische Notwendigkeit und die Wirtschaftlichkeit einer Leistung nach Kapitel 2, §§ 7, 10, 11 Verfo (G-BA 2017).

Für den Nachweis des Nutzens sind im Gesetz (§ 35b SGB V) drei relevante Parameter genannt: Mortalität (Sterblichkeit infolge einer Erkrankung), Morbidität (Beschwerden oder Komplikationen bei einer Erkrankung) und Lebensqualität. Diese Parameter sind sogenannte patientenrelevante Endpunkte einer Studie, also die Outcomes, die in der Studie zu einem definierten Zeitpunkt gemessen werden. Meistens werden sie in einer Interventionsgruppe durchgeführt und mit einer Kontrollgruppe verglichen. So wird untersucht, welche Effekte (bessere Lebensqualität, längeres Überleben, weniger Nebenwirkungen etc.) in der Interventionsgruppe aufgrund der Intervention festzustellen sind im Vergleich zu der Kontrollgruppe mit Standardtherapie. Finanzielle Aspekte spielen hier zunächst keine Rolle.

Die medizinische Notwendigkeit ist abhängig von der Relevanz der medizinischen Problematik, dem Verlauf der Erkrankung und vorhandenen Alternativen. Für die Wirtschaftlichkeit müssen zusätzlich Unterlagen zur Kostenschätzung und zur Kosten-Nutzen-Abwägung in Bezug auf einzelne Patienten oder auf die Gesamtheit der Versicherten – auch im Vergleich zu Alternativen – eingereicht werden.

Der G-BA differenziert dabei zwischen diagnostischen und therapeutischen Methoden. Die gerade dargestellten entsprechen, aufgrund der größeren Relevanz für Digital-Health-Anwendungen, den Evidenzstufen therapeutischer Methoden. Bei diagnostischen Methoden wird zusätzlich ein Fokus auf die Analyse diagnostischer Kenngrößen zur Testgenauigkeit gelegt. Diese Testgüte wird vor allem bemessen durch die Sensitivität (Rate der richtig als krank Identifizierten) und Spezifität (Rate der richtig als gesund Identifizierten), durch Wahrscheinlichkeitsverhältnisse sowie einen positiven und negativen prädiktiven Wert (quasi die Vorhersagegenauigkeit). Für die exakte Klassifizierung der Evidenzstufen siehe die Verfahrensordnung des G-BA (2017), § 11.

Für die Nutzenbewertung beim G-BA sind also unterschiedliche Studienformen als Nutzenbeleg möglich. Allerdings muss, je niedriger die Evidenzstufe ist, die Auswahl der Studienform begründet werden. Nachweise des Nutzens durch Unterlagen mit niedriger Evidenz sind tendenziell eher möglich, wenn es sich um besonders seltene Erkrankungen handelt oder wenn keine Behandlungsalternativen bestehen.

Die Nutzenbewertungen des G-BA setzen den methodischen Standard für Technologiefolgeabschätzungen in Deutschland. Das IQWiG ermittelt dabei den belegbaren Nutzen hinsichtlich Morbidität, Mortalität und Lebensqualität aufgrund der Studiendaten. Der G-BA entscheidet anschließend über Nutzen, Notwendigkeit und Wirtschaftlichkeit und kann dabei von den Empfehlungen des IQWiG abweichen. Diese Anforderungen gelten somit auch für Digital-Health-Anwendungen, die als Medizinprodukte zugelassen und in den Leistungskatalog der gesetzlichen Krankenversicherung aufgenommen werden sollen.

Nicht zu unterschätzen ist die Bedeutung von medizinischen Fachgesellschaften und Einrichtungen wie der Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften in Deutschland (AWMF e. V.). In seiner Verfahrensordnung bezieht sich der G-BA mehrfach auf medizinische Leitlinien als relevante Unterlagen für die Darstellung des medizinischen Hintergrundes und des Standes der medizinischen Erkenntnis. Die Aufnahme einer Methode oder eines Verfahrens in eine nationale oder internationale medizinische Leitlinie kann somit als hilfreiche bzw. notwendige, wenn auch nicht hinreichende Voraussetzung für eine positive Vergütungsentscheidung des G-BA gesehen werden. Die Beteiligungsrechte der Fachgesellschaften an Verfahren des G-BA sind in dessen Verfahrensordnung geregelt (G-BA 2017, Anlage III zu Kapitel 1).

2.3 Evidenzklassen für Studien zur Wirksamkeit

Im Zuge der Weiterentwicklung und Stärkung der evidenzbasierten Medizin (EbM) hat sich über die Jahre eine Systematik herausgebildet, um die Güte von Evidenz und der zugrundeliegenden Studien einzuordnen. Die Evidenzklassen definieren, welche Anforderungen an die Qualität der Evidenz bei Studien gestellt werden. Diese werden sowohl bei den zuvor beschriebenen Institutionen und deren Methodik bei Interventionsstudien und Technologiefolgeabschätzungen, bei den Verfahren zur Medizinproduktezertifizierung und Vergütungsregelung als auch bei der Definition von medizinischen Leitlinien berücksichtigt.

Das Netzwerk evidenzbasierte Medizin definiert die Evidenzklassen wie folgt: Die höchste Evidenzklasse Ia bildet solche Evidenz, die auf systematischen Übersichtsarbeiten randomisierter, kontrollierter Studien beruht; Evidenzstufe Ib ist erreicht, wenn die Evidenz anhand von mindestens einer hoch qualitativen randomisierten, kontrollierten Studie erhoben wurde; Evidenzstufe IIa liegt vor, wenn mindestens eine gut angelegte kontrol-

Tabelle 1: **Evidenzklassen**

Evidenzklasse	Anforderung
Ia	Evidenz aufgrund einer systematischen Übersichtsarbeit von randomisierten, kontrollierten Studien (evtl. mit Metaanalyse)
Ib	Evidenz aufgrund mindestens einer hoch qualitativen, randomisierten, kontrollierten Studie
IIa	Evidenz aufgrund mindestens einer gut angelegten, kontrollierten Studie ohne Randomisierung
IIb	Evidenz aufgrund mindestens einer gut angelegten, quasi experimentellen Studie
III	Evidenz aufgrund gut angelegter, nicht experimenteller deskriptiver Studien
IV	Evidenz aufgrund von Berichten, Meinungen, Expertenkreisen, Konsensuskonferenzen und/oder klinischer Erfahrung anerkannter Autoritäten

Quelle: Deutsches Netzwerk für evidenzbasierte Medizin

lierte Studie ohne Randomisierung vorliegt; Evidenzstufe IIb besagt, dass die Evidenz aufgrund mindestens einer gut angelegten, quasi experimentellen Studie beigebracht wird, und Evidenzstufe III ist mit gut angelegten, nicht experimentellen, deskriptiven Studien erreicht (Deutsches Netzwerk für evidenzbasierte Medizin 2017).

Der G-BA (2017) folgt in seiner Verfahrensordnung einer vergleichbaren Systematik, ebenso wie das IQWiG (2014) in seinem Methodenpapier.

Digital-Health-Anwendungen müssen somit – sofern sie das Potenzial bzw. den Anspruch haben, das klassische Versorgungsgeschehen substanziell zu revolutionieren (z. B. das Ersetzen von Therapien, Medikamenten, Verfahren oder Leistungen) – ihre Wirksamkeit und ihren Nutzen nach den vorherrschenden Methoden belegen. Der methodische Standard ist im höchsten und international üblichen Evidenzlevel gesetzt durch randomisierte kontrollierte (Interventions-)Studien oder systematische Übersichtsarbeiten zu ebensolchen. Randomisiert bedeutet hier, dass es eine Interventionsgruppe und eine Kontrollgruppe (z. B. Placebo) gibt, wobei die Studienteilnehmer per Zufallsverfahren der einen oder der anderen Gruppe zugeordnet werden. Systematische Übersichtsarbeiten zu randomisierten kontrollierten (Interventions-)Studien sind sogenannte Metastudien, die die Ergebnisse aller gefundenen randomisierten Studien zum gleichen Thema miteinander vergleichen.

2.4 Bedarf an Nachweis der Wirksamkeit und Nutzenbewertung bei Digital-Health-Anwendungen

Wirksamkeitsnachweise und Nutzenbewertungen sind im Bereich Digital Health noch eher unüblich. Allerdings ist die aktuelle Diskussion teilweise geprägt durch Unklarheit darüber, für welche Anwendungstypen solche Nachweise und Bewertungen erforderlich sind. Hierfür kann die Einordnung anhand von sieben Anwendungstypen (siehe Anhang) und der Medizinproduktklassifizierung herangezogen werden. Diese wurde in einem Teilbericht 3 dieser Reihe im Detail vorgestellt.

Bei digitalen Gesundheitsanwendungen kann die Abgrenzung zwischen Leistungen, die für die Versorgung relevant sind, und Lifestyleprodukten grundsätzlich auf Grundlage der Ziel-

Tabelle 2: Versorgungsrelevanz der digitalen Anwendungstypen nach Zielgruppe

Anwendungstypen	Zielgruppe / Gesundheitszustand				
	Gesund ohne Risikofaktoren	Gesund mit Risikofaktoren, Verdacht auf Erkrankung, unklaren Symptomen, einschließlich Empfängnis		Krank akut und chronisch	
		Allgemeine Information	Personalisierte Information	Allgemeine Information	Personalisierte Information
I Typ 1: Stärkung der Gesundheitskompetenz					
II Typ 2: Analyse und Erkenntnis					
III Typ 3: Indirekte Intervention					
IV Typ 4: Direkte Intervention					
V Typ 5: Dokumentation Gesundheits- und Krankheitsgeschichte					
VI Typ 6: Organisation und Verwaltung					
VII Typ 7: Einkauf und Versorgung					

■ versorgungsrelevant ■ nicht versorgungsrelevant
 Quelle: Eigene Darstellung

gruppe und der sieben Anwendungstypen erfolgen. Tabelle 2 zeigt die Versorgungsrelevanz von sieben digitalen Anwendungstypen nach Zielgruppe:

Digital-Health-Anwendungen der Typen 2 (Analyse und Erkenntnis), 3 (indirekte Intervention) und 4 (direkte Intervention für kranke Menschen und Menschen mit Risikofaktoren) sowie Anwendungen vom Typ 1 (Stärkung der Gesundheitskompetenz mit personalisierten Informationen für Menschen mit spezifischem Erkrankungsrisiko) sind in der Regel als Medizinprodukte einzustufen – und bedürfen demnach einer europäischen Zulassung mit CE-Kennzeichen (siehe Teil 3 der Reihe, „Medizinproduktezertifizierung“, Bertelsmann Stiftung 2017b). Die gleichen Anwendungen haben auch Bedarf an Wirksamkeitsnachweis und Nutzenbewertung, da sie für eine Vergütung durch Krankenkassen oder Krankenversicherungen infrage kommen. Die Vergütungsmöglichkeiten für versorgungsrelevante Leistungen unterliegen verschiedenen Anforderungen und teilweise komplexen Rahmenbedingungen – sie wurden in Teil 5 dieser Reihe, „Vertrags- und Vergütungsmodelle in der GKV“, bereits betrachtet.

3 Wirksamkeitsnachweis

Im Folgenden werden für den Wirksamkeitsnachweis klassische und neue Methoden einander gegenübergestellt. Das Kapitel bietet somit einen Überblick über die methodischen Möglichkeiten, die hierzulande und international noch wenig etabliert sind.

3.1 Klassische Studiendesigns

Ähnlich wie bei den Arzneimitteln hat für den Bereich der Medizinprodukte die Frage nach Evidenz für den Nutzen an zentraler Bedeutung gewonnen. Innovationen in der Medizintechnik und auch im Digital-Health-Bereich müssen zunehmend nachweisen, dass sie für die Patienten und das Gesundheitswesen Wirkung und Nutzen entfalten, wenn sie u. a. im Rahmen der gesetzlichen Krankenversicherung eingesetzt und vergütet werden sollen. Diese Anforderungen werden künftig eher noch zunehmen. Umso wichtiger ist es, frühzeitig zu analysieren und zu planen, welcher Grad der Evidenz für ein Produkt oder ein neues Verfahren notwendig und möglich sein wird. Davon leitet sich ab, mit welcher Methodik diese Evidenz gewonnen werden muss.

Evidenz im Gesundheitswesen kann in unterschiedlicher Form und mit unterschiedlichem Aufwand gewonnen werden. Röhrig et al. erläutern im Deutschen Ärzteblatt anschaulich, dass dabei zwischen der Forschung an Primärdaten und an Sekundärdaten unterschieden werden kann (Röhrig, Prel und Blettner 2009). Die Festlegung des Studientyps ergibt Klarheit über die Forschungskategorie. Entscheidend für die Studienplanung ist die Festlegung des Studiendesigns. Dies hängt wiederum eng mit der eigentlichen Forschungsfrage zusammen: Was ist das Ziel einer Studie und welche Frage soll beantwortet werden? Diese muss dann mess- und beurteilbar gemacht werden. Das Studiendesign soll hier als Gesamtkonzept aller Vorgehensweisen im Rahmen einer Studie verstanden werden (ebd.). Sechs Aspekte gehören zu diesem Gesamtkonzept: Fragestellung, Studienpopulation, Studientyp, Beobachtungseinheit, Messverfahren und Fallzahlplanung (ebd.).

Unter klinischer Forschung werden sowohl interventionelle als auch nicht interventionelle Studien zusammengefasst. Mit interventionellen klinischen Studien (klinische Prüfungen) sollen klinische Wirkungen von Arzneimitteln oder Medizinprodukten nachgewiesen werden (vgl. Röhrig, Prel und Blettner 2009). Interventionelle Studien mit einer Randomisierung überwiegen in der Praxis und haben den höheren Evidenzgrad (s. u.).

Interventionelle Studien sind Therapiestudien, in denen eine neue Behandlungsmethode erprobt wird. Eine kontrollierte Studie gilt als Goldstandard. Die neue Behandlung wird mit einer anderen bestehenden Behandlung oder einer Scheinbehandlung (Placebo) verglichen; dazu werden Studienteilnehmer in zwei Gruppen eingeteilt. Es gibt randomisierte und nicht randomisierte Interventionsstudien sowie solche mit und ohne Verblindung.

Randomisierung: Studienteilnehmer werden per Zufallslos einer der Gruppen zugeordnet. In der Behandlungsgruppe kommt die neue Behandlung zum Einsatz; die andere Gruppe ohne Einsatz der neuen Behandlung ist die Kontrollgruppe.

Verblindung einer Studie bedeutet, dass die Studienteilnehmer nicht wissen, ob sie in der Studiengruppe oder in der Kontrollgruppe sind. Bei einer sogenannten Doppelblindstudie wissen die durchführenden Ärzte dies ebenfalls nicht (vgl. Röhrig, Prel und Blettner 2009).

Nicht interventionelle Studien beinhalten keine gezielte Intervention – es wird lediglich beobachtet und dokumentiert. Drei wichtige Studientypen werden unterschieden: Fall-Kontroll-Studie, Kohortenstudie und Anwendungsbeobachtung.

Fall-Kontroll-Studie: Form einer retrospektiven Studie, bei der Patienten, die eine Behandlung erhalten haben, im Nachhinein mit einer zweiten Gruppe ohne die Behandlung verglichen werden. Es wird ein bestimmtes Zielereignis definiert. Eine Gruppe enthält Probanden, bei denen das Zielereignis eingetreten ist; bei der anderen Gruppe ist es ausgeblieben.

Kohortenstudie: Eine Gruppe von Patienten („Kohorte“), die eine Behandlung erhalten haben, werden hinsichtlich ihres Krankheitsverlaufs prospektiv über einen längeren Zeitraum beobachtet. Die Kohorte wird nach bestimmten Charakteristika (z. B. Risikofaktor) ausgewählt. Ziel ist es, ein bestimmtes Ereignis, das zu Beginn der Untersuchung noch nicht vorhanden war, nach Ausmaß und Häufigkeit des Auftretens zu untersuchen und zu beschreiben.

Anwendungsbeobachtung: Ein neues bereits zugelassenes Medikament oder Medizinprodukt wird bei seiner Anwendung über einen längeren Zeitraum beobachtet, um Informationen über den Einsatz im Alltag zu sammeln. Aspekte wie die Therapietreue des Patienten oder Begleiterkrankungen sind hier eine Größe, die das Ergebnis beeinflussen und beobachtet werden können (vgl. Röhrig, Prel und Blettner 2009).

Die Anforderungen an Interventionsstudien sind im Bereich der Medizinprodukte abhängig vom Produkttyp und von der Risikoklasse. Im Vergleich zu denen an andere Produkte wie etwa Arzneimittel und den Anforderungen im Rahmen der Vergütung in der gesetzlichen Krankenversicherung sind sie niedriger. Soweit eine Interventionsstudie im Kontext der Medizinproduktezertifizierung gefordert ist, können die Kernelemente einer typischen Interventionsstudie wie in Abbildung 1 zusammengefasst werden.

Solche Studien werden klassischerweise sequenziell nach der Produktentwicklung und vor dem Konformitätsverfahren durchgeführt. Klassische Medizinprodukte werden mehrere Jahre unverändert am Markt vertrieben.

3.2 Limitationen klassischer Methoden

RCT (Randomized Controlled Trial) gilt als der Goldstandard bei der Gewinnung von klinischer Evidenz. Bei den Arzneimitteln haben sich RCTs zur Bewertung etabliert. Aufgrund der Heterogenität von Medizinprodukten ist die Durchführung von RCTs schon in diesem Bereich schwieriger als bei den Arzneimitteln. Die Entwicklungszyklen sind andere, der Patentschutz ist nicht vergleichbar und eine Randomisierung oder gar Verblindung gestaltet sich komplizierter bzw. ist teilweise nicht möglich. Dennoch ist auch bei den Medizinprodukten erkennbar, dass sich RCTs als maßgebliches Studiendesign für die Evidenzge-

Abbildung 1: Kernelemente einer Interventionsstudie



winnung durchsetzen: Das IQWiG hat im Januar 2018 eine „Analyse klinischer Prüfungen bei Medizinprodukten“ vorgelegt, die auf Anträgen bei der Ethikkommission des Landes Berlin in den Jahren 2010 bis 2013 beruht (vgl. IQWiG 2018). In dem betrachteten Zeitraum wurden 122 Anträge auf klinische Prüfung eines Medizinproduktes bzw. eines In-Vitro-Diagnostikums bei der Ethikkommission gestellt. 57 Prozent davon (70) waren als RCT geplant. 62 Studien hatten das Ziel, Wirksamkeit nachzuweisen.

Bei Digital-Health-Anwendungen sind die Voraussetzungen noch schwieriger. Eine besondere Herausforderung ist es, bestehende Forschungsdesigns mit der hohen Innovationsgeschwindigkeit dieser Anwendungen zusammenzubringen (vgl. Pham et al. 2016). RCTs laufen den Anforderungen von Digital-Health-Anwendungen an Evaluationsmethoden zuwider und entsprechen den Besonderheiten noch weniger als bei Medizinprodukten.

Folgende Punkte sprechen dagegen, dass klassische RCTs im Kontext von Digital Health unmittelbar angewendet werden sollten (vgl. Pham, Wiljer und Cafazzo 2016; Albrecht et al. 2018; Graalman, Czaja und Rödiger 2017):

- **Release-Zyklen:** Digital-Health-Anwendungen basieren auf Software. Ein Charakteristikum von Software ist, dass sie sich schnell und kontinuierlich verändert und weiterentwickelt. Digital-Health-Anwendungen haben daher kurze, agile Entwicklungszyklen, einhergehend mit sich schnell verändernden Nutzerwünschen und -bedarfen.

- **Produktlebenszyklus:** RCTs dauern lange (im Schnitt fünf Jahre von der Einschreibung der Teilnehmer bis zur Veröffentlichung der Ergebnisse). Sie passen mit ihren strikten Studienprotokollen nicht zur Dynamik von Digital-Health-Anwendungen, sondern stehen ihr vielmehr entgegen – die Anwendung könnte in der evaluierten Form obsolet sein, bevor die Evidenz vorliegt.
- **Dynamisierung und Personalisierung:** Information und Intervention bei den Digital-Health-Anwendungen sind stark personalisiert und dynamisiert. Bei den RCTs liegt der primäre Fokus auf einer Standardisierung unter Laborbedingungen; dadurch wird versucht, alle störenden externen Einflussgrößen möglichst kleinzuhalten – hierin liegt ein Zielkonflikt.
- **Verhaltensänderung und soziale Interaktion:** Digital-Health-Anwendungen haben einen starken Fokus auf Verhaltensänderungen und auf das Gesundheitshandeln im Kontext des sozialen Umfelds bzw. der sozialen Interaktion. Diese ist nicht vorab im Detail festgelegt bzw. definiert. Zudem bestehen die Anwendungen häufig aus unterschiedlichsten Modulen. Durch diese teils indirekten Effekte ist der Zusammenhang zwischen Ursache und Effekt schlechter isoliert darstellbar. Diese Eigenschaften sind mit strikten Studienprotokollen bei RCTs wenig vereinbar.

Bei Arzneimitteln hingegen und mit Abstrichen auch bei Medizinprodukten sind Ursache und Effekt bisher leichter isolierbar gewesen. Außerdem haben diese längere Lebenszyklen als eine Digital-Health-Anwendung. Lang laufende RCTs sind daher hier eher durchführbar und sinnvoll.

Vor diesem Hintergrund überrascht es, dass Pham et al. in ihrer Analyse zeigen, dass auch bei der Evaluation von Digital-Health-Anwendungen überwiegend (80 %) RCTs genutzt wurden. Dies zeigt, dass Forscher ihr Studiendesign bisher nicht anpassen an die besonderen Qualitäten von Digital-Health-Anwendungen, die sie untersuchen. Studien werden weiterhin überwiegend an Studienzentren und in Krankenhäusern durchgeführt, das heißt, neue Methoden zur Datenerhebung werden bisher nur eingeschränkt genutzt (vgl. Pham, Wiljer und Cafazzo 2016).

Im Folgenden werden vier neue Methoden vorgestellt, die den oben genannten Limitationen entgegenwirken. Sie werden den Besonderheiten von Digital-Health-Anwendungen gerechter und können als agile Weiterentwicklung klassischer RCTs gesehen werden.

3.3 Neue Methoden

Für Digital-Health-Anwendungen haben sich in der internationalen Forschungsgemeinde neue Methoden der Evidenzgewinnung entwickelt. Diese sind stärker auf die Besonderheiten der neuen Technologien zugeschnitten als klassische RCT-Studiendesigns und können ohne substantielle Abstriche bei der Evidenzgüte eingesetzt werden. Obwohl sie somit das Potenzial haben, den Einstieg für Digital-Health-Anwendungen in den ersten Gesundheitsmarkt zu erleichtern, werden sie in Deutschland bislang wenig angewendet. Eine Auswahl dieser Methoden – CEEBIT, MOST, SMART und Micro-Randomized Trials – wird im Folgenden den klassischen, RCT-dominierten Studiendesigns exemplarisch gegenübergestellt.

Continuous Evaluation of Evolving Behavioral Intervention Technologies (CEEBIT)

CEEBIT (vgl. Mohr et al. 2013) bzw. die „Kontinuierliche Evaluierung sich entwickelnder Verhaltensinterventionstechnologien“ ist eine alternative Bewertungsmethode für technick-

basierte Verhaltensinterventionen (Behavioral Intervention Technologies – BIT). Behavioral Intervention Technologies (BIT) sind die Umsetzung von Verhaltens- und psychologischen Interventionsstrategien durch technische Applikationen. Durch die Nutzung von BITs sollen Verhaltens-, kognitive und affektive Ziele erreicht werden, welche die körperliche und mentale Gesundheit sowie Verhaltensänderungen unterstützen (ebd.). Damit sind BITs eine spezielle Digital-Health-Anwendung. CEEBIT stellt für diese Anwendungen eine Alternative zu RCTs dar, ohne RCT-spezifische Überlegungen zu vernachlässigen: Randomisierung, statistische Methoden zur Verbesserung nicht randomisierter Selektionen und hoch entwickelte Einschluss- bzw. Ausschlusskriterien werden kombiniert.

Ein RCT dauert oft Jahre und ist damit ein für die Entwicklungsdynamik der sich ständig weiterentwickelnden BITs ungeeignetes Bewertungsinstrument. BITs können jedoch strategisch und geplant kontinuierlich Daten sammeln und damit durch prospektive systematische Analyse Evidenz erzeugen. Dadurch entsteht ein „lernendes System“, das die dynamischen Veränderungen aufgreift, vorherige Versionen mit schlechten Outcomes aussortiert und neuen ermöglicht, in die Evaluation einbezogen zu werden (ebd.).

Die Methode nutzt also aus, dass die BITs routinemäßig Daten in Echtzeit sammeln und durch deren kontinuierliche prospektive Analyse Evidenz bereits während der Entwicklungsphase der Interventionen geschaffen werden kann. So beseitigt CEEBIT die Notwendigkeit mehrerer einzelner RCTs.

Die Methode schließt Zuordnungsverfahren ein, die es erlauben, nicht nur die Akzeptanz und Motivation der Studienteilnehmer hochzuhalten, sondern auch die Nutzerpräferenzen als vollständig beobachtbare Prä-Randomisierungsfaktoren zu berücksichtigen und den Einfluss der Konsumentenwahl zu untersuchen.

Klassische RCTs prüfen einzelne Anwendungen gegenüber einer Standardtherapie. In einer CEEBIT-Studie werden verschiedene Anwendungen oder Komponenten im Verhältnis zueinander getestet und unterlegene bzw. minderwertige aus der Studie eliminiert. Neben der Erhöhung der Studieneffizienz und der Nutzung des Real-Life-Settings besteht daher eine weitere Stärke der Methode darin, weniger Probanden zu benötigen. Darüber hinaus ist die Methode in der Lage, Aktualisierungen der Technologie ebenso wie die Veränderungen der Population und der Umgebung zu managen (ebd.).

Multiphase Optimization Strategy (MOST)

MOST (vgl. Collins, Murphy und Strecher 2007) oder Mehrphasenoptimierungsstrategie ist ein neuer methodischer Ansatz zur Bewertung von Verhaltensinterventionen (Behavioral Intervention – BIT). Er besteht aus den drei Phasen Vorbereitung, Optimierung und Bewertung. Ziel ist die Identifizierung der optimierten Kombination von Interventionskomponenten vor der Durchführung eines ressourcenintensiven RCTs. In der Vorbereitungsphase werden Komponenten auf Basis ihrer Performance identifiziert, um sie bei einer Intervention zu nutzen. Bei der Optimierungsphase werden die ausgewählten Komponenten weiterentwickelt und Elemente wie die optimale „Dosierung“ festgelegt. In der Bewertungsphase werden die optimierten Interventionen, basierend auf optimal dosierten Einzelkomponenten, in ihrer Kombination in einem Standard-RCT untersucht (ebd.).

Während ein klassisches RCT die gesamte Intervention bewertet, lassen sich mit der MOST-Methode die Hauptwirkung jeder einzelnen Interventionskomponente sowie die Wechselwirkungen zwischen den Komponenten untersuchen. Diese Informationen können

verwendet werden, um die einzelnen Teilinterventionen zu optimieren (im Sinne einer Eliminierung minderwertiger Komponenten), bevor die gesamte Intervention in einem RCT untersucht wird. So sind auch Aussagen möglich über optimale Kombinationen einzelner Komponenten bzw. über die Wirksamkeit einzelner Komponenten und deren Anteil an der Gesamtwirkung der Anwendung. Daher besteht mit MOST die Möglichkeit, auch indirekte Outcomes eindeutiger zu ermitteln.

Zum Einsatz kommen hier fraktionale, faktorielle, experimentelle Designs (Teilfaktorplan) mit sorgfältig ausgewählten Abschnitten der einzelnen Untersuchungen. Bei einem solchen Teilfaktorplan werden Teile eines vollständigen Versuchsplans durchgespielt, um die wirksamen Einzelkomponenten zu identifizieren. Zudem werden randomisierte Experimente eingesetzt, um Zusammenhänge mit Ko-Varianten zu untersuchen, und Modellierungsanalysen zur Bestimmung wirksamer Komponenten-Kombinationen verwendet. Nach solchen Optimierungsschritten wird die Bewertung mittels einer klassischen RCT vorgenommen, die nun jedoch deutlich ressourcenschonender erfolgen kann.

Sequential Multiple Assignment Randomized Trial (SMART)

SMART (vgl. Collins, Murphy und Strecher 2007) ist eine sequenzielle randomisierte Mehrfacheinschluss-Studie, in der eine Randomisierung in jeder Stufe, also sequenziell, erfolgt. Eine Randomisierung in verschiedene Behandlungsoptionen zu einem späteren Zeitpunkt kann dabei als Konsequenz aus der Wirksamkeit oder Unwirksamkeit einer vorherigen Intervention vorgenommen werden. Dies ermöglicht die einzelne Beurteilung verschiedener Variablen.

Die SMART-Methode erlaubt, im Rahmen eines experimentellen Designs Entscheidungsregeln für adaptive Interventionen zu operationalisieren. Der regelhafte und fundierte Einsatz von adaptiven Eingriffen in das Studiendesign gewinnt an Bedeutung angesichts der kurzen Entwicklungs- und Lebenszyklen von Digital-Health-Anwendungen sowie der einwirkenden instabilen sozialen und persönlichen Komponenten bei der Untersuchung der Anwendungen.

SMARTs liefern Daten, die die Entwicklung von qualitativ hochwertigen adaptiven Interventionen ermöglichen: Herangezogen werden dafür Merkmale und Reaktionen des Nutzers sowie Zwischenergebnisse, die während der Studie gesammelt wurden. So kann die Intervention dem individuellen Fortschritt der Nutzer angepasst und möglichst effektiv gestaltet werden.

Eine SMART-Studie besteht aus mehreren Interventionsstufen, wobei jede Stufe einer kritischen Entscheidung bei der adaptiven Intervention entspricht. Da die Teilnehmer randomisiert werden, können die Wissenschaftler die Wirksamkeit jeder einzelnen Stufe beurteilen. Die besonderen Vorteile von SMART: Mit diesem Design kann prospektiv über den Typ und die Intensität einer Intervention entschieden werden. In einstufigen randomisierten Studien können nur retrospektiv Informationen zu den Einflüssen von Interventionen ausgewertet und für die Entscheidung herangezogen werden. Die Analyse, wann die Wirkung einer Intervention durch nachfolgende oder vorherige Eingriffe verstärkt wird, hat mit dem SMART-Design eine höhere Gültigkeit. Zudem lassen sich Kohorteneffekte minimieren.

Micro-Randomized Trial

Ähnlich wie bei den zuvor beschriebenen neuen Methoden setzt das mikro-randomisierte Testverfahren auf eine sequenzielle Randomisierung während einer Studie. Micro-Randomized Trials (vgl. Klasnja et al. 2015) werden eingesetzt, um Prädiktoren der zeitveränder-

lichen Verfügbarkeit zu modellieren, was etwa bei der Evaluierung dynamischer Therapiepläne wichtig ist.

Diese Methode ermöglicht es, Teilnehmer zu den jeweiligen Entscheidungszeitpunkten in einer Studie, abhängig von ihrer Verfügbarkeit, den einzelnen Behandlungsoptionen randomisiert zuzuweisen. Mit einem Verfügbarkeitsindikator können Untergruppen identifiziert werden, die aus Studiensicht für eine Evaluation der Hauptbehandlungseffekte relevant sind. Teilnehmer, die sich aus welchen Gründen auch immer entziehen, werden ignoriert.

Bestandteil der Methode ist auch eine Anweisung zur Berechnung der Stichprobengröße, die den Verfügbarkeitsindikator nutzt. Neben der Verbesserung der Studieneffizienz ist es so möglich, Wechselwirkungen zwischen Behandlungen zu bewerten.

3.4 Gegenüberstellung der Methoden

Ein wesentlicher Unterschied zwischen den Methoden besteht darin, dass klassische Ansätze auf Primärdaten unter Laborbedingungen basieren und neue Ansätze auf eben-
solchen aus dem Real-Life-Setting. Alle neuen Methoden erheben selbst Primärdaten und beziehen Patientenberichte mit ein.

Darüber hinaus werden die neuen Verfahren nicht bei der Betrachtung von Sicherheitsaspekten herangezogen, sondern primär für klinische Evidenz und Wirksamkeit. Ihr Vorteil insbesondere bei der Evaluation von Digital-Health-Anwendungen ist, dass sie deren Entwicklungsdynamik durch iterative Verfahren gerechter werden als klassische RCTs.

Es ist davon auszugehen, dass der Grad der Evidenz bei den neuen Verfahren nicht geringer ist als bei den etablierten Verfahren.

Die vier neuen Methoden werden nachfolgend hinsichtlich der Dimensionen Studiendesign, Studienpopulation, Beobachtungseinheit, Beobachtungszeitraum und -zeitpunkt, Studienumgebung, Daten und Datenerhebung, Outcomes sowie ihres Umgangs mit Hürden und Störfaktoren mit klassischen RCTs verglichen.

Tabelle 3: **Klassische und neue Methoden – Gegenüberstellung nach Forschungsdimensionen und Datentypen**

	klassischer RCT	CEEBIT, MOST, SMART, Micro-Randomized Trial
Klinische Evidenz	Primärdaten (Labor-Setting)	Primärdaten (Real-Life-Setting)
Wirksamkeit	Primärdaten (Labor-Setting)	Primärdaten (Real-Life-Setting)
Sicherheit	Primärdaten (Labor-Setting)	
Wirtschaftlichkeit		
Soziale Aspekte		
Ethische Aspekte		
Rechtliche Aspekte		
Organisatorische Aspekte		

Quelle: Eigene Darstellung

Wesentliche Unterschiede klassischer und neuer Methoden

Studiendesign

Bei den neuen Methoden handelt es sich im Grunddesign genau wie bei klassischen RCTs um Interventionsstudien. Die Designs enthalten alle ebenfalls Randomisierung und Kontrollgruppen. Hinsichtlich der besonderen Anforderungen, die Digital-Health-Anwendungen als Forschungsgegenstand stellen, liegen die Stärken der neuen Methoden darin, dass sie u. a. innovative Designelemente wie ein iteratives Studiendesign und/oder präferenzbasierte Randomisierung enthalten können.

Studienpopulation

Die Studienpopulation beschreibt die Zahl der Teilnehmer einer Studie und welche Einschlusskriterien dieser Auswahl zugrunde liegen. Die neuen Methoden unterscheiden sich auch hinsichtlich der erforderlichen Probandenzahl von einem klassischen RCT. Die teils kleinen Stichprobennzahlen in einzelnen Iterationen können ggf. zunächst zu Akzeptanzproblemen führen, weil bisher gilt, dass eine kleine Stichprobengröße eine zu geringe statistische Power beinhaltet und die Gefahr birgt, Unterschiede nicht nachweisen zu können (vgl. Röhrig, Prel und Blettner 2009).

Grundsätzlich kann jedoch gerade bei Digital-Health-Anwendungen bedarfsweise die Studienpopulation rasch erweitert werden. Bei der Verifizierung der Einschluss- und Ausschlusskriterien bleibt jedoch eine Unsicherheit, wenn die Studien ohne Einbeziehung von Studienzentren durchgeführt werden und die Kriterien nur mit persönlichem Kontakt nachweisbar sind.

Beobachtungseinheit

Die Beobachtungseinheit beschreibt einen weiteren Teil des Umfangs einer Studie. Sie zeigt den Ausschnitt der Anwendung, der evaluiert werden soll, das heißt, es muss geklärt werden, ob eine ganze Anwendung oder nur einzelne Komponenten untersucht werden. Anders als bei klassischen RCTs können bei den neuen Methoden auch einzelne Komponenten und das Zusammenspiel von Komponenten untersucht werden.

Beobachtungszeitraum und -zeitpunkt

Der Beobachtungszeitraum und -zeitpunkt beschreibt, wie lange und zu welchem Zeitpunkt eine Studie durchgeführt wird. Bei den neuen Methoden sind die Beobachtungszeiträume im Vergleich zu einem klassischen RCT kürzer. Dies wird jedoch dadurch relativiert, dass mehrere kurze Beobachtungsphasen kumuliert werden können. Der gesamte Beobachtungszeitraum kann am Ende durchaus mit dem eines RCT vergleichbar sein; allerdings wird die Beobachtungseinheit ggf. jeweils angepasst. Zudem kann die Studie parallel zur Produktentwicklung durchgeführt werden. Die Zwischenergebnisse aus der Erhebung können für die weitere Optimierung der Entwicklung genutzt werden.

Studienumgebung

Studien können wie oben beschrieben unter Laborbedingungen durchgeführt werden oder in einem Real-Life-Setting – darin liegt ein wesentlicher Unterschied zwischen den neuen Methoden und einem klassischen RCT und der besondere Mehrwert der neuen Methoden für die Digital-Health-Anwendungen. Eine Herausforderung liegt allerdings in der Herstellung einer Vergleichbarkeit der Ergebnisse bei den neuen Methoden, insbesondere vor dem Hintergrund möglicher Voreingenommen-

heit (Bias) und von Störfaktoren (Confounder). Da jedoch Digital-Health-Anwendungen zunehmend das Gesundheitshandeln der Patienten im Alltag beeinflussen und hier auch das soziale Umfeld teils mit aktiviert wird, ist es anzuraten, die Intervention im Real-Life-Setting zu evaluieren und diese Faktoren im Studiendesign und in der Auswertung mit zu berücksichtigen.

Daten und Datenerhebung

Diese Dimension beschreibt, welche Art von Daten in welcher Form erhoben werden. Alle Methoden erheben selbst Primärdaten. Bei der MOST-Methode werden zusätzlich aggregierte Daten einbezogen. Alle Methoden beziehen auch Patientenselbstberichte mit ein. Klinische Daten werden jedoch lediglich bei RCTs und bei CEE-BIT berücksichtigt. Anders als bei den anderen Methoden werden bei CEEBIT und beim Micro-Randomized Trial die Daten kontinuierlich und nicht nur punktuell erhoben. Dies ermöglicht eine bessere Verlaufskontrolle.

Outcomes

Bei den Outcomes handelt es sich um die Ergebnisse einer Studie. Anders als beim Output geht es bei den Outcomes nicht um Quantität, sondern um die Qualität der Ergebnisse. Die Outcomes ergeben sich aus dem Zusammenspiel und der Auswertung der oben beschriebenen anderen Kriterien.

Der Hauptunterschied zwischen klassischen und neuen Studienansätzen ist, dass bei den neuen auch indirekten Outcomes wie Patientenautonomie, Komfort, Adhärenz, Nutzen, Selbstwirksamkeit, Werturteile und/oder Präferenzen mit aufgenommen werden.

Umgang mit Hürden und Störfaktoren

Hürden und Störfaktoren können die Durchführung von Studien erschweren und die Ergebnisse verfälschen. Sie müssen beim klassischen Design und bei der Durchführung der Studie genau bedacht und systematisch ausgeschlossen werden, da das Studienprotokoll in der Regel nicht veränderbar ist.

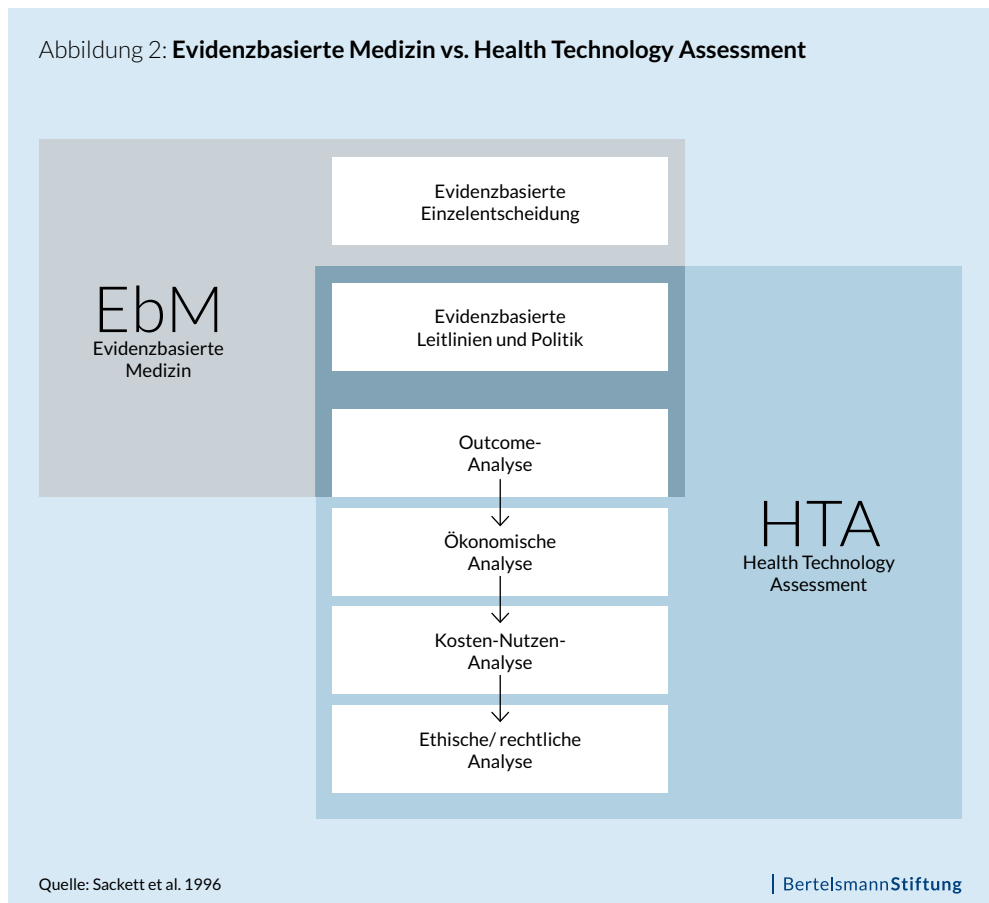
Bei neuen Ansätzen ist die Anpassung des Studiendesigns möglich durch veränderte Gewichtungen, die Einbeziehung von Baseline-Variablen, die Berücksichtigung von Anpassungen in der Intervention und die Präferenzen der Nutzer.

4 Nutzenbewertung

Neben Studien zum Nachweis der Wirksamkeit im Kontext der Zulassung bzw. Zertifizierung ist der Nutzen auch vor dem Hintergrund der Vergütung zu bewerten. Die geeignete Methode dafür ist das sogenannte Health Technology Assessment. Im Folgenden wird zunächst der klassische Ansatz dargelegt und dann ein neues Vorgehen, das prospektive HTA, dem gegenübergestellt.

4.1 Klassisches Health Technology Assessment

Health Technology Assessment (HTA) ist ein systematischer Prozess zur Analyse, Evaluierung und Bewertung wissenschaftlicher Informationen (vorwiegend publizierter Studien) über Auswirkungen bereits eingeführter und neuer gesundheitsrelevanter Verfahren, Leis-



tungen, Technologien und Prozeduren auf die Gesundheit bzw. Gesundheitsversorgung der Bevölkerung (DAHTA 2013). HTA ist also die systematische Analyse und Bewertung von Eigenschaften und Effekten oder anderen Einflüssen einer Technologie mit Gesundheitsbezug, um Entscheidungsträger zu informieren und die Anwendung der Technologie zu verbessern (ebd.).

Während die Evidenzbasierte Medizin (EbM) die Suche nach einer bestmöglichen Lösung für Patienten in den Vordergrund stellt, verfolgt HTA das Ziel, die beste Lösung für eine Entscheidung hinsichtlich der Gesundheitsversorgung insgesamt wissenschaftlich zu unterstützen (ebd.).

Die untersuchten inhaltlichen Dimensionen sind – analog zu klinischen Studien – Wirksamkeit und Sicherheit, darüber hinaus jedoch auch Wirtschaftlichkeit sowie soziale, ethische, rechtliche und organisatorische Aspekte (Gesundheit Österreich 2012). In HTA-Berichten wird angestrebt, die wesentliche verfügbare externe Evidenz zur jeweiligen Indikation oder Fragestellung aufzuführen und kritisch zu bewerten (vgl. Cochrane Deutschland).

Tabelle 4 gibt einen Überblick über Themen innerhalb der einzelnen Dimensionen eines HTA-Berichtes, welche die Deutsche Agentur für HTA beim Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) für wichtig erachtet.

Tabelle 4: Dimensionen und Themen bei HTA

Ethik	Gesundheitsökonomie	Medizin	Organisation	Recht	Sozialwissenschaften
Verteilungsgerechtigkeit	Systemimpact	Epidemiologie	Standort	Rechtliche Anforderungen	Zugangsgerechtigkeit
Fairness	Nachhaltigkeit	Krankheitslast	Infrastruktur	Grundrechte	Compliance
Autonomie	Folgekosten	Sicherheit	Ausbildung	Angewandetes Recht	Akzeptanz
Partizipation	Effektivität	Studienwirksamkeit	Qualitätssicherung		Nachhaltigkeit
Patientenwürde		Alltagswirksamkeit			
		Patientennutzen			

Quelle: DAHTA 2013

HTA-Berichte sind klassische Grundlagen für Entscheidungsträger bei Behörden, Kassen oder anderen Institutionen über die experimentelle Wirksamkeit (Efficacy), die Wirksamkeit unter Alltagsbedingungen (Effectiveness), die Sicherheit (Safety) und den gesundheitsökonomischen Stellenwert (Cost-Effectiveness) sowie den sozialen, ethischen und rechtlichen Rahmen der jeweiligen Fragestellung (Cochrane Deutschland). Originäres Ziel von HTA ist es, „Value for Money“ in der Gesundheitsversorgung herzustellen (ebd.).

Ein HTA-Bericht kann verschiedene Adressaten haben und zu unterschiedlichen Zeitpunkten während des Lebenszyklus eines Produktes oder einer Technologie durchgeführt werden. Seit 2016 hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) den gesetzlichen Auftrag, einen Themencheck Medizin zu von Bürgerinnen und Bürgern vorgeschlagenen Themen durchzuführen (§ 139b SGB V) und pro Jahr fünf ausgewählte HTA-Berichte vorzulegen.

Einzelne Krankenkassen können ebenfalls HTAs durchführen (lassen), wenn es darum geht zu entscheiden, ob eine Technologie selektivvertraglich angeboten werden soll. Auch

der Medizinische Dienst des Spitzenverbandes Bund der Krankenkassen (MDS) führt HTAs durch und trägt damit zur Meinungsbildung über eine neue Technologie innerhalb des GKV-Spitzenverbandes bei. Nutzenbewertungen auf der Ebene des G-BAs sind der formalste Ansatz, bei dem HTA genutzt wird. Sie folgen strengen gesetzlichen und untergesetzlichen Regeln (vgl. G-BA 2017).

Der Zugang zu diesem System ist sehr formalisiert und eingeschränkt. Ein Hersteller von Medizinprodukten kann lediglich einen Anstoß dazu geben, eine neue Methode ins kollektive Vergütungssystem zu bringen. Dazu kann er beim G-BA einen Antrag auf Methodenbewertung nach § 137e SGB V stellen. Erst wenn der G-BA einer beantragten neuen Methode das Potenzial bescheinigt, eine erforderliche Behandlungsalternative darstellen zu können, ihr Nutzen aber noch nicht hinreichend belegt ist, kann er eine Erprobungsstudie initiieren. Darin soll dann der Nutzen der neuen Methode bewertet werden. Erst nach einem positiven Nutznachweis ist eine Methodenbewertung nach § 135 SGB V und die Aufnahme in den Leistungskatalog möglich.

4.2 Limitationen klassischer HTA

Klassische HTAs sind ein gutes Instrument, um bestehende Methoden und Verfahren zu analysieren und zu bewerten. Dies setzt jedoch einen gewissen Reifegrad einer Anwendung voraus, das heißt, klinische Studien liegen bereits vor und die Anwendung wird partiell schon genutzt und erprobt. HTAs sind damit wie beschrieben ein Bewertungsinstrument für Institutionen und Entscheidungsträger. Ein klassisches HTA wird jedoch nicht bereits in der Entwicklungsphase eines Produktes genutzt – es dient also nicht als strategisches Entscheidungsinstrument und zur Weichenstellung in der Entwicklung. Insbesondere wird es nicht als Planungsinstrument für den Evidenzaufbau verwendet. Dadurch besteht das Risiko, dass nach der Entwicklung bei der Prüfung des Marktzugangs Dimensionen und Parameter abgefragt und bewertet werden, die in der Entwicklung gar nicht bedacht wurden. Entwicklungssteuerung und Bewertung fallen damit auseinander, was im Lebenszyklus eines Produktes zu Komplikationen und Verzögerungen führen kann.

4.3 Prospective Health Technology Assessment

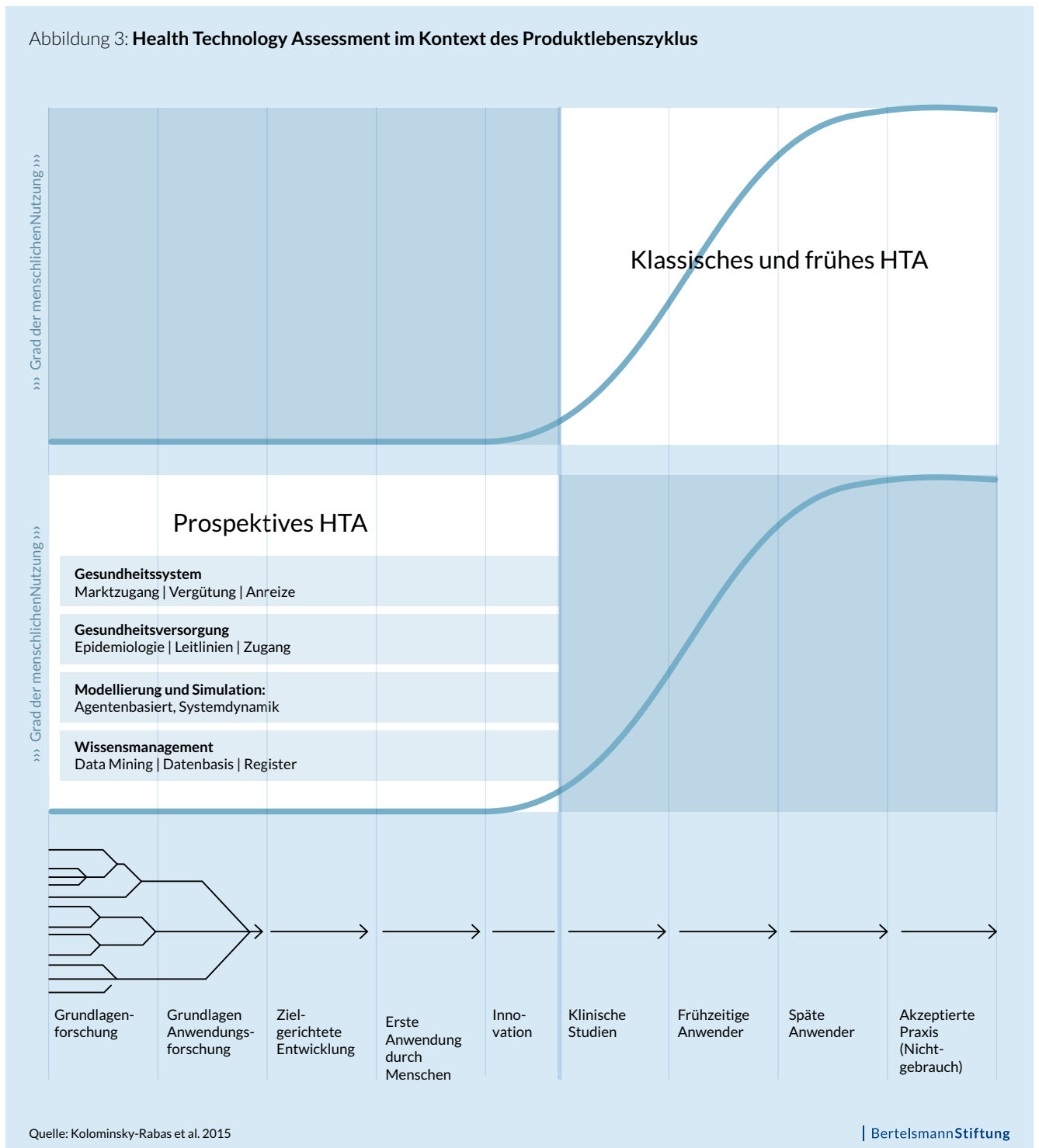
HTAs können (und sollten) auch von Unternehmen selbst im frühen Stadium einer Produktentwicklung durchgeführt werden. Ein frühes, prospektives HTA kann Einsatzpotenziale und Schwachstellen einer Innovation aufzeigen und helfen, Chancen und Risiken am Markt frühzeitig zu bewerten, und damit als strategisches Entscheidungsinstrument dienen (vgl. Kolominsky-Rabas et al. 2015).

Ein prospektives HTA zeigt Potenziale, aber auch Schwachstellen einer Innovation und beantwortet die Fragen: Was sind die Auswirkungen einer Innovation und wie muss eine Technologie aussehen, um den gewünschten Effekt hervorzurufen? Das Schaubild in Abbildung 3 verdeutlicht den ProHTA-Ansatz und seine zeitliche Positionierung im Entwicklungsprozess.

Ein Prospective Health Technology Assessment kann (und sollte) bereits sehr früh – deutlich vor den ersten klinischen Studien – während des Entwicklungsprozesses stattfinden. Innovationen können so frühzeitig an den Marktbedürfnissen ausgerichtet, Risiken für Fehlentwicklungen minimiert und Absatzchancen des Produktes verbessert werden (vgl.

Kolominsky-Rabas 2018). Auf Basis dieser Informationen können dann die Investitionsentscheidungen, auch für klinische Studien, fundierter getroffen werden. Die betrachteten Dimensionen eines Prospektiven HTA sind: Zielpopulation, Outcomes, ökonomischer Nutzen und Public-Health-Nutzen. Kontextfaktoren wie die demographische Entwicklung, Krankheitsverläufe, Gesundheitsleistungen und Finanzierung werden ebenfalls berücksichtigt.

Abbildung 3: Health Technology Assessment im Kontext des Produktlebenszyklus



4.4 Gegenüberstellung der Methoden

Um die Nutzbarkeit der neuen Methoden besser bewerten zu können, werden hier die Dimensionen mit den Verfahren abgeglichen. So kann auf einen Blick erfasst werden, ob ein Verfahren alle relevanten Aspekte abdeckt. Tabelle 3 stellt die Dimensionen eines HTA nach Anwendungsebenen dar.

Bei MDS und G-BA werden vor allem die klinische Evidenz, die Wirksamkeit, die Wirtschaftlichkeit und ethische Aspekte bewertet. Bei den Anbietern und Einzelkassen geht die Bewertung – je nach Anwendung – auch über diese Dimensionen hinaus.

Der wesentliche Unterschied zwischen klassischem und prospektivem HTA liegt in der Modellrechnung sowie der Basierung auf Annahmen.

Tabelle 5: Darstellung der HTA-Dimensionen nach Anwendungsebenen

Dimensionen	Anwendungsebene			
	Unternehmen (Prospektiver HTA)	Einzelkasse	MDS	G-BA
Klinische Evidenz				
Wirksamkeit				
Sicherheit				
Wirtschaftlichkeit				
Soziale Aspekte				
Ethische Aspekte				
Rechtliche Aspekte				
Organisatorische Aspekte				

■ Typischer Umfang von HTA nach Anwendungsebene (Akteur) und Dimension
 Quelle: Eigene Darstellung

Tabelle 6: Klassische und neue Methoden – Gegenüberstellung nach Forschungsdimensionen und Datentypen

	klassischer HTA	prospektiver HTA
Klinische Evidenz	Sekundärdaten	Annahmen und Sekundärdaten
Wirksamkeit	Sekundärdaten	Annahmen und Sekundärdaten
Sicherheit	Sekundärdaten	Annahmen und Sekundärdaten
Wirtschaftlichkeit	Sekundärdaten (u. a. Routinedaten)	Sekundärdaten (u. a. Routinedaten)
Soziale Aspekte	Sekundärdaten	Sekundärdaten
Ethische Aspekte	Sekundärdaten	Sekundärdaten
Rechtliche Aspekte	Sekundärdaten	Sekundärdaten
Organisatorische Aspekte	Sekundärdaten	Sekundärdaten

Quelle: Eigene Darstellung

5 Datentypen, Datengewinnung und Setting

Im Kontext der Digitalisierung kommt nicht nur agilen Methoden etwa für Produktentwicklung, Interventionsstudien und Technologiefolgeabschätzung eine große Bedeutung zu. Auch Daten wird eine größere Bedeutung zugeschrieben als in der Vergangenheit. Im Folgenden werden die wesentlichen klassischen Datentypen, die Formen der Datengewinnung und Settings mit neuen verglichen.

5.1 Datentypen

Datentypen, die bei Interventionsstudien und HTA in der Versorgungsforschung im Gesundheitswesen vorwiegend genutzt und unterschieden werden, sind Primärdaten sowie Sekundärdaten in Form von Studien und sogenannten Routinedaten. Als Gegensatz zu diesen überwiegend sehr strukturierten Daten sind Massendaten zu sehen, die auch unter dem Begriff „Big Data“ geführt werden:

Primärdaten: Werden eigene Studien aufgelegt und Daten zu einem Produkt oder Verfahren erhoben, sind dies sogenannte Primärdaten. Damit sollen wissenschaftliche Fragestellungen beantwortet und neue Erkenntnisse gewonnen werden (vgl. Röhrig, Prel und Blettner 2009).

Sekundärdaten: Bei der Forschung an Sekundärdaten werden bereits vorliegende Primärdaten, also schon durchgeführte und publizierte Studien analysiert. Damit wird ein Überblick über den aktuellen Stand der Forschung gegeben und daraus werden Schlüsse gezogen (vgl. Röhrig, Prel und Blettner 2009).

Routinedaten: Routinedaten (vgl. Springer Gabler) im Gesundheitswesen sind standardisierte Informationen, die vor allem zu Zwecken der Abrechnung mit den Leistungserbringern erhoben werden. Dies sind beispielsweise Daten der gesetzlichen Krankenversicherung (GKV), der Renten- und Unfallversicherung (Sozialdaten), aber auch der amtlichen Statistik (z. B. Krankenhausdiagnose- oder DRG-Statistik, siehe auch Vergütungssystem). Im Gegensatz zu Primärdaten, die eigens für die wissenschaftliche Verwendung erzeugt werden, handelt es sich bei Routinedaten um bereits vorliegende Daten, die zunächst für andere, nicht primär wissenschaftliche Zwecke erhoben wurden. Für die wissenschaftliche Nutzung können die versorgungsrelevanten Daten nach projektspezifischen Vorgaben ausgewählt und in der Regel pseudonymisiert verwendet werden.

Die Vorteile der GKV-Routinedaten liegen zunächst in der Größe der in wissenschaftlichen Auswertungen einschließbaren Population, in der Länge (Kontinuität) der beobachtbaren Zeiträume sowie in der Tatsache begründet, dass diese Daten einen weitgehend unverzerrten Blick auf die Versorgungsrealität gestatten. Da Routinedaten aus dem Versorgungsalltag

stammen, weisen sie in der Regel eine hohe externe Validität auf, sodass allgemeine Rückschlüsse auf die Grundgesamtheit bei einer ausreichend großen Stichprobe grundsätzlich möglich sind.

Routinedaten werden teilweise auch Sekundärdaten genannt, da sie aus vorherigen Erhebungen, Beobachtungen oder Experimenten stammen. Die Sekundärdatenanalyse beschreibt somit die wissenschaftliche Auswertung von Daten, die primär zu anderen Zwecken erhoben wurden.

Die Nutzung von Routinedaten im Gesundheitswesen hat in den vergangenen Jahren stetig zugenommen. Mittlerweile beschäftigen sich die unterschiedlichsten Fachdisziplinen mit Routinedaten. Ihr Einsatzspektrum reicht von der Versorgungsforschung über gesundheitsökonomische Analysen bis hin zur Verwendung für Modellierungen im Medical Decision Making und Health Technology Assessment (HTA). Auch der Gesetzgeber hat die Bedeutung dieser Datenquellen für die Gesundheitsberichtserstattung, Evaluation und Steuerung unseres Gesundheitssystems erkannt. Sekundärdatenanalysen werden zunehmend als Grundlage der Gesundheitswissenschaft und der Gesundheitspolitik genutzt.

Massendaten: Neben den klassischen werden im Zuge der Digitalisierung zunehmend neue Datentypen erschlossen. Hier entwickelt sich unter dem Namen „Big Data“ das Verständnis bzw. die Definition dieser neuen Datentypen ständig weiter. Derzeit kann der Begriff wie folgt definiert werden:

Der aus dem englischen Sprachraum stammende Begriff „Big Data“ bezeichnet Datenmengen, die beispielsweise zu groß, zu komplex, zu schnelllebig oder zu schwach strukturiert sind, um sie mit manuellen und herkömmlichen Methoden der Datenverarbeitung auszuwerten (vgl. Christl 2014). Im deutschsprachigen Raum ist der traditionellere Begriff „Massendaten“ gebräuchlich.

„Big Data“ wird häufig als Sammelbegriff für digitale Technologien verwendet, die in technischer Hinsicht für eine neue Ära digitaler Kommunikation und Verarbeitung und in sozialer Hinsicht für einen gesellschaftlichen Umbruch verantwortlich gemacht werden (vgl. Reichert 2014).

In der Definition von Big Data bezieht sich das „Big“ auf die drei Dimensionen Volume (Umfang, Datenvolumen), Velocity (Geschwindigkeit, mit der die Datenmengen generiert und transferiert werden) sowie Variety (Bandbreite der Datentypen und -quellen).¹ Erweitert wird diese Definition um die zwei Vs Value und Validity, die für einen unternehmerischen Mehrwert und die Sicherstellung der Datenqualität stehen (vgl. Bachmann, Gerzer und Kemper 2014).

5.2 Datenerhebung

Neben den klassischen Wegen der Datenerhebung wie klinischen Studien, Literaturrecherche für Sekundärdaten oder die Routinedatenerhebung der gesetzlichen Krankenversicherung haben sich im Kontext der Digitalisierung und der Massendaten auch neue Formen der Datenerhebung entwickelt. Hier gibt es viele Beispiele, wie Social-Media-Daten,

¹ „Big data is high-volume, high-velocity and high-variety in formation assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making“ (Gartner IT Glossary).

Tracking-Daten, Geo-Daten, Ambient-Assisted-Living-Daten und Befunddaten aus elektronischen Gesundheitsakten, die alle auch in der Versorgungsforschung und bei Digital Health relevant sein können.

Im Hinblick auf die Erhebung von Primärdaten und die Gewinnung von Studienteilnehmern werden hier exemplarisch das Apple HealthKit und ResearchKit vorgestellt, um die neuen Möglichkeiten zur Gewinnung von Probanden und Daten im Kontext von Digital Health zu illustrieren.

Das Open Source Framework HealthKit sammelt gesundheitsrelevante Daten der Nutzer und führt sie in einer App zusammen. Die Daten sollen nach Auskunft von Apple anonymisiert von renommierten universitären Forschungspartnern ausgewertet werden.

Dies soll Ärzten und Wissenschaftlern helfen, mehr und genauere Daten für die gesundheitliche und medizinische Forschung zu sammeln. Die so recht einfach zu realisierenden großen Fallzahlen, die breite Diversität der Teilnehmer und die lückenlosen Verläufe bei der Datenerhebung sind die Stärken dieses Verfahrens – vor allem im Hinblick auf Probanden mit seltenen Krankheiten.

Das iPhone kann dank seiner Elemente (Beschleunigungsmesser, Mikrofon, Gyroskop und GPS) ein leistungsstarkes Instrument der medizinischen Forschung sein, das durch die zahlreichen Apps, die über das HealthKit miteinander kommunizieren können, erweitert wird.

Mit dem ResearchKit ist es einfach, Studienteilnehmer zu rekrutieren und gleichzeitig ihre Einverständniserklärung einzuholen. Das Kit besteht aus drei grundlegenden Modulen: Einverständniserklärung (einholen), Umfrage, aktive Aufgabe.

Ein Nachteil bei der Nutzung des ResearchKits ist die Gefahr von Bias. Und offen bleibt die Frage, wie Basisinformationen eingeholt und überprüft werden können (Hat der Proband tatsächlich diese Krankheit)? Es fehlt grundsätzlich ein persönlicher Kontakt, bei dem Informationen überprüft und offene Fragen geklärt werden können. Ein Vorteil ist die Fähigkeit, jüngere Probanden zu erreichen sowie Gruppen, die traditionell eher klinisch isoliert sind (z. B. Teilnehmer in ländlichen Gebieten, Teilnehmer mit Geisteskrankheit).

Erforderlich ist weitere ethische und regulatorische Arbeit, bevor ResearchKit-Apps für den Einsatz in interventionellen, randomisierten Studien in Betracht gezogen werden können. Das größte Entwicklungspotenzial liegt bei der Verknüpfung dieser Apps mit elektronischen Gesundheitsakten, mit klinischen und Labor-Testwerten sowie genetischen Screens (vgl. Mohammadi 2015).

5.3 Setting

In der klassischen Welt der klinischen Studien werden Primärdaten unter Idealbedingungen erhoben und das Ziel ist, störende externe Faktoren möglichst zu eliminieren. Das bedeutet: Primärdaten werden unter Studienbedingungen in Studienzentren gewonnen – mit dem „wirklichen Leben“ („Real-Life“) von Patienten hat das wenig zu tun. Die Daten werden in der Regel manuell erhoben, also aktiv abgefragt und archiviert. Sie werden nicht automatisch mit einem technischen Hilfsmittel gesammelt. Die Datenerhebung erfolgt auch nicht kontinuierlich, sondern immer zu bestimmten Zeitpunkten.

Die neue Studienwelt unterscheidet sich deutlich von der klassischen. Untersucht wird jetzt in einer „Real-Life“-Umgebung, also in der Lebenswirklichkeit der Studienteilnehmer bzw. der Nutzer. Daten werden nicht mehr zu festgelegten Zeitpunkten in Studienzentren erhoben, sondern kontinuierlich automatisch mithilfe einer App gesammelt. Es entsteht eine neue Datengrundlage, die beispielsweise Geo-Daten und selbst erhobene Bewegungs- und Vitaldaten mit einbeziehen kann.

6 Implikationen

Digital Health stellt in vielerlei Hinsicht regulatorische Grundlagen, Prozesse und Strukturen des Gesundheitswesens infrage. In dieser Studienreihe wurde das bereits zu verschiedenen Themen sichtbar, wie den Vergütungssystemen, der Medizinproduktezertifizierung und der Markttransparenz. Eine Ursache liegt in der Andersartigkeit der Produkte selbst: Sie sind oftmals ein Hybrid aus Prozess und Produktinnovation, adressieren in vielen Fällen das Gesundheitshandeln der Versicherten und suchen entlang dieses Weges nach Integration in die Versorgungsprozesse des ersten Gesundheitsmarktes. Sie zeichnen sich durch sehr kurze Release-Zyklen aus, in denen die Lösungen häufig quartalsweise, teilweise laufend aktualisiert werden.

Neben diesen Produkteigenschaften liegt die eigentliche Ursache für den disruptiven Charakter jedoch tiefer und auch in der Organisation der Anbieter: Diese gehen in der Regel nach agilen und nicht nach klassischen, sequenziellen Modellen in der Produktentwicklung vor.

Das klassische Vorgehen zeichnet sich dadurch aus, dass beispielsweise eine Produktdefinition, eine Marktanalyse, eine ausführliche Konzeption, die Entwicklung, die Testung und das Qualitätsmanagement, die Zulassung/Zertifizierung und die Markteinführung von separaten Teams und sequenziell, also nacheinander, durchgeführt werden.

In agilen Modellen verorten Anbieter diese Prozessschritte in integrierten Teams, die viele der Schritte aus Analyse, Konzeption, Entwicklung, Testung, Qualitätssicherung gleichzeitig umsetzen. Hierzu wird das Gesamtvorhaben in Module gegliedert und in einzelnen Bearbeitungszeiträumen (Sprints) jeweils ein Modul bis zur fertigen Software entwickelt.

Diese innovative Arbeitsweise stößt nun bei der Markteinführung auf die überwiegend klassisch arbeitenden Organisationsstrukturen und Verfahren des Gesundheitswesens. Denn die anbieterinterne agile Produktentwicklung wurde bisher noch nicht integriert in die Evaluationsprozesse ausgewählter Forschungsinstitute, in die Verhandlungsprozesse streng hierarchisch organisierter Krankenkassen oder in die Zertifizierungsprozesse des Gesundheitswesens.

Aus den zwölf Grundprinzipien des agilen Vorgehens sind die folgenden im Zusammenhang mit Digital Health und klassischen Organisationsprinzipien des Gesundheitswesens als besonders relevant hervorzuheben: Interaktion zwischen den Menschen aller erforderlichen Disziplinen, ein iterativer Arbeitsprozess, eine kontinuierliche Erfolgsmessung und Kooperation mit den Kunden (Beck et al. 2001).

In der Medizinproduktezertifizierung und besonders im Bereich der Software sind bereits erste Methoden zur fortwährenden, iterativen Durchführung von Konformitätserklärungen in der Gesetzgebung angelegt: Durch die Etablierung eines Qualitätsmanagementsystems

kann der Anbieter die Konformität selbst kontinuierlich und im Einklang mit eigenen agilen Entwicklungs- und Release-Zyklen absichern und erklären. Das ist jedoch bei der Vergütung und vor allem beim Nachweis der Wirksamkeit sowie der Bewertung des Nutzens noch nicht etabliert – weder bei Anbietern noch bei den zuständigen Spitzenorganisationen der GKV oder bei Krankenkassen.

Im Folgenden werden Handlungsempfehlungen gegeben, die in der Summe die Umstellung der Begleitforschung für den Nachweis der Wirksamkeit und des Nutzens auf agile Prinzipien ermöglichen. Hier ist eine enge organisatorische Verzahnung erforderlich, die über die Organisationsstruktur des Anbieters hinausgeht. Sie muss die Organisation der Kunden, also in der Regel der Krankenkassen und begleitenden Forschungseinrichtungen, mit beinhalten. Darüber hinaus sind aber auch entsprechende Organe der gesetzlichen Krankenversicherung einzubeziehen, die für die Aufnahme von Leistungen in den GKV-Katalog verantwortlich sind. Diese müssen Ergebnisse aus agilen Studiendesigns für den Nachweis der Wirksamkeit und der Bewertung des Nutzens – hier: Methoden der Technologiefolgeabschätzung – akzeptieren. Im Idealfall werden sie bereits zu Beginn der Produktentwicklung in Abstimmung mit Interventionsstudie und Technologiefolgeabschätzung mit einbezogen; der Freigabeprozess für die Vergütung, wie etwa die Nutzenbewertung durch den G-BA, startet gleichzeitig.

In Bezug auf klassische Produkte im Gesundheitswesen wie Arzneimittel und Medizinprodukte ist der Standard für die Zulassung/Zertifizierung bzw. die Vergütung der Nachweis der Wirksamkeit in Form von Interventionsstudien mit Randomisierung und Kontrollgruppe (RCT). Dieser Standard wird meistens im Anschluss an die Produktentwicklung angewendet und ist dann Grundlage für eine darauffolgende Nutzenbewertung im Rahmen der Vergütungsverhandlung.

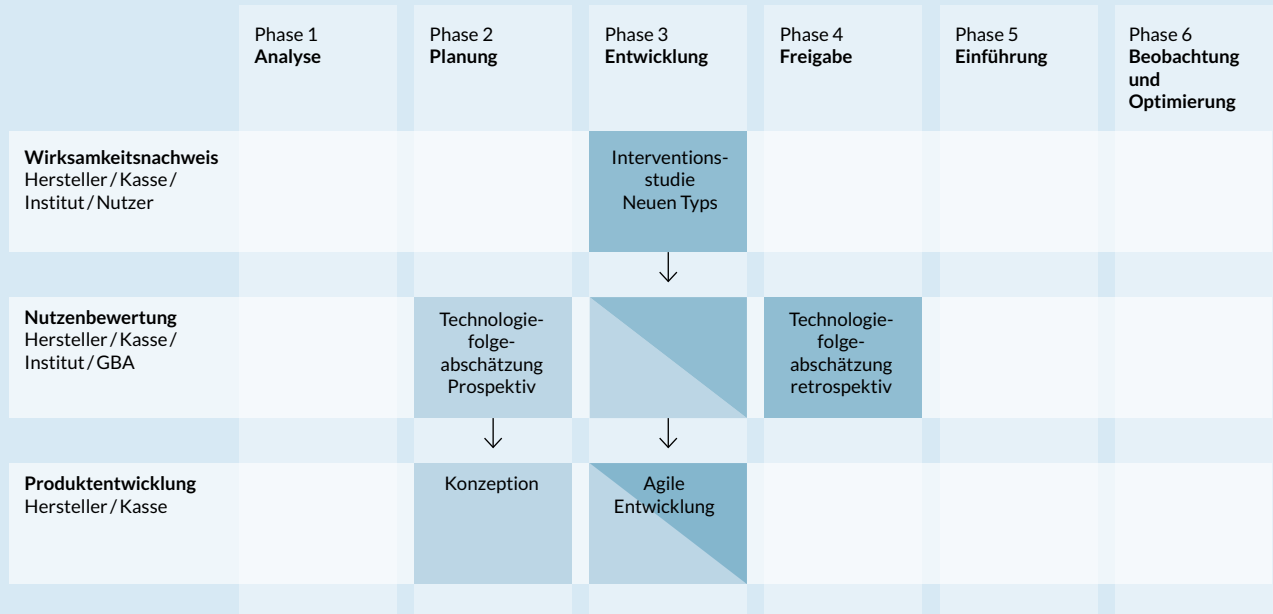
Insgesamt ist hier im Zusammenspiel aller relevanten Akteure ein grundsätzlich neues Verfahren erforderlich, das sich in bestimmten Punkten an den Werten und Grundprinzipien des agilen Vorgehens orientiert. Wesentlich ist dabei, die Bewertung des Nutzens und den Nachweis der Wirksamkeit in den Prozess der Produktentwicklung zu integrieren und auch die Zusammenarbeit mit Kunden (Krankenkasse) sowie einer Forschungseinrichtung für die Interventionsstudie (auch die Technologiefolgeabschätzung) in dieser Phase bereits mit einzubeziehen. Die hierfür wesentlichen Grundprinzipien der Agilität sind in den folgenden Eckpunkten abstrakt erfasst:

- **Ausrichtung an Ergebnissen:** Ausrichtung auf eine ganzheitliche Sicht der Ergebnisse – das heißt: Wirksamkeit und Nutzen vom ersten Schritt der Produktentwicklung an
- **Iteratives Vorgehen:** Prospektive und iterative Schätzung, Messung und Optimierung von Wirksamkeit und Nutzen in der Produktentwicklung
- **Interaktion:** Organisatorische Integration der Entwicklungsprozesse der Hersteller in die Begleitforschung von Instituten
- **Kooperation mit Kunden:** Kooperation mit den Kunden – das heißt in der Regel Krankenkassen – über die Produktentwicklung bis hin zur Ausgestaltung der Vergütung

Die wesentlichen Vorteile des Vorgehens:

- **Optimierung von Nutzerpräferenzen und Wirksamkeit:** Die Nutzerpräferenzen werden in den Interventionsstudien neuen Typs deutlich stärker mit einbezogen und somit auch in die Produktentwicklung. Dies unterstützt gerade bei verhaltensänderungsbezogenen Anwendungen die Adhärenz und insgesamt die Wirksamkeit.

Abbildung 4: Agile Produktentwicklung mit integriertem Wirksamkeitsnachweis und Nutzenbewertung



■ Niedrige Evidenz – Modellrechnungen und Entwicklungstätigkeit überwiegend basierend auf „Sekundärdaten anderer Studien und Annahmen“
■ Hohe Evidenz – Studie, Modellrechnung und Entwicklungstätigkeiten basierend auf „Sekundärdaten anderer Studien und Primärdaten aus eigener Studie“
 Quelle: Eigene Darstellung

| BertelsmannStiftung

Tabelle 7: Klassische und neue Methoden – Gegenüberstellung nach Forschungsdimensionen und Datentypen

	klassischer RCT	CEEBIT, MOST, SMART Micro-Randomized Trial	klassischer HTA	prospektiver HTA
Klinische Evidenz	Primärdaten (Labor-Setting)	Primärdaten (Real-Life-Setting)	Sekundärdaten	Annahmen und Sekundärdaten
Wirksamkeit	Primärdaten (Labor-Setting)	Primärdaten (Real-Life-Setting)	Sekundärdaten	Annahmen und Sekundärdaten
Sicherheit	Primärdaten (Labor-Setting)		Sekundärdaten	Annahmen und Sekundärdaten
Wirtschaftlichkeit			Sekundärdaten (u. a. Routinedaten)	Sekundärdaten (u. a. Routinedaten)
Soziale Aspekte			Sekundärdaten	Sekundärdaten
Ethische Aspekte			Sekundärdaten	Sekundärdaten
Rechtliche Aspekte			Sekundärdaten	Sekundärdaten
Organisatorische Aspekte			Sekundärdaten	Sekundärdaten

Quelle: Eigene Darstellung

- **Optimierung des Nutzens:** Die Ergebnisorientierung wird bei einer Konzeption, die maßgeblich von prospektiver Technologiefolgeabschätzung gelenkt wird, verbessert und der Nutzen so maximiert.
- **Kurze Entwicklungszeit:** Die Parallelisierung und Integration von Interventionsstudie, Technologiefolgeabschätzung und Freigabe ermöglichen einen zeitoptimierten Ablauf und somit eine Absicherung der Innovationsfähigkeit im Gesundheitswesen.
- **Niedrige Entwicklungskosten:** Durch die überwiegend automatisierte, digitale Datenerhebung, den direkten ortsunabhängigen Zugang zu Probanden und den integrativen Ansatz erhöht sich die Effektivität in der Produktentwicklung und die Entwicklungskosten werden minimiert.

Anhand der Forschungsdimensionen lässt sich der Unterschied zwischen den einzelnen Methoden verdeutlichen. Das Spektrum reicht von ausschließlich klinischer Evidenz bis hin zu einer breiten Betrachtung im Rahmen eines HTA. Klinische Evidenz wird in allen Verfahren behandelt. Am weitesten darüber hinaus geht naturgemäß ein HTA. Allerdings baut der HTA auf den Ergebnissen der primärdatenbasierten Studien auf.

Im Folgenden wird das Vorgehen mittels eines Sets an Handlungsempfehlungen operationalisiert und nach Methoden bzw. relevanten Akteuren geordnet.

Handlungsempfehlung 1 (an Hersteller und Forschungseinrichtungen): Neue, agile Designs für Interventionsstudien als integralen Bestandteil der Produktoptimierung anwenden – anstelle nachgelagerter, klassischer RCTs

Mögliche agile Studiendesigns zur Messung der Wirksamkeit sind u. a. MOST, CEEBIT, SMART oder Micro-Randomized Trials (siehe oben). Sie sind geeignet, entlang der Release-Zyklen eine Ergebnisorientierung auch wissenschaftlich zu fundieren.

Die Methoden sollten mit den kurzen Entwicklungszyklen und den agilen Entwicklungsmethoden von Digital-Health-Anwendungen synchronisiert werden. Dies kann über eine Reflexion der Ergebnisse entlang der Release- und Sprintplanung erfolgen. Geachtet werden sollte auf eine regelmäßige Lieferung von Zwischenergebnissen zu einzelnen Modulen und Modulkombinationen an die Produktentwicklung und zur Aktualisierung der Nutzenbewertung.

Die kontinuierliche Ergebnismessung liefert – im Gegensatz zu klassischen Methoden – bereits während der laufenden Entwicklung Feedback, schützt vor Fehlentwicklungen und hilft trotz anfänglichem Mehraufwand, die begrenzten Ressourcen von Start-ups optimiert einzusetzen.

Handlungsempfehlung 2 (an Hersteller und Forschungseinrichtungen): Neue Kanäle zur Rekrutierung von Studienteilnehmern nutzen und Anwendungen im Real-Life-Setting erproben

Digitale Vertriebskanäle wie App-Stores, Apps und soziale Netzwerke sollten genutzt werden, um – im Gegensatz zur klassischen Ansprache von Studienprobanden über feste Studienzentren – einen ortsungebundenen Einschluss von Patienten zu ermöglichen.

Damit geht auch eine Erprobung der Anwendungen im Real-Life-Setting einher. Die Erprobung der Wirksamkeit unter Laborbedingungen wie in klassischen Studien wird hier

übersprungen, da die Digital-Health-Anwendungen ohnehin über die Beeinflussung des Gesundheitshandelns stark mit dem Real-Life-Setting interagieren. Diese Effekte sollten bereits im Studiendesign mitberücksichtigt und gemessen werden.

Die Studienteilnehmer über digitale Vertriebskanäle anzuwerben, kann Aufwände für die Rekrutierung verringern.

Handlungsempfehlung 3 (an Hersteller und Forschungseinrichtungen): Daten über die digitalen Anwendungen erheben

In klassischen Studiendesigns werden die Daten im Allgemeinen durch medizinisches Fachpersonal erhoben. Üblicherweise sind zwei bis fünf Messpunkte geplant, die über relativ weite Beobachtungszeiträume verteilt sind.

Neue digitale Instrumente wie das „Apple ResearchKit“ sowie Tracking- und Aktenlösungen ermöglichen unter Einbezug der Sensoren der mobilen Endgeräte, die Probanden effizient und engmaschig zu beobachten. Es können sehr viel engere Zusammenhänge zwischen der Intervention und den Effekten hergestellt werden.

Diese Möglichkeiten sollten – auch um Aufwände bei der Datenerhebung zu reduzieren – systematisch genutzt werden und bei der Ausgestaltung der Studiendesigns mit einfließen.

Handlungsempfehlung 4 (an Hersteller, Forschungseinrichtungen, Krankenkassen, G-BA): Prospektive Technologiefolgeabschätzung als Ausgangspunkt und integralen Bestandteil der Produktentwicklung einsetzen

Bei der klassischen Herangehensweise folgt die Technologiefolgeabschätzung im Anschluss an den Wirksamkeitsnachweis, zum Beispiel durch eine RCT-Interventionsstudie. Diese dient primär dazu, bei der Vergütungsverhandlung in der GKV Preise zu bestimmen.

Es ergibt sich aber erhebliches Potenzial, wenn die Erkenntnisse aus der Technologiefolgeabschätzung als Ausgangspunkt für die Produktentwicklung genutzt werden. Hierzu kann das Grundgerüst eines klassischen HTA schon in der frühen Phase der Produktentwicklung aufgesetzt werden. Die Wirksamkeit von vergleichbaren Interventionen liefert wichtige Anhaltspunkte für den Zielbereich und für die Wirksamkeit der jeweiligen Lösung.

Die zunächst fehlenden Daten und Informationen zur Wirksamkeit der eigenen Lösung können in einer Modellrechnung zum Nutzen ersetzt werden durch Annahmen. Diese sind so weit wie möglich an Vergleichswerten aus Studienergebnissen orientiert, die die Wirksamkeit vergleichbarer Produkte und/oder in dem Indikationskontext üblicher Produkte und Verfahren nachweisen. Diese „Grundbefüllung“ mit Wahrscheinlichkeiten und anderen Werten kann schrittweise mit jeder Iteration der eigenen Wirksamkeitsstudien ersetzt werden. So kann bei jedem Turnus die stringente Optimierung auf den Nutzen umgesetzt werden.

Handlungsempfehlung 5 (an Hersteller, Forschungseinrichtungen, Fachgesellschaften und Krankenkassen): Entwicklungspartnerschaften eingehen und neue Verfahren für Wirksamkeit und Technologiefolgeabschätzung akzeptieren

Bei der Produktentwicklung sollten Krankenkassen, Forschungseinrichtungen und Fachgesellschaften stärker als bisher mit einbezogen werden. Ihre Rolle kann u. a. sein:

- Beratung bei der Konzeption und Entwicklung hinsichtlich der Produkthanforderungen
- Beratung zur Methodik der Technologiefolgeabschätzung und Bereitstellung von Datengrundlagen (GKV-Routinedaten) für die prospektive Technologiefolgeabschätzung hinsichtlich der Optimierung des Nutzens
- Beratung zum Studiendesign für agile Interventionsstudien und zu Zwischenergebnissen im Kontext der Optimierung der Wirksamkeit und Präferenzgerechtigkeit

Handlungsempfehlung 6 (an den G-BA): Erprobungsrichtlinie § 137e und § 137h SGB V für den ambulanten und stationären Sektor weiterentwickeln und ein frühes, iteratives Verfahren zur Nutzenbewertung ermöglichen

Bei den Verfahren zur Methodenbewertung beim Gemeinsamen Bundesausschuss im Rahmen des § 137e SGB V und des § 137h SGB V sollten neue, agile Studiendesigns berücksichtigt werden. Dies gilt insbesondere bei der Bewertung des Potenzials einer neuen Untersuchungs- und Behandlungsmethode nach § 137e Abs. 1 SGB V und § 137h Abs. 1 SGB V, wenn diese Methode auf einer Digital-Health-Anwendung beruht. Die innovativen Studiendesigns sollten auch bei der Durchführung von Erprobungsstudien berücksichtigt werden. Damit wird die Basis für den Zugang von Digital-Health-Anwendungen zum ersten Gesundheitsmarkt erweitert. Hersteller von neuen, agilen Anwendungen haben so bessere Marktzugangschancen, ohne dass beim Nachweis des Nutzens für die Versicherten und Patienten Abstriche gemacht werden müssen.

Handlungsempfehlung 7 (an Politik und Selbstverwaltung): Forschungseinrichtungen für die Anwendung und Weiterentwicklung von agilen Interventionsstudien und HTAs in Kombination fördern

In Teil 2 dieser Reihe – „Bedarfsgerechte Innovations- und Forschungsförderung“ – wurde bereits der Förderbedarf im Bereich Digital Health ausführlich beleuchtet; der Fokus lag auf einem Programm für eine bedarfsgerechte Innovations- und Forschungsförderung. Das Programm sieht eine Förderung der Begleitforschung für Digital Health vor. Die Methodik für die Begleitforschung kann diesem Teilbericht 4 entnommen werden. Sie umfasst die oben genannten Methoden zu agilen Interventionsstudien und prospektiver Technologiefolgeabschätzung. Hierzu sollten Anbieter und Forschungseinrichtungen gezielt gefördert werden. Eine Etablierung dieser Methoden unterstützt die Innovations- und Wettbewerbsfähigkeit am Standort Deutschland im Bereich Digital Health.

7 Literatur

- Albrecht, U.-V., B. Kuhn, J. Land, V. E. Amelung und U. von Jan (2018). Nutzenbewertung von digitalen Gesundheitsprodukten (Digital Health) im gesellschaftlichen Erstattungskontext. Bundesgesundheitsblatt 24.1.2018.
- Bachmann, R., T. Gerzer und D. G. Kemper (2014). Big Data – Fluch oder Segen? – Unternehmen im Spiegel gesellschaftlichen Wandels. Heidelberg u. a.
- Baethge, C. (2014). Evidenzbasierte Medizin: In der Versorgung angekommen, aber noch nicht heimisch. Dtsch Arztebl 111 (39): A-1636 / B-1416 / C-1348.
- Banta, D. (1976). Development of Medical Technology: Opportunities for Assessment. United States Congress: Office of Technology Assessment.
- Beck, K., et al. (2001). Manifesto for Agile Software Development. <http://agilemanifesto.org> (Zugriff 14.2.2018).
- Bertelsmann Stiftung (Hrsg.) (2016). Digital-Health-Anwendungen für Bürger. Kontext, Typologie und Relevanz aus Public-Health-Perspektive. Entwicklung und Erprobung eines Klassifikationsverfahrens. Gütersloh.
- Bertelsmann Stiftung (Hrsg.) (2017a). Transfer von Digital-Health-Anwendungen in den Versorgungsalltag. Teil 2: Bedarfsgerechte Innovations- und Forschungsförderung: Innovationspotenzial, Förderbedarf und Implikationen. Gütersloh.
- Bertelsmann Stiftung (Hrsg.) (2017b). Transfer von Digital-Health-Anwendungen in den Versorgungsalltag. Teil 3: Medizinproduktezertifizierung – Rechtsgrundlage, Risikoklassifizierung und Implikationen. Gütersloh.
- Bertelsmann Stiftung (Hrsg.) (2017c). Transfer von Digital-Health-Anwendungen in den Versorgungsalltag. Teil 5: Vertrags- und Vergütungsformen in der gesetzlichen Krankenversicherung – Gegenstand, Verfahren und Implikationen. Gütersloh.
- Christl, W. (2014). Kommerzielle digitale Überwachung im Alltag. Wien. http://crackedlabs.org/dl/Studie_Digitale_Ueberwachung.pdf (Zugriff 14.2.2018).
- Cochrane Deutschland. Health Technology Assessment (HTA). www.cochrane.de/de/hta (Zugriff 14.2.2018).
- Collins, L. M., S. A. Murphy und V. Strecher (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent ehealth interventions. American Journal of Preventive Medicine 32 (5): 112–118.
- Deutsches Netzwerk für evidenzbasierte Medizin. www.ebm-netzwerk.de/was-ist-ebm/images/evidenzklassen.jpg/view (Zugriff 14.2.2018).
- DAHTA – Deutsche Agentur für Health Technology Assessment des DIMDI. Methodenhandbuch HTA, Draft-Version 2013. www.dimdi.de (Zugriff 14.2.2018).
- Ex, P., R. Busse und C. Henschke (2016). Die Nutzenbewertung von nichtmedikamentösen Untersuchungs- und Behandlungsmethoden. G&S Gesundheits- und Sozialpolitik 70 (2): 48–55.
- Gartner: IT Glossary. Big Data. www.gartner.com/it-glossary/big-data (Zugriff 15.1.2018).

-
- G-BA – Gemeinsamer Bundesausschuss (2017). Verfahrensordnung des Gemeinsamen Bundesausschusses, in Kraft getreten am 5.8.2017. www.g-ba.de/downloads/62-492-1436/VerfO_2017-04-20_iK-2017-08-05.pdf (Zugriff 15.1.2018).
- Gesundheit Österreich GmbH (2012). Methodenhandbuch für Health Technology Assessment, Version 1. Wien. <https://jasmin.goeg.at/121/1/Methodenhandbuch%20f%C3%BCr%20Health%20technology%20Assessment%20Version%201.2012.pdf> (Zugriff 15.10.2017).
- Graalman, J., M. Czaja und T. Rödiger (2017). Diskussionspapier: Digital-AMNOG mit CE+Zertifizierung. Wie können wir das GKV-Innovationsdilemma für Digital-Health-Innovationen auflösen? BzG 9-17 Beiträge zur Gesellschaftspolitik.
- Gwadz, M. V., L. M. Collins, C. M. Cleland, N. R. Leonard, L. Wilton, M. Gandhi, R. Scott Braithwaite, D. C. Perlman, A. Kutnick und A. S. Ritchie (2017). Using the multiphase optimization strategy (MOST) to optimize an HIV care continuum intervention for vulnerable populations: a study protocol. *BMC Public Health* 17 (1): 383. www.ncbi.nlm.nih.gov/pubmed/28472928 (Zugriff 15.1.2018).
- Hendela, T. Apple Press Info. 2015 Mar 09. Apple Introduces ResearchKit, Giving Medical Researchers the Tools to Revolutionize Medical Studies. www.apple.com/ca/pr/library/2015/03/09Apple-Introduces-ResearchKit-Giving-Medical-Researchers-the-Tools-to-Revolutionize-Medical-Studies.html (Zugriff 15.1.2018).
- Hermann-Frank, A., und M. Lelgemann (2013). Nutzenbewertung von Medizinprodukten: Datenlage bei der Einführung in die Versorgung. 12. Deutscher Kongress für Versorgungsforschung. Berlin, 23.-25.10.2013. Düsseldorf.
- IQWiG – Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2014). Allgemeine Methoden, Version 4.2.
- IQWiG – Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2018). IQWiG-Berichte – Nr. 569: Analyse klinischer Prüfungen von Medizinprodukten; Arbeitspapier GA 14-05; Version 1.0, Stand 15.12.2017.
- Jardine, J., J. Fisher und B. Carrick (2015). Apple's ResearchKit: smart data collection for the smartphone era? *Journal of the Royal Society of Medicine* 108 (8): 294–296.
- Klasnja, P., E. B. Hekler, S. Shiffman, A. Boruvka, D. Almiral, A. Tewari und S. A. Murphy (2015). Micro-Randomized Trials: An experimental design for developing just in time adaptive interventions. *Health Psychol* 34 (0): 1220–1228.
- Kolominsky-Rabas, P., A. Djanatljev, P. Wahlster, M. Gantner-Bär, B. Hofmann, R. German, M. Sedlmayr, E. Reinhardt, J. Schüttler und C. Kriza (2015). Technology foresight for medical device development through hybrid simulation – the ProHTA project, *Technological Forecasting and Social Change* 97: 105.
- Kolominsky-Rabas, P. (2018): ProHTA. Prospective Health Technology Assessment. Ein innovativer Ansatz für die strategische Planung Ihres Medizinprodukteportfolios. www.prohta.de/index.html (Zugriff 15.1.2018).

-
- Lei, H., I. Nahum-Shani, K. Lynch, D. Oslin und S. A. Murphy (2012). A „SMART“ design for building individualized treatment sequences. *Annual Review of Clinical Psychology* 8: 21–48.
- Liao, P., P. Klasnja, A. Tewari und S. A. Murphy (2015). Micro-Randomized Trials in mHealth. Department of Statistics, University of Michigan, Ann Arbor, April 2, 2015. <https://arxiv.org/pdf/1504.00238.pdf> (Zugriff 15.9.2017).
- Mohr, D. C., K. Cheung, S. M. Schueller, C. Hendricks Brown und N. Duan (2013). Continuous evaluation of evolving behavioral intervention technologies. *American Journal of Preventive Medicine* 45 (4): 517–523.
- Mohammadi, D. (2015). ResearchKit: A clever tool to gather clinical data. *The Pharmaceutical Journal* 294: 781–782.
- Pham, Q., D. Wiljer und J. A. Cafazzo (2016). Beyond the Randomized Controlled Trial: A Review of Alternatives in mHealth Clinical Trial Methods. *JMIR mhealth uhealth* 4(3): e107. <http://mhealth.jmir.org/2016/3/e107/> (Zugriff 15.10.2017).
- Reichert, R. (2014). Big Data: Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie. Bielefeld.
- Röhrig, B., J.-B. Prel und M. Blettner (2009). Studiendesign in der medizinischen Forschung. *Dtsch Arztebl Int* 106 (11): 184–189.
- Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen (2014). Bedarfsgerechte Versorgung – Perspektiven für ländliche Regionen und ausgewählte Leistungsbereiche. Berlin.
- Sackett, D. L., W. Rosenberg, J. Gray et al. (1996). Evidence based medicine: what it is and what it isn't. *BMJ* 312c (7023): 7. www.pharma-economics.com/search/label/HTA (Zugriff 15.1.2018).
- Sauerland, S. (2017). Medizinprodukte zwischen schneller Innovation und fraglichem Nutzen. *GGW* 17 (4): 25–30.
- Springer Gabler Verlag (Hrsg.). Gabler Wirtschaftslexikon. Routinedaten im Gesundheitswesen. <http://wirtschaftslexikon.gabler.de/Archiv/-2046631400/routinedaten-im-gesundheitswesen-v2.html> (Zugriff 15.1.2018).
- Statista. Statistik Lexikon. Bias – Verzerrung. https://de.statista.com/statistik/lexikon/definition/35/bias_verzerrung/ (Zugriff 15.1.2018).
- Wilbur, J. E., A. M. Kolanowski und L. M. Collins (2016). Utilizing MOST frameworks and SMART designs for intervention research. [www.nursingoutlook.org/article/S0029-6554\(16\)30054-9/pdf](http://www.nursingoutlook.org/article/S0029-6554(16)30054-9/pdf) (Zugriff 15.1.2018).
- Windeler J., und S. Lange (2015). Nutzenbewertung medizinischer Leistungen im deutschen Gesundheitswesen – rechtlicher Rahmen, historische und internationale Perspektive. *Bundesgesundheitsblatt* 58 (3): 220–226.
- Windeler, J. (2006). Nutzen und Nutzenbewertung. *Dtsch Med Wochenschr* 131: 12–15.

8 Anhang

Sieben Typen von Digital-Health-Anwendungen

Tabelle 8: **Digital-Health-Anwendungen – sieben Typen**

Anwendungstyp	Erläuterung
I Stärkung der Gesundheitskompetenz	Dieser Typ umfasst primär Anwendungen, die den Bürgern Wissen rund um Gesundheits- oder Krankheitsanliegen verfügbar machen. Dadurch kann ein selbstbestimmtes Gesundheitshandeln innerhalb und außerhalb der Leistungssektoren des Gesundheitswesens gefördert werden. Es handelt sich dabei überwiegend um kostenlose Web-Anwendungen, die sich gleichermaßen an Gesunde und Kranke richten, z. B. gesundheitsinformation.de, weisse-liste.de, aidshilfe.de und labtestsonline.de.
II Analyse und Erkenntnis	Anwendungen dieses Typs erfassen gesundheits- und umweltbezogene Informationen punktuell und werten sie im persönlichen Kontext aus. Die Anwender erhalten eine differenzierte Entscheidungsgrundlage in Form von Einschätzungen, Bewertungen und Diagnosen. Es sind überwiegend kostenlose mobile Anwendungen mit Sensor- und Cloud-Anbindung, z. B. Mimi Hörtest und LärmApp.
III Indirekte Intervention	Förderung von Selbstwirksamkeit, Adhärenz und Sicherheit: Dieser Typ erfasst kontinuierlich gesundheits- und umweltbezogene Informationen und wertet sie im persönlichen Kontext aus. Üblich sind die Überwachung im Zeitverlauf, automatisierte Motivation und die Einbindung in soziale Netzwerke. So kann eine Stärkung der Selbstwirksamkeit, Adhärenz und Sicherheit bewirkt werden. Es sind vornehmlich kostenlose mobile Anwendungen mit Sensor- und Cloud-Nutzung, z. B. runtastic, Stimmungstagebuch (Depression), Pillreminder und Babytracker Pro.
IV Direkte Intervention	Veränderung von Fähigkeiten, Verhalten und Zuständen: Anwendungen dieses Typs umfassen therapeutische Funktionen wie eine technisch unterstützte Erweiterung der menschlichen Fähigkeiten (z. B. Sehen – elektronische Lupe), ein Anleiten und Lehren (z. B. Psychoedukation, Onlinekurse, Tutorials) und/oder eine Veränderung von Gesundheitszuständen etwa durch audiovisuelle Stimulation. Es sind größtenteils kostenpflichtige mobile Anwendungen mit Sensor-Anbindung, z. B. Tinnitracks, caterna.de und deprexis.de.
V Dokumentation	Anwendungen dieses Typs bieten eine elektronische Gesundheits- und/oder Krankheitsakte, in der etwa Befunde, Diagnosen, Therapien und Vitaldaten gespeichert, verwaltet und in Bezug gesetzt werden können. Es handelt sich vorrangig um kostenlose mobile Web- und Systemanwendungen mit Sensor- und Cloud-Anbindung, z. B. Apple Health, S Health und healthvault.com.
VI Organisation und Verwaltung	Dieser Typ umfasst vor allem Anwendungen, die der Organisation und Verwaltung der Versorgung dienen, etwa Leistungserbringertermine, Prozesse der Sozialverwaltung wie Genehmigungen und Abrechnungen. Es sind überwiegend kostenlose mobile und Web-Anwendungen, z. B. tk.de (Online-Geschäftsstelle), APP zum Leistungserbringer und Arzttermine.de.
VII Einkauf und Versorgung	Dieser Typ umfasst primär Anwendungen, die der Versorgung einzelner Personen mit gesundheits- und krankheitsbezogenen Produkten – Hilfsmittel, Nahrungsergänzungsmittel, Arzneimittel etc. – dienen. In der Regel sind darunter Onlineshops (E-Commerce) zu verstehen. Es sind überwiegend kostenlose mobile und Web-Anwendungen wie z. B. docmorris.de.

Quelle: Bertelsmann Stiftung (2016)

Autoren und Autorin

Karsten Knöppler, Diplom-Betriebswirt, ist Geschäftsführer der fbeta GmbH, Experte und Berater für die Themen „Gesundheits- und Versorgungsmanagement“ sowie „Gesundheits-IT“. Im Projekt war er als Leiter und Experte tätig. Zuvor war er u. a. Geschäftsbereichsleiter der DV-Steuerung im AOK-Bundesverband, Geschäftsbereichsleiter der gevko in der AOK Systems und Berater im IGES Institut mit den Schwerpunkten „Krankenkassen“ und „Neue Versorgungsformen“. Zudem hat er im Kontext der Disease-Management-Programme in der Versorgungsforschung, Entwicklung und Einführung u. a. bei ANYCARE gearbeitet. Karsten Knöppler hat Internationale Betriebswirtschaft studiert.
kk@karstenknoeppler.de

Sebastian Hesse, Jurist und MSc Public Health, ist Berater für strategische Public Affairs, Gesundheitspolitik und Gesundheits-IT sowie für Marktzugangsfragen bei Medizinprodukten bei der fbeta GmbH. Im Projekt war er als Experte tätig. Zuvor war er u. a. Senior Director Government Affairs bei der Siemens AG in Berlin und strategischer Berater für Public Affairs in Brüssel. Außerdem war als EU Liaison & Research Officer beim European Health Forum Gastein und als Projektmanager im Internationalen Netzwerk Gesundheitspolitik bei der Bertelsmann Stiftung tätig. Sebastian Hesse hat Rechtswissenschaften an der Freien Universität Berlin und Public Health an der London School of Hygiene & Tropical Medicine und der LSE studiert.

Patricia Ex, Erasmus Mundus Master in Euroculture, ist Expertin für das Thema „Finanzierung im Gesundheitswesen“ sowie den Marktzugang von neuen Gesundheitstechnologien. Im Projekt war sie als Expertin tätig. Zuvor war sie u. a. Senior Consultant für Public Affairs, wissenschaftliche Mitarbeiterin im Deutschen Bundestag und Leiterin des Hauptstadtbüros eines Facharztverbandes. Sie ist Doktorandin am Fachbereich Management im Gesundheitswesen der TU Berlin. Patricia Ex hat englische Linguistik und internationale Politik an den Universitäten in Göttingen, Bilbao und Bremen sowie am Dickinson College in Pennsylvania/USA studiert.

Impressum

© März 2018
Bertelsmann Stiftung

Verantwortlich
Uwe Schwenk

Ansprechpartner
Timo Thranberend
Dr. Thomas Kostera

Titelbild
Getty Images/iStock-
photo/Cecilie_Arcurs,
Shutterstock/Billion
Photos, Shutterstock/
Dean Drobot

Gestaltung
Dietlind Ehlers

Adresse | Kontakt

Bertelsmann Stiftung
Carl-Bertelsmann-Straße 256
Postfach 103
33311 Gütersloh

Timo Thranberend
Senior Project Manager
Telefon: +49 5241 81-81117
Telefax: +49 5241 81-681117
timo.thranberend@bertelsmann-stiftung.de

Dr. Thomas Kostera
Project Manager
Telefon: +49 5241 81-81204
Telefax: +49 5241 81-681204
thomas.kostera@bertelsmann-stiftung.de

www.bertelsmann-stiftung.de