

# Digitale Primärprävention in der GKV: Möglichkeiten, Grenzen und Ansätze zur Wirkungsbewertung

---

**Krankenkassen- und Gesundheitsökonomie Masterclass**

Datum: 21.07.25

Name Verfasser:in: Dr. Paul Schmelzer

1 Einleitung .....	- 1 -
2. Prävention in der Gesetzlichen Krankenversicherung (GKV) .....	- 1 -
3. Digitale Technologien in der Prävention .....	- 2 -
3.1 Definition und Einordnung digitaler Präventionstechnologien .....	- 2 -
3.2 Anwendungsmöglichkeiten im Kontext des GKV-Präventionsauftrags .....	- 3 -
3.3 Abgrenzung zu DiGA und therapeutischen Anwendungen .....	- 3 -
4. Herausforderungen und Probleme bei der Umsetzung digitaler Präventionsangebote .....	- 4 -
5. Methodische Anforderungen und bestehende Evidenz zur Wirkungsmessung digitaler Präventionsangebote .....	- 6 -
5.1 Herausforderungen bei der Wirkungsmessung .....	- 6 -
5.1.1 Auswahlverzerrung und Selbstselektion .....	- 6 -
5.1.2 Heterogene Zielgruppen und Nutzungsintensität .....	- 7 -
5.1.3 Mangelnde Langfristigkeit und eingeschränkte Datenverfügbarkeit .....	- 7 -
5.2 Methodische Ansätze zur Wirkungsmessung .....	- 8 -
5.2.1 Wirkungslogik als Ausgangspunkt .....	- 8 -
5.2.2 Übersicht über Evaluationsmethoden .....	- 9 -
5.3 Relevante empirische Befunde zu DiGA .....	- 10 -
5.3.1 Meta-Analyse von Mäder et al. (2023) .....	- 11 -
5.3.2 Ergänzende Studienanalyse .....	- 13 -
6. Chancen, Grenzen und Perspektiven digitaler Prävention in der GKV .....	- 14 -
6.1 Einleitung .....	- 14 -
6.2 Kritische Bewertung der Evidenzlage .....	- 15 -
6.3 Kosten-Nutzen-Überlegungen .....	- 15 -
7 Empfehlungen für zukünftige Forschung .....	- 16 -
8 Fazit und Ausblick .....	- 17 -
Literaturverzeichnis: .....	- 19 -

## **1 Einleitung**

Die Digitalisierung verändert das deutsche Gesundheitswesen grundlegend – von der medizinischen Versorgung über die Verwaltung bis hin zur Prävention. Digitale Gesundheitsanwendungen (DiGA) nehmen dabei eine immer bedeutendere Rolle ein. Sie versprechen nicht nur eine zeitgemäße, niedrigschwellige Unterstützung von Patient:innen im Alltag, sondern gelten auch als Instrumente zur Effizienzsteigerung im Gesundheitssystem. Gerade in der Primärprävention bieten sie theoretisch das Potenzial, gesundheitsförderliches Verhalten zu fördern und langfristig Kosten im GKV-System zu senken.

Vor diesem Hintergrund stellt sich die Frage, ob und wie diese digitalen Interventionen tatsächlich einen nachweisbaren Mehrwert bieten – insbesondere im Hinblick auf ihre Wirksamkeit und Zielgruppenpassung. Denn eine gelungene Prävention erfordert nicht nur technische Innovation, sondern auch eine empirisch belastbare Evidenzbasis: Wer profitiert von welchen digitalen Angeboten – und unter welchen Bedingungen? Für gesetzliche Krankenkassen stellen sich damit nicht nur strategische, sondern auch wirtschaftliche Fragen: Lohnt sich der Einsatz solcher Technologien im Sinne der Versichertengemeinschaft? Oder droht eine ineffiziente Mittelverwendung ohne belastbare Wirksamkeitsnachweise?

Ziel dieser Arbeit ist es daher, Möglichkeiten und Grenzen digitaler Primärprävention in der GKV zu analysieren – mit besonderem Fokus auf die methodischen Anforderungen der Wirkungsmessung. Im Mittelpunkt steht die Frage, unter welchen Bedingungen digitale Präventionsangebote evidenzbasiert bewertet werden können. Auf Basis aktueller Studien und gesetzlicher Rahmenbedingungen sollen zentrale Herausforderungen aufgezeigt, relevante methodische Ansätze diskutiert und praxisrelevante Empfehlungen formuliert werden.

Der Aufbau der Arbeit gliedert sich wie folgt: Kapitel 2 skizziert die gesetzlichen und institutionellen Rahmenbedingungen der Primärprävention in der GKV. Kapitel 3 stellt digitale Technologien vor, die derzeit oder potenziell für präventive Zwecke genutzt werden. Kapitel 4 beleuchtet zentrale Herausforderungen bei der praktischen Umsetzung solcher Angebote. Kapitel 5 widmet sich den methodischen Grundlagen der Wirkungsmessung und analysiert exemplarisch ausgewählte nationale und internationale Studien. Kapitel 6 diskutiert Chancen und Grenzen digitaler Prävention aus GKV-Perspektive. Kapitel 7 fasst die wesentlichen Erkenntnisse zusammen und gibt einen Ausblick auf zukünftige Entwicklungen.

## **2. Prävention in der Gesetzlichen Krankenversicherung (GKV)**

Die gesetzliche Krankenversicherung (GKV) erfüllt nicht nur eine versichernde, sondern auch eine gesundheitspolitisch gestaltende Funktion. Ein zentrales Handlungsfeld ist dabei die Prävention. Wie Bodner (2016) betont, ist Prävention eine öffentliche Aufgabe mit hoher gesundheits- und

gesellschaftspolitischer Bedeutung. Ziel primärpräventiver Maßnahmen ist es, Krankheiten gar nicht erst entstehen zu lassen, um Folgekosten und Krankheitslast zu reduzieren.

Im Rahmen des deutschen Gesundheitssystems ist die primäre Prävention in § 20 SGB V gesetzlich geregelt (vgl. Busse, Blümel & Spranger, 2017). Dieser Paragraph verpflichtet die Krankenkassen dazu, Leistungen zur Gesundheitsförderung und Prävention bereitzustellen, insbesondere in den *Lebenswelten* ihrer Versicherten, wie Schulen, Betrieben oder Pflegeeinrichtungen. Dabei handelt es sich um Maßnahmen, die *vor dem Eintritt einer Erkrankung* ansetzen – etwa durch Bewegungsförderung, Ernährungsberatung oder Stressbewältigung.

Zur Abgrenzung: Sekundärprävention – also Maßnahmen zur Früherkennung von Krankheiten wie etwa Krebs-Screenings – ist im § 25 SGB V geregelt (vgl. Bodner, 2016). Die tertiäre Prävention wiederum zielt auf die Vermeidung von Rückfällen oder Komplikationen bei bestehenden Krankheiten ab. Dazu zählen etwa Rehabilitationsprogramme oder Schulungen bei chronischen Erkrankungen. Obwohl diese nicht in den Präventionsauftrag nach § 20 SGB V fällt, ist sie ein integraler Bestandteil des Gesundheitssystems.

Wie Wasem, Matusiewicz, Neumann und Noweski (2019) herausstellen, gewinnen Krankenkassen zunehmend an Bedeutung als Akteure einer *aktiven* Gesundheitspolitik. Dabei steht nicht nur die Einhaltung gesetzlicher Vorgaben im Vordergrund, sondern auch die Ausgestaltung innovativer, effizienter und zielgruppenspezifischer *Präventionsstrategien*.

Insgesamt ist Prävention ein zentraler Baustein der gesetzlichen Krankenversicherung – nicht nur im Hinblick auf die Verbesserung der Gesundheitschancen der Versicherten, sondern auch im Sinne der langfristigen Steuerung von Ausgaben im Gesundheitssystem. Die Digitalisierung eröffnet hierbei neue Möglichkeiten, den gesetzlichen Auftrag innovativ weiterzuentwickeln.

Die Digitalisierung eröffnet hierbei neue Möglichkeiten, den gesetzlichen Auftrag innovativ weiterzuentwickeln – insbesondere durch digitale Gesundheitsanwendungen, die zunehmend Eingang in präventive Strategien der GKV finden.

### **3. Digitale Technologien in der Prävention**

#### **3.1 Definition und Einordnung digitaler Präventionstechnologien**

Digitale Technologien haben in den letzten Jahren zunehmend Einzug in den Gesundheitssektor gehalten – auch im Bereich der Primärprävention. Darunter werden sämtliche digitalen Anwendungen verstanden, die dazu dienen, *gesundheitsförderndes Verhalten* zu unterstützen oder gesundheitliche Risiken zu verringern. Dazu zählen insbesondere mobile Gesundheits-Apps, digitale Coaching-Programme, Online-Kurse, Wearables sowie interaktive Plattformen zur Gesundheitsinformation und -kommunikation (WHO, 2021; Albrecht & Von Jan, 2019).

Diese Anwendungen zeichnen sich oft durch personalisierte Inhalte, Echtzeit-Feedback oder Gamification-Elemente aus und werden in der Regel über mobile Endgeräte wie Smartphones oder Tablets genutzt. Sie adressieren typische Risikofaktoren wie Bewegungsmangel, unausgewogene Ernährung, chronischen Stress oder Suchtmittelkonsum – also Bereiche, in denen präventives Handeln besonders wirkungsvoll sein kann (BZgA, 2022).

### **3.2 Anwendungsmöglichkeiten im Kontext des GKV-Präventionsauftrags**

Gemäß § 20 SGB V ist es Aufgabe der gesetzlichen Krankenkassen, Maßnahmen der Primärprävention in *Lebenswelten* ihrer Versicherten zu unterstützen. Digitale Technologien eröffnen hierfür eine Reihe von innovativen Möglichkeiten. Sie lassen sich zielgerichtet in zentrale Präventionsfelder wie Bewegung, Ernährung, Stressbewältigung und Suchtmittelvermeidung integrieren (GKV-Spitzenverband, 2022).

*Bewegungsförderung* kann beispielsweise durch Schrittzähler-Apps, personalisierte Trainingspläne oder Wearables erfolgen, die Bewegungsziele erfassen und Rückmeldung geben. Im Bereich *Ernährung* bieten digitale Ernährungstagebücher, Barcode-Scanner oder Rezept-Coaches niederschwellige Unterstützung im Alltag. *Stressbewältigung* wird etwa durch Achtsamkeits-Apps, geführte Meditationen oder Biofeedback-Systeme adressiert, während *Suchtmittelvermeidung* zunehmend durch digitale Begleitprogramme – etwa zur Tabakentwöhnung – ergänzt wird (Albrecht & Von Jan, 2019).

Diese Angebote ermöglichen es Krankenkassen, gesundheitsförderndes Verhalten alltagsnah, flexibel und kosteneffizient zu fördern. Gleichzeitig erreichen sie insbesondere technikaffine oder jüngere Zielgruppen, bei denen klassische Gruppenprogramme häufig auf geringere Resonanz stoßen (BZgA, 2022). Gleichzeitig stellt sich die Frage, wie sich diese Anwendungen rechtlich und funktional von den Digitalen Gesundheitsanwendungen (DiGA) abgrenzen lassen.

### **3.3 Abgrenzung zu DiGA und therapeutischen Anwendungen**

Wesentlich für die Einordnung digitaler Gesundheitsanwendungen ist die *rechtliche* und *funktionale* Abgrenzung zwischen *präventiven Angeboten nach § 20 SGB V* und den sogenannten *Digitalen Gesundheitsanwendungen* (DiGA), die gemäß § 33a SGB V bei bestimmten Indikationen verordnet und erstattet werden können. DiGA sind digitale Medizinprodukte niedriger Risikoklassen (I oder IIa), die beim Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) gelistet werden und deren *positiver Versorgungseffekt* wissenschaftlich belegt sein muss (BfArM, 2023). Im Unterschied zu Medikamenten ist für DiGA kein Nachweis einer kausalen Wirksamkeit erforderlich – obwohl dieser dem gesetzlichen

Anspruch implizit zugrunde liegt. Dadurch bleibt der tatsächliche Kausalitätsnachweis häufig aus, obwohl er dem gesetzlichen Anspruch implizit zugrunde liegt.<sup>1</sup>

Anders als reine Präventions-Apps richten sich DiGA in der Regel an Menschen mit bestehenden gesundheitlichen Einschränkungen oder erhöhtem Risiko und dienen medizinischen Zwecken wie Diagnostik, Therapie oder Nachsorge. Beispiele hierfür sind Anwendungen zur Unterstützung bei Depressionen, Migräne oder Tinnitus. Diese Apps werden häufig in ärztlich geleitete Behandlung integriert und können bei entsprechender Indikation verschrieben werden (BMG, 2021).

Obwohl viele DiGA auch verhaltenspräventive Komponenten enthalten, gelten sie rechtlich nicht als Primärprävention im Sinne des § 20 SGB V, da sie auf behandlungsbedürftige Zustände oder Risikokonstellationen abzielen und ein medizinisches Ziel verfolgen. Apps, die ausschließlich präventive Zwecke bei gesunden Personen verfolgen – etwa zur Bewegungsförderung oder Stressbewältigung – gelten nicht als DiGA und fallen stattdessen unter den § 20 SGB V. Ihre Finanzierung erfolgt – sofern sie bestimmte Qualitätskriterien erfüllen – über das Präventionsbudget der Krankenkassen. Der Leitfaden Prävention des GKV-Spitzenverbands gibt hierfür die relevanten Rahmenbedingungen vor (GKV-Spitzenverband, 2022).

Digitale Gesundheitsanwendungen (DiGA) im Sinne der §§ 33a und 139e SGB V lassen sich inhaltlich in verschiedene Anwendungsbereiche gliedern. Ein großer Teil entfällt auf *psychische* Erkrankungen wie Depression, Angststörungen oder Stressbewältigung, beispielsweise durch kognitive Verhaltenstherapie per App. Weitere Schwerpunkte bilden *muskuloskelettale* Erkrankungen (z. B. Rückenschmerzen), *Stoffwechsel-* und *Ernährungskrankheiten* (z. B. Adipositas, Typ-2-Diabetes) sowie neurologische und internistische Indikationen wie Tinnitus oder Herzinsuffizienz. In der Regel richten sich DiGA an *chronisch Erkrankte* und zielen auf die Förderung der Selbstwirksamkeit, die Unterstützung der Therapieadhärenz oder die Symptomreduktion. Voraussetzung für ihre Aufnahme ins DiGA-Verzeichnis ist der Nachweis eines positiven Versorgungseffekts – entweder im Sinne eines medizinischen Nutzens oder einer patientenrelevanten Struktur- und Verfahrensverbesserung.

#### **4. Herausforderungen und Probleme bei der Umsetzung digitaler Präventionsangebote**

Trotz ihres vielfach betonten Potenzials stehen digitale Präventionsangebote in der gesetzlichen Krankenversicherung (GKV) vor erheblichen Umsetzungsproblemen. Diese betreffen sowohl die technische und organisatorische Integration als auch Fragen der Evidenzbasis, Akzeptanz und Standardisierung.

---

<sup>1</sup> Die Einteilung in Risikoklassen erfolgt nach der europäischen Medizinprodukteverordnung (MDR). Klasse I umfasst Produkte mit sehr geringem Risiko (z. B. Tagebuch- oder Dokumentations-Apps), während Klasse IIa bereits Anwendungen mit leicht erhöhtem Risiko umfasst (z. B. Apps zur Unterstützung bei psychischen Erkrankungen oder chronischen Schmerzen). Für DiGA sind nur Produkte dieser beiden niedrigeren Risikoklassen zugelassen – Produkte höherer Risikoklassen (z. B. zur Steuerung von Implantaten) sind ausgeschlossen.

Ein zentrales Problem stellt die *geringe Nutzung entsprechender Angebote* dar. Studien zeigen, dass viele digitale Präventionsanwendungen lediglich eine kleine, oft ohnehin gesundheitsaffine Zielgruppe erreichen, während die breite Versichertenbasis kaum erreicht wird. So verweisen Groene und Schneck (2023) in ihrer Analyse zur Verordnungspraxis Digitaler Gesundheitsanwendungen (DiGA) darauf, dass trotz wachsender Zahl zugelassener Produkte die tatsächliche Verschreibung und Nutzung weit hinter den Erwartungen zurückbleiben. Zwar beziehen sich diese Ergebnisse primär auf DiGA, doch lassen sich ähnliche Muster auch für nicht-verordnungsfähige Präventions-Apps feststellen.

Ein weiterer kritischer Punkt ist die **unzureichende wissenschaftliche Fundierung** vieler Angebote. Dittrich et al. (2023) zeigen in einem systematischen Review, dass zahlreiche Gesundheits-Apps methodische Schwächen aufweisen – insbesondere im Hinblick auf Studiendesigns, Endpunktdefinitionen und Evaluationszeiträume. Dies erschwert eine belastbare Nutzenbewertung und führt zu Unsicherheiten bei den Entscheidungsträger:innen in der GKV.

Auch die **fehlende Standardisierung und Qualitätssicherung** stellt ein Problem dar. Während für DiGA ein strukturierter Bewertungsprozess durch das Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) existiert, fehlt ein vergleichbares Verfahren für Präventionsanwendungen, die nicht unter die DiGA-Regelung fallen. Krankenkassen müssen hier häufig auf Eigenbewertungen oder freiwillige Zertifizierungen zurückgreifen, deren Aussagekraft stark variieren kann.

Hinzu kommt das Problem **sozialer und digitaler Ungleichheiten**. Ältere Menschen, sozial benachteiligte Gruppen oder Personen mit geringer digitaler Gesundheitskompetenz können durch technische Anforderungen ausgeschlossen werden. Diese „digitale Spaltung“ gefährdet nicht nur die Reichweite präventiver Maßnahmen, sondern birgt das Risiko, bestehende gesundheitliche Ungleichheiten zu verstärken (vgl. JMIR Human Factors, 2023).

Nicht zuletzt mangelt es vielen Anwendungen an **Integration in den Versorgungskontext**. Häufig werden digitale Tools unabhängig von bestehenden Versorgungsstrukturen entwickelt und lassen sich nur schwer in etablierte Prozesse einbinden. Das betrifft etwa Abrechnungsmodalitäten, Schnittstellen zur Kommunikation mit Leistungserbringern oder die technische Kompatibilität mit IT-Systemen der Kassen. Gerade kleinere Krankenkassen sehen sich hier mit erheblichen Hürden konfrontiert (vgl. BMG, 2021).

Zusammenfassend lässt sich festhalten: Digitale Präventionsangebote stoßen in der GKV grundsätzlich auf Offenheit, ihre Umsetzung wird jedoch durch eine Vielzahl struktureller, methodischer und sozialer Barrieren erschwert. Diese Herausforderungen müssen systematisch adressiert werden, um das Potenzial digitaler Prävention wirksam in die Regelversorgung zu überführen.

## 5. Methodische Anforderungen und bestehende Evidenz zur Wirkungsmessung digitaler Präventionsangebote

Wie in den vorangegangenen Kapiteln bereits dargelegt, ist die Evidenzlage zu digitalen Präventionsangeboten vielfach unzureichend. Eine methodisch fundierte Wirkungsmessung ist daher unerlässlich, um belastbare Aussagen zur tatsächlichen Wirksamkeit und Wirtschaftlichkeit solcher Interventionen treffen zu können.

Im Zentrum jeder Evaluation steht die Frage, ob eine beobachtete Veränderung kausal auf eine digitale Gesundheitsanwendung (z. B. eine DiGA) zurückzuführen ist – oder lediglich mit ihrer Nutzung korreliert. In der ökonometrischen Wirkungsmessung bezeichnet der Begriff „Wirkung“ (engl. *treatment effect*) stets einen ursächlichen Zusammenhang: Die Intervention führt nachweislich zu einem veränderten Ergebnis, unter sonst gleichen Bedingungen. Eine Korrelation hingegen beschreibt lediglich das gleichzeitige Auftreten zweier Merkmale, ohne dass daraus automatisch eine Ursache-Wirkung-Beziehung abgeleitet werden kann (vgl. Angrist & Pischke, 2009; Imbens & Rubin, 2015; Hernán & Robins, 2020).

Diese Unterscheidung ist für digitale Präventionsangebote besonders relevant: So kann etwa die Nutzung einer Gesundheits-App mit einer Verbesserung des Gesundheitsverhaltens einhergehen – doch ob diese Verbesserung tatsächlich auf die App zurückzuführen ist, lässt sich nur mit geeigneten methodischen Ansätzen klären. Genau hierin liegen die zentralen Herausforderungen empirischer Evaluationsforschung im Gesundheitswesen.

Im Folgenden werden zunächst typische Verzerrungen und methodische Fallstricke beschrieben. Anschließend werden verschiedene Studiendesigns zur Kausalanalyse vorgestellt und bestehende empirische Befunde eingeordnet.

### 5.1 Herausforderungen bei der Wirkungsmessung

#### 5.1.1 Auswahlverzerrung und Selbstselektion

Ein zentrales methodisches Problem bei der Evaluation digitaler Präventionsangebote besteht in **systematischen Auswahlverzerrungen**. Nutzer:innen solcher Anwendungen sind häufig technikaffin, gesundheitsbewusst und überdurchschnittlich gebildet – Merkmale, die generell mit einem gesünderen Lebensstil und besseren Gesundheitsergebnissen einhergehen. Die beobachtete Wirksamkeit digitaler Interventionen könnte daher nicht primär auf die Intervention selbst, sondern auf Eigenschaften der Teilnehmenden zurückzuführen sein. Diese **Selbstselektion** erschwert kausale Aussagen über den tatsächlichen Interventionseffekt.

Randomisierte kontrollierte Studien (RCTs) gelten als Goldstandard, da sie systematische Unterschiede zwischen Interventions- und Kontrollgruppen minimieren. In der Praxis greifen jedoch viele DiGA-



Studien auf nicht-randomisierte oder quasi-experimentelle Verfahren zurück – etwa **Matching-Methoden** oder **Instrumentvariablenansätze**, um beobachtete und unbeobachtete Heterogenität statistisch zu kontrollieren (vgl. Kapitel 5.3).

Ein Beispiel für eine kritische Auseinandersetzung mit diesem Problem bietet die Analyse von Dittrich et al. (2023), die methodische Schwächen zahlreicher Studien im DiGA-Kontext dokumentieren. Auch bei Angeboten wie ViViRA zur Rückenschmerztherapie zeigt sich, dass Nutzer:innen selektiv und eigenmotiviert mit der Anwendung interagieren – was die Vergleichbarkeit zusätzlich erschwert.

### 5.1.2 Heterogene Zielgruppen und Nutzungsintensität

Digitale Präventionsangebote adressieren sehr unterschiedliche Zielgruppen – von chronisch Kranken über psychisch Belastete bis hin zu gesunden Personen mit Präventionsinteresse. Diese Gruppen unterscheiden sich erheblich in Alter, Gesundheitszustand, digitaler Kompetenz und Motivation. Hinzu kommt eine große **Varianz in der Nutzungsintensität**: Manche verwenden die App regelmäßig, andere brechen sie frühzeitig ab.

Um der Heterogenität gerecht zu werden, sind **Subgruppenanalysen** oder **interaktive Modelle** erforderlich – etwa zur Untersuchung von Wechselwirkungen zwischen der Intervention und Merkmalen wie Alter, Geschlecht oder Gesundheitskompetenz. Damit lassen sich **bedingte Behandlungseffekte** (Conditional Average Treatment Effects, CATE) schätzen – also die Frage beantworten, für wen die Intervention besonders wirksam ist.

Viele Studien berichten lediglich den **Average Treatment Effect on the Treated (ATT)**, also den mittleren Effekt bei jenen, die die Intervention genutzt haben. Dieser ist jedoch nicht auf die Gesamtbevölkerung übertragbar, da Nutzer:innen selektiv sind. Der **Average Treatment Effect (ATE)** beschreibt dagegen den durchschnittlichen Effekt in der gesamten Zielpopulation. Wird dieser ohne Berücksichtigung von Heterogenität interpretiert, kann das zu Fehlschlüssen führen – etwa wenn die Intervention bei bestimmten Gruppen nicht wirkt oder besonders stark wirkt, was sich im Durchschnitt verwässert.

Die Studie in *Nature Digital Health* (2024) zu einer DiGA gegen Depressionen weist eine hohe Effektstärke (Cohen's  $d = 1,63$ ) aus, ohne differenziert auf die Übertragbarkeit der Ergebnisse auf verschiedene Subgruppen einzugehen. Auch Mäder et al. (2023) betonen, dass Effekte digitaler Interventionen stark vom Nutzungskontext und der Zielgruppe abhängig sind.

### 5.1.3 Mangelnde Langfristigkeit und eingeschränkte Datenverfügbarkeit

Ein weiteres Problem besteht in der **kurzen Beobachtungsdauer** vieler Studien. Effekte werden meist über Zeiträume von 6 bis 12 Wochen gemessen – zu kurz, um Rückschlüsse auf nachhaltige Verhaltensänderungen zuzulassen. Gerade bei präventiven Interventionen ist aber entscheidend, ob sie langfristig wirksam sind.

Zudem ist die Nutzungsdauer selbst **nicht exogen**, sondern hängt mit Merkmalen der Nutzer:innen zusammen. Wer eine DiGA lange nutzt, ist meist motivierter, gesünder oder technikaffiner – was zu einem **Attrition Bias** führt: Werden nur die Verläufe jener analysiert, die „durchhalten“, können Interventionseffekte überschätzt werden.

Erschwerend kommt hinzu, dass viele Studien **herstellerfinanziert** sind und sich auf firmeneigene Daten stützen. Der Zugang zu **unabhängigen Routinedaten** – etwa von Krankenkassen – ist selten. Dadurch sind Validität und Replizierbarkeit der Ergebnisse eingeschränkt.

Eine stärkere **Einbindung der Krankenkassen** in die Evaluation – z. B. über gemeinsame Studien, Zugänge zu Sekundärdaten oder Registerlösungen mit Follow-up-Erhebungen – könnte helfen, die Evidenzlage deutlich zu verbessern. Nur so lassen sich reale Versorgungseffekte und fundierte Aussagen zum Kosten-Nutzen-Verhältnis treffen.

## 5.2 Methodische Ansätze zur Wirkungsmessung <sup>2</sup>

Während die bisherigen Abschnitte zentrale Herausforderungen bei der Evaluation digitaler Gesundheitsanwendungen (DiGA) identifiziert haben, richtet sich der Blick nun auf mögliche methodische Lösungen. Dabei stellt sich die Frage, was unter idealen Bedingungen notwendig wäre, um kausale Effekte verlässlich zu messen – und inwieweit sich diese Anforderungen im Praxisrahmen der gesetzlichen Krankenversicherung (GKV) realisieren lassen.

### 5.2.1 Wirkungslogik als Ausgangspunkt

Im Zentrum jeder Wirkungsmessung steht die Frage, ob eine beobachtete Veränderung tatsächlich auf eine Intervention – im vorliegenden Fall eine DiGA – zurückzuführen ist oder lediglich eine statistische Korrelation vorliegt. In der Wirkungsevaluation wird unter einem *treatment effect* (Behandlungseffekt) streng genommen ein kausaler Zusammenhang verstanden, nicht bloß eine Assoziation (vgl. Angrist & Pischke, 2009; Imbens & Rubin, 2015; Hernán & Robins, 2020).

Um eine belastbare Wirkungsmessung zu ermöglichen, ist es unerlässlich, die Intervention in eine plausible Wirkungslogik einzubetten – also in ein theoretisch fundiertes Modell, das nachvollziehbar macht, **was** beeinflusst werden soll, **durch welche Maßnahme** (Treatment) und **bei welcher Zielgruppe**. Eine DiGA zur Rückenschmerztherapie etwa zielt auf eine Reduktion der Schmerzintensität (Ziel) durch regelmäßige physiotherapeutische Übungen via App (Behandlung) bei Erwachsenen mit chronischen Rückenschmerzen (Zielgruppe).

---

<sup>2</sup> Ergänzend zu den in diesem Kapitel behandelten quantitativen Evaluationsmethoden werden in der Praxis häufig auch qualitative Ansätze wie Interviews, Fokusgruppen oder digitale Nutzerbewertungen eingesetzt. Diese liefern wichtige Einblicke in Nutzungserfahrungen, Barrieren und subjektiv empfundene Wirkung. Solche Verfahren eignen sich besonders in frühen Phasen der Entwicklung oder zur Hypothesengenerierung. Für belastbare Aussagen zur Wirksamkeit im kausalen Sinne sind sie jedoch nicht ausreichend. Da diese Arbeit den Fokus auf methodisch fundierte Kausalanalysen legt, werden qualitative Methoden im Folgenden nicht vertiefend behandelt.

## 5.2.2 Übersicht über Evaluationsmethoden

Zur Abschätzung kausaler Effekte digitaler Präventionsangebote stehen verschiedene methodische Ansätze zur Verfügung. Im Mittelpunkt steht stets die Frage, wie valide und belastbar sich der Beitrag der Intervention zur beobachteten Veränderung bestimmen lässt. Die Wahl der Methode hängt dabei nicht nur von wissenschaftlichen Standards, sondern auch von praktischen und institutionellen Rahmenbedingungen – insbesondere im Kontext der GKV – ab. Im Folgenden werden die wichtigsten Verfahren vorgestellt, jeweils mit Blick auf ihre methodische Fundierung, praktische Anwendbarkeit und ihre Grenzen im Versorgungskontext.

**Randomisierte kontrollierte Studien (RCTs)** gelten als Goldstandard der Wirkungsmessung. Durch die zufällige Zuweisung von Teilnehmenden zu Interventions- und Kontrollgruppen lassen sich systematische Verzerrungen minimieren, was eine hohe interne Validität und klare kausale Aussagen ermöglicht.

Methodisch sind RCTs oft weniger komplex als angenommen: Im Gegensatz zu vielen Beobachtungsdesigns benötigen sie keine aufwendigen Kontrollvariablenmodelle. Bei sauberer Randomisierung reichen häufig einfache statistische Verfahren – etwa Mittelwertvergleiche oder *t*-Tests –, um belastbare Aussagen zum Behandlungseffekt zu treffen. Auch der Variablenbedarf bleibt überschaubar, da primär die Gruppenzuordnung und das Outcome entscheidend sind.

In der praktischen Umsetzung erweisen sich RCTs jedoch als aufwendig: Sie erfordern detaillierte Studienprotokolle, ethische Genehmigungen, hohe Fallzahlen (insbesondere bei heterogenen Gruppen oder geplanter Subgruppenanalyse) sowie eine strukturierte Datenerhebung. Im GKV-Kontext sind sie daher meist nur im Rahmen klar abgegrenzter Interventionen oder Pilotprojekte umsetzbar – häufig in Kooperation mit externen Forschungspartnern oder unter aktiver Beteiligung der Krankenkassen.

Hinzu kommen ethische Herausforderungen: Personen in der Kontrollgruppe wird ein potenziell hilfreicher Zugang zur DiGA zunächst verwehrt – obwohl sie formal infrage kämen. Dies erfordert entweder sorgfältige Abwägung oder die Nutzung alternativer Designs, etwa durch spätere Nachreichung der Intervention (*crossover design*).

Auch bei RCTs bleibt **selektiver Studienabbruch (Attrition)** ein Problem: Personen, die vorzeitig ausscheiden, unterscheiden sich oft systematisch von jenen, die die Studie abschließen – was die Aussagekraft verzerren kann.

**Quasi-experimentelle Designs** bieten eine pragmatische Alternative, wenn randomisierte Studien nicht realisierbar sind. Zu den etablierten Verfahren zählen:

- **Propensity Score Matching (PSM):** Vergleichbare Gruppen werden auf Basis beobachtbarer Merkmale gebildet.

- **Difference-in-Differences (DiD):** Veränderungen im Zeitverlauf zwischen behandelten und unbehandelten Gruppen werden analysiert.
- **Instrumentvariablenansätze (IV):** Exogene Einflussfaktoren dienen als Instrumente zur Abschätzung unbeobachteter Selektionseffekte.

Diese Methoden setzen eine gute Datenlage voraus – etwa Routinedaten der GKV – und müssen sorgfältig implementiert werden, um Verzerrungen zu minimieren. Ihr Vorteil liegt in der Nähe zur Versorgungspraxis und der Möglichkeit, große Stichproben einzubeziehen. Gleichwohl bleiben sie anfällig für Verzerrungen durch **unbeobachtete Heterogenität** – vor allem, wenn relevante Einflussgrößen wie Motivation, Technikaffinität oder Gesundheitsverhalten nicht im Datensatz enthalten sind.

Viele dieser Verfahren entfalten ihr volles Potenzial erst bei **längsschnittlich erhobenen Daten**. Paneldaten erlauben es, individuelle Veränderungen im Zeitverlauf zu modellieren und zeitinvariante Störgrößen (*Fixed Effects*) auszuschalten. DiD kann als Spezialfall eines Fixed-Effects-Modells verstanden werden, das gezielt die Wirkung einer spezifischen Intervention isoliert. Auch **Ereignisanalysen** wie *interrupted time series* oder *event study*-Designs bauen auf Panelstrukturen auf und identifizieren Effekte vor und nach dem Interventionszeitpunkt. Bei aggregierten oder unvollständigen Daten stoßen diese Methoden jedoch an ihre Grenzen.

**Beobachtungsstudien auf Basis von Sekundärdaten** – etwa aus Abrechnungs- oder Versorgungsdaten der GKV – bieten Potenzial zur Wirkungsmessung unter realen Versorgungsbedingungen. Ihre Stärke liegt in der **hohen externen Validität**, da sie große Subpopulationen über längere Zeiträume in Alltagskontexten abbilden. Ihre Schwäche liegt in der **fehlenden internen Validität**: Vergleichsgruppen entstehen nicht zufällig, sondern unterliegen Selektionseffekten. Zudem fehlen häufig wichtige Kovariaten wie digitale Kompetenz, Motivation oder psychische Belastung.

Solche Studien ermöglichen dennoch eine **unabhängige Einschätzung der Studienbedingungen**: Wie realistisch sind Nutzungsraten, Abbruchquoten oder Wirkungsmuster im Vergleich zur Versorgungspraxis? Lassen sich die untersuchten DiGA an reale Zielgruppen und Versorgungssettings anschließen? Auch wenn die Kausalaussagen begrenzt bleiben, leisten Beobachtungsstudien einen wichtigen Beitrag zur Einordnung von Studienergebnissen in den Versorgungskontext.

### 5.3 Relevante empirische Befunde zu DiGA

Digitale Gesundheitsanwendungen (DiGA) werden zunehmend durch wissenschaftliche Studien gestützt. Dabei stellt sich die Frage, wie belastbar die bislang vorliegende Evidenz ist – insbesondere im Hinblick auf Studiendesign, Stichprobengröße, Beobachtungszeitraum und externe Validität. Im Folgenden werden zentrale Studien exemplarisch vorgestellt und hinsichtlich ihrer methodischen Aussagekraft eingeordnet.

Aktuell sind 71 DiGAs gelistet (Stand: 15.07.2025), darunter 44 dauerhaft (BfArM, 2025) (Quelle: Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). (2025). DiGA-Verzeichnis. Abgerufen von <https://diga.bfarm.de/de>).

Im Rahmen dieser Arbeit wurden ausschließlich DiGA-Studien berücksichtigt, die den Kriterien randomisierter kontrollierter Studien (RCTs) entsprechen und ausreichende Angaben zu Methodik, Daten, Kontrollgruppen und Analyse enthalten. Es wurde darauf geachtet, dass die Studien ordnungsgemäß beschrieben sind, um eine belastbare Grundlage für die Analyse der Ergebnisse zu gewährleisten. Besonders wichtig war, dass relevante Details zu Drop-out-Raten und Beobachtungszeiträumen klar dokumentiert sind, um eine fundierte Bewertung der Studienqualität und der Resultate zu ermöglichen. Diese Vorgehensweise wurde gewählt, um die Verlässlichkeit und Vergleichbarkeit der ausgewerteten Studien sicherzustellen. Es ist jedoch zu beachten, dass die meisten der ausgewählten Studien potenzielle Interessenkonflikte aufweisen, was die Interpretation der Ergebnisse beeinflussen kann.

Von 59 Studien (im Jahr 2023) identifizierten Mäder et al. 13 Studien als belastbar. Mäder et al. (2023). Nach einer Recherche habe ich fünf weitere Studien als belastbar identifiziert und deren Ergebnisse kurz zusammengefasst. Die restlichen Studien sind entweder nur teilweise oder gar nicht öffentlich zugänglich.

### **5.3.1 Meta-Analyse von Mäder et al. (2023)**

Mäder et al. (2023) analysieren alle zum Untersuchungszeitpunkt dauerhaft im DiGA-Verzeichnis gelisteten Anwendungen (n = 13). Die Mehrzahl dieser Anwendungen ist dem Bereich psychische Gesundheit zuzuordnen, und alle Hersteller stützten ihren Evidenznachweis auf randomisierte kontrollierte Studien (RCTs). Bei den meisten DiGA lag eine dokumentierte Verbesserung des Gesundheitszustands im primären Endpunkt vor, wenngleich in Einzelfällen – etwa bei Fokus auf Lebensqualität – kein signifikanter Effekt nachweisbar war.

Die Autor:innen identifizieren mehrere zentrale methodische Schwächen, die auch aus evaluativer Perspektive relevant erscheinen:

#### **(1) Subjektive Endpunkte und fehlender Placebovergleich:**

Wie Mäder et al. feststellen, bleibt häufig unklar, was genau unter einer „Verbesserung des Gesundheitszustands“ zu verstehen ist. Viele Studien verwenden subjektive Endpunkte wie Selbstauskünfte zu Stimmung oder Schmerzempfinden, ohne systematisch Placeboeffekte auszuschließen. Ergänzend lässt sich kritisch anmerken, dass bei psychischen oder verhaltensorientierten Interventionen aktive Placebokontrollgruppen – z. B. neutrale digitale Inhalte – notwendig wären, um Erwartungseffekte auszuschließen. Gerade bei subjektiven Zielgrößen kann die Erwartungshaltung der Nutzer:innen den gemessenen Effekt erheblich verzerren.

## **(2) Definition und Aussagekraft des primären Endpunkts:**

Laut Mäder et al. beruht die Wirksamkeit meist auf einem vorab definierten primären Endpunkt. Wird dieser zum Messzeitpunkt signifikant verbessert, gilt der medizinische Nutzen formal als erbracht. Darüber hinaus erscheint es aus evaluativer Sicht problematisch, dass es sich meist um punktuelle Zustandsveränderungen handelt – nicht jedoch um langfristige oder nachhaltige Verbesserungen. Rückfälle oder dynamische Verläufe bleiben in den meisten Fällen unbeachtet.

## **(3) Hohe Drop-out-Raten und Selektionsprobleme:**

Drop-out-Raten stellen ein zentrales methodisches Problem in der Evaluation digitaler Gesundheitsanwendungen dar – sowohl bei DiGA als auch bei anderen digitalen Interventionen. Wie Mäder et al. (2023) zu Recht betonen, „lässt eine hohe Drop-out-Rate [zwischen 7 % und 30 %] nicht auf den Erfolg der Anwendung schließen und ist nur eingeschränkt hinnehmbar“ (S. 12; eigene Übersetzung). Ergänzend lässt sich argumentieren, dass insbesondere Patient:innen ohne wahrgenommene Verbesserung zum Abbruch neigen. Solange nicht belegt ist, dass diese Ausfälle zufällig erfolgen, besteht ein erhebliches Risiko systematischer Verzerrungen – etwa in Form eines *downward bias*, wenn vor allem Nicht-Ansprechende abbrechen, oder eines *upward bias*, wenn Ausfälle ignoriert und nur erfolgreiche Verläufe berücksichtigt werden. Eine differenzierte Analyse der Abbruchgründe wäre daher methodisch zwingend erforderlich, um Verzerrungen zu erkennen und gegebenenfalls durch Gewichtungsverfahren zu korrigieren.

Auch unabhängig von DiGA stellen Drop-out-Raten ein zentrales methodisches Problem in der Evaluation digitaler Gesundheitsanwendungen dar. Besonders in Bereichen wie psychischen Erkrankungen und onkologischen Indikationen zeigen viele Studien überdurchschnittlich hohe Abbruchquoten. Diese führen regelmäßig zu einer systematischen Überschätzung der Wirksamkeit, da schwerer belastete, weniger technikaffine oder kränkere Patient:innen häufiger abbrechen, während motivierte, gesündere Personen in der Stichprobe verbleiben und die Ergebnislage dominieren (Karyotaki et al., 2021; Meyerowitz-Katz et al., 2020; Hernández-Rodríguez et al., 2022). Bei Apps für Krebspatient:innen reichen die dokumentierten Drop-out-Raten von 16 % bis 37 % – meist infolge gesundheitlicher Verschlechterung oder fehlender Motivation (CanRelax2-Studie; *PLOS ONE*, 2023). In der psychischen Gesundheitsversorgung werden vereinzelt sogar Raten von bis zu 48 % berichtet, wobei selektive Ausfälle nachweislich zu Verzerrungen führen (Nwosu et al., 2022). Demgegenüber erscheinen digitale Anwendungen bei muskuloskelettalen Beschwerden wie Rückenschmerzen oder Knieproblemen weniger verzerrungsanfällig, da die Drop-out-Raten hier meist unter 20 % liegen (Mannion et al., 2024; *JMIR mHealth*, 2024).

## **(4) Externe Validität und fehlende Subgruppenanalysen:**

Mäder et al. kritisieren zudem die begrenzte externe Validität: Die Studien berichten fast ausschließlich den Average Treatment Effect on the Treated (ATT), beruhend auf stark selektiven, oft digital affinen Stichproben. Aus evaluativer Perspektive bleibt ergänzend festzuhalten, dass damit keine belastbaren Aussagen über die Wirksamkeit in der breiten Versichertenpopulation (ATE) getroffen werden können. Auch Subgruppenanalysen – etwa nach Alter, Geschlecht, Bildungsstatus oder Vorerkrankungen – fehlen weitgehend, was eine differenzierte Wirkungseinschätzung erschwert.

#### **(5) Replizierbarkeit und Transparenz:**

Schließlich zeigen Mäder et al., dass teils ergänzende Recherchen in Studienregistern notwendig waren, um die Ergebnisse nachvollziehbar einordnen zu können. In mindestens einem Fall wichen die Angaben im DiGA-Verzeichnis von jenen im Studienbericht ab – ein Hinweis auf mögliche Transparenzdefizite.

### **5.3.2 Ergänzende Studienanalyse**

Nachdem die Mäder-Studie (2023) und ihre methodischen Schwächen diskutiert wurden, sollen nun die Ergebnisse weiterer relevanter DiGA-Studien einbezogen werden, die ich zusätzlich recherchiert habe:

#### **SinCephalea-Studie (Grosse et al., 2024)**

Untersuchte die Wirksamkeit einer Migräneprophylaxe-App. Die monatlichen Migränetage konnten in der Interventionsgruppe signifikant stärker reduziert werden als in der Kontrollgruppe. Zu beachten ist der mögliche Interessenkonflikt, da die Perfood GmbH kommerziell beteiligt war.

#### **Mawendo-Studie (Mayer et al., 2024)**

Evaluierte Mawendo bei patellofemorale Schmerzen in einer prospektiven RCT. Die App zeigte eine signifikante Überlegenheit gegenüber der Standardtherapie, insbesondere hinsichtlich Funktionsverbesserung und Schmerzreduktion. Methodische Einschränkungen bestanden durch hohe Drop-out-Rate und fehlende Verblindung. Nur 2,4% der Datensätze waren unvollständig = drop outs, und die **Imputation fehlender Werte** zeigt, dass der Einfluss auf die Ergebnisse minimal war.

#### **Vantis-Studie (Rödel & Schröder, 2023)**

Untersuchte die Wirkung einer DiGA bei Patienten mit KHK und Bluthochdruck. Die Intervention führte zu einer systolischen Blutdrucksenkung von 15,5 mmHg gegenüber 6,0 mmHg in der Kontrollgruppe – vergleichbar mit medikamentöser Kombinationstherapie. Auch hier gibt es mögliche Interessenkonflikte durch kommerzielle Sponsoren. Warden keine dropouts berichtet.

#### **Oviva Direkt – Adipositas (Gemasi et al., 2024)**

Evaluierte Oviva Direkt in einer 12-wöchigen RCT bei Adipositas. Es zeigte sich eine signifikante Gewichtsreduktion von etwa 3 % im Vergleich zur Kontrollgruppe bei einer Drop-out-Rate von 20 %.

Obwohl ein post-hoc Power-Test eine hohe statistische Power (99,9 %) ergab, wurden Langzeiteffekte nur innerhalb der Interventionsgruppe erfasst, ohne Kontrolle über 24 Wochen hinweg.

## **Fazit**

Die bisher vorliegende Evidenz zu DiGA basiert überwiegend auf herstellerfinanzierten randomisierten kontrollierten Studien (RCTs) mit kurzen Laufzeiten und selektiven Stichproben, was die Externe Validität und die Langfristigkeit der Ergebnisse einschränkt. Während diese Studien wichtige Hinweise zur Wirksamkeit der DiGA geben, bleibt ihre Aussagekraft hinsichtlich der Nachhaltigkeit, der Zielgruppendifferenzierung und der Übertragbarkeit auf die allgemeine Versorgung begrenzt. Insbesondere die hohe Drop-out-Rate und die finanzielle Unterstützung durch kommerzielle Akteure werfen Fragen zu möglichen Interessenkonflikten auf, die die Unabhängigkeit und Objektivität der Ergebnisse beeinträchtigen können. Darüber hinaus fokussieren sich viele Studien auf den Average Treatment Effect on the Treated (ATT), wodurch keine belastbaren Aussagen über die Wirksamkeit in der breiten Versichertenpopulation (Average Treatment Effect, ATE) getroffen werden können. Die Heterogenität der Stichproben und das Fehlen von Subgruppenanalysen erschwert eine differenzierte Beurteilung der Wirkung auf verschiedene Patientengruppen. Zukünftige Studien sollten daher verstärkt Real-World-Daten einbeziehen, längere Beobachtungszeiträume erfassen und unabhängige Evaluationen durch Institutionen wie Krankenkassen oder Registerstrukturen anstreben, um eine objektivere und umfassendere Bewertung der DiGA zu ermöglichen. Dies würde nicht nur die Präsentierbarkeit und Anwendbarkeit der Ergebnisse verbessern, sondern auch eine verlässlichere Grundlage für die Implementierung von DiGA in die Regelversorgung bieten.

## **6. Chancen, Grenzen und Perspektiven digitaler Prävention in der GKV**

### **6.1 Einleitung**

Digitale Präventionsangebote versprechen, gesundheitsförderndes Verhalten auf innovative Weise zu unterstützen – ortsunabhängig, niedrighschwellig und potenziell kosteneffizient. Durch die breite Nutzung digitaler Technologien könnten nicht nur individuelle Gesundheitskompetenzen gestärkt, sondern auch strukturelle Engpässe in der Gesundheitsversorgung abgefedert werden. Gleichzeitig stellen digitale Interventionen hohe Anforderungen an wissenschaftliche Evidenz, Datenschutz, Nutzerakzeptanz und regulatorische Steuerung.

Gerade im GKV-System, das auf Solidarität und Effektivität ausgelegt ist, darf der Einsatz digitaler Prävention nicht allein auf technologischer Machbarkeit beruhen. Vielmehr bedarf es eines systematischen, evidenzbasierten und patientenzentrierten Ansatzes. Die folgenden Abschnitte beleuchten daher zentrale Erkenntnisse zur Wirkung digitaler Anwendungen, benennen bestehende methodische Lücken und leiten Handlungsperspektiven für Forschung und Versorgung ab.



## 6.2 Kritische Bewertung der Evidenzlage

Die bisherigen Studien zu DiGA weisen erhebliche methodische Schwächen auf. Zu den zentralen Problemen zählen kurze Beobachtungszeiträume, subjektive Endpunkte, fehlende Placebo-Kontrollen sowie eine begrenzte Transparenz der Studien. In vielen Fällen ist kein Zugang zu vollständigen Studienberichten möglich, was die Nachvollziehbarkeit der Ergebnisse einschränkt. Auch die Replizierbarkeit bleibt häufig ungesichert.

Wie gezeigt wurde, basieren viele Ergebnisse auf stark selektiven, digital affinen Nutzergruppen. Dadurch ist die Aussagekraft hinsichtlich der breiten Versichertenpopulation begrenzt. Die meisten Studien berichten lediglich den Average Treatment Effect on the Treated (ATT), also den durchschnittlichen Effekt unter jenen, die die Intervention tatsächlich genutzt haben. Aussagen über den Average Treatment Effect (ATE) – also den Effekt auf die Gesamtpopulation – sind damit nicht möglich.

Um der Heterogenität der Zielgruppen gerecht zu werden, wären Subgruppenanalysen oder explizite Interaktionstests erforderlich, etwa mit Merkmalen wie Alter, Geschlecht, digitaler Gesundheitskompetenz oder Gesundheitsstatus. Solche Verfahren ermöglichen die Schätzung von Conditional Average Treatment Effects (CATE) und damit die Beantwortung der Frage, für wen eine digitale Intervention besonders wirksam ist – und für wen nicht. Zwar lassen sich solche Analysen bei großen Fallzahlen prinzipiell auch innerhalb von RCTs durchführen, in der Praxis bleiben sie jedoch die Ausnahme.

Besonders kritisch zu bewerten sind die hohen Drop-out-Raten: Je nach Studie liegen diese zwischen 7 % und 30 %, bei psychischen Erkrankungen teilweise sogar deutlich höher. Da insbesondere weniger motivierte, stärker belastete oder kränkere Personen zum Abbruch neigen, besteht die Gefahr systematischer Verzerrungen – etwa in Form eines upward bias, wenn nur die „Responder“ in der Analyse verbleiben. Die meisten Studien unterlassen eine differenzierte Analyse der Abbrecher:innen, sodass unklar bleibt, wie stark diese Selektivität die Ergebnisse beeinflusst.

Ein Blick auf die aktuelle Entwicklung im DiGA-Verzeichnis bestätigt diese methodischen Bedenken: Von den bis Ende 2024 gelisteten 68 DiGA konnten lediglich 12 (18 %) bereits bei Aufnahme einen belastbaren Nutzen nachweisen. Die Mehrheit wurde nur zur Erprobung zugelassen, wobei viele Anwendungen entweder mangels Evidenz wieder gestrichen oder nicht dauerhaft in die Regelversorgung übernommen wurden – obwohl die GKV bereits während der Erprobung für die Kosten aufkam (GKV-Spitzenverband, 2025). Diese Zahlen unterstreichen die Dringlichkeit methodisch fundierter und transparenter Evaluationsstandards.

## 6.3 Kosten-Nutzen-Überlegungen

Eine fundierte Kosten-Nutzen-Analyse digitaler Gesundheitsanwendungen setzt verlässliche Wirksamkeitsnachweise voraus. Solange diese Evidenz – etwa durch belastbare Langzeitstudien mit geringen Drop-out-Raten und hoher externer Validität – nicht vorliegt, bleiben ökonomische

Bewertungen spekulativ. Derzeit mangelt es jedoch häufig an solchen Grundlagen: Die Studienlage ist durch kurze Beobachtungszeiträume, selektive Stichproben und methodische Unsicherheiten geprägt (vgl. Kapitel 5.4 und 6.2). Entsprechend vorsichtig müssen Aussagen zum potenziellen Nutzen im Verhältnis zu den Kosten getroffen werden.

Dabei ist die Notwendigkeit ökonomischer Bewertungen unbestritten – insbesondere im Hinblick auf eine nachhaltige Integration von DiGA in die Regelversorgung der GKV. Hierfür sind insbesondere Real-World-Daten aus der Versorgungspraxis erforderlich, die bislang jedoch kaum systematisch erfasst und ausgewertet werden. Auch Fragen der Verteilungsgerechtigkeit, der langfristigen Kostenentwicklung und der Entlastung klassischer Versorgungsstrukturen bleiben weitgehend unbeantwortet.

Im Vergleich zu anderen staatlich regulierten Bereichen – etwa der Arzneimittelzulassung oder der arbeitsmarktpolitischen Förderung durch die Bundesagentur für Arbeit – erscheinen die Anforderungen an die Evaluation digitaler Gesundheitsanwendungen bislang deutlich geringer. Zwar sind die derzeitigen Ausgaben für DiGA im Verhältnis noch vergleichsweise niedrig. Dennoch ist eine robuste Wirkungsevaluation auch hier essenziell, um Fehlinvestitionen zu vermeiden, Vertrauen in digitale Gesundheitsinnovationen zu sichern und eine evidenzbasierte Weiterentwicklung der Regelversorgung zu ermöglichen. Insbesondere wenn der GKV-Einsatz künftig ausgeweitet und mit höheren Ausgaben verbunden sein sollte, wird eine solide, unabhängige Kosten-Nutzen-Analyse unverzichtbar.

## **7 Empfehlungen für zukünftige Forschung**

Die bisherigen Analysen zeigen: Trotz der wachsenden Verbreitung digitaler Gesundheitsanwendungen fehlt es vielfach an methodisch überzeugender Evidenz, die eine fundierte Bewertung von Wirksamkeit, Zielgruppenpassung und Kosteneffizienz erlaubt. Um das Potenzial digitaler Prävention realistisch einzuschätzen und gezielt weiterzuentwickeln, sind mehrere Forschungsdesiderate zu adressieren:

**Langzeitstudien und nachhaltige Effekte:** Viele DiGA-Studien beschränken sich auf kurze Beobachtungszeiträume von wenigen Wochen oder Monaten. Um die langfristige Wirksamkeit und Stabilität digitaler Interventionen – insbesondere bei chronischen Erkrankungen – beurteilen zu können, sind Studien mit längerer Laufzeit dringend erforderlich.

**Integration von Real-World-Daten (RWD):** Der Einbezug von Routinedaten gesetzlicher Krankenkassen oder Versorgungsdaten aus der Praxis kann helfen, die externe Validität zu erhöhen und die Übertragbarkeit auf die breite Versichertenpopulation zu verbessern. Solche Real-World-Evidenz (RWE) ist bislang kaum systematisch verfügbar, obwohl sie essenziell für eine evidenzbasierte Versorgung ist. Werden Studien jedoch mit Routinedaten verknüpft, können methodische Probleme wie selektive Drop-outs oder Verzerrungen durch geeignete Verfahren – etwa gewichtete Regressionsmodelle, Inverse Probability Weighting, Sensitivity Analyses oder Multiple Imputation – erkannt

und zum Teil korrigiert werden. Dazu bedarf es allerdings fortgeschrittener Methodenkompetenz und eines transparenten Studiendesigns.

**Subgruppenanalysen und Heterogenität der Effekte:** Digitale Interventionen wirken nicht bei allen Nutzer:innen gleichermaßen. Künftige Studien sollten vermehrt Subgruppenanalysen durchführen, um bedingte Effekte (Conditional Average Treatment Effects, CATE) zu identifizieren – also zu untersuchen, für wen eine DiGA besonders geeignet ist und unter welchen Voraussetzungen. Dabei sind Merkmale wie Alter, Geschlecht, Gesundheitskompetenz, digitales Nutzungsverhalten oder Vorerkrankungen gezielt zu berücksichtigen.

**Evaluation außerhalb der Herstellerperspektive:** Derzeit sind viele Studien herstellerfinanziert oder unterliegen potenziellen Interessenkonflikten. Um Verzerrungen zu minimieren und unabhängige Evidenz zu schaffen, sind öffentlich geförderte oder durch Krankenkassen getragene Studien notwendig – idealerweise in Kombination mit Studienregistern und offenen Datenstandards.

**Transparenz, Replizierbarkeit und Zugang zu Studiendaten:** Eine solide wissenschaftliche Grundlage erfordert nachvollziehbare Studienprotokolle, öffentlich zugängliche Ergebnisse sowie eine transparente Dokumentation von Daten, Methoden und Auswertungen. In vielen Fällen besteht hier noch erheblicher Nachholbedarf.

Darüber hinaus sollten künftige Arbeiten auch internationale und nationale Studien einbeziehen, die sich mit digitalen Interventionen außerhalb des DiGA-Kontexts befassen, jedoch ähnliche gesundheitsbezogene Wirkungsversprechen machen – etwa in der Behandlung von Depression, Adipositas oder Rückenschmerzen. Wenn solche Studien methodisch robust sind, könnten sie wertvolle Hinweise auf übertragbare Wirkmechanismen und Kontextbedingungen liefern. Im Rahmen dieser Arbeit war eine systematische Analyse entsprechender Studien jedoch nicht möglich.

Insgesamt gilt: Eine konsequent evaluative Ausrichtung künftiger Forschungsprojekte ist entscheidend, um digitale Präventionsangebote gezielt zu verbessern, Ressourcen wirksam einzusetzen und Vertrauen bei Patient:innen wie auch bei Leistungsträgern zu stärken.

## **8 Fazit und Ausblick**

Digitale Gesundheitsanwendungen (DiGA) und weitere digitale Präventionsangebote eröffnen neue Möglichkeiten, gesundheitsförderliches Verhalten zu unterstützen und Versorgungslücken zu schließen. Insbesondere im System der gesetzlichen Krankenversicherung (GKV) werden sie zunehmend als potenzielle Instrumente zur Verbesserung der Versorgungsqualität, zur Stärkung der Gesundheitskompetenz und zur Entlastung klassischer Versorgungsstrukturen diskutiert.

Gleichzeitig zeigt die bisherige Evidenzlage erhebliche Schwächen: Viele Studien basieren auf kurzen Beobachtungszeiträumen, kleinen und selektiven Stichproben sowie unzureichender methodischer

Transparenz. Die Aussagekraft hinsichtlich Wirksamkeit, Zielgruppenpassung und Kosteneffizienz bleibt daher begrenzt. Auch hohe Drop-out-Raten und fehlende Subgruppenanalysen erschweren eine belastbare Bewertung der tatsächlichen Effekte.

Für eine sinnvolle Integration digitaler Prävention in die Regelversorgung bedarf es nicht nur technologischer Innovation, sondern vor allem evidenzbasierter Entscheidungsgrundlagen. Dazu gehören belastbare Wirkungsnachweise, differenzierte Analysen heterogener Zielgruppen sowie unabhängige, replizierbare Studien mit hoher externer Validität. Ebenso müssen ethische und datenschutzrechtliche Anforderungen konsequent berücksichtigt werden.

Derzeit sind viele DiGA nur zur Erprobung zugelassen. In zahlreichen Fällen fehlen belastbare Nachweise eines medizinischen Nutzens. Die Tatsache, dass dennoch GKV-Mittel zur Finanzierung dieser Anwendungen eingesetzt werden, unterstreicht die Dringlichkeit strenger Evaluationsstandards und einer transparenten Nutzenbewertung.

Zukünftig sollten digitale Präventionsangebote stärker im Rahmen öffentlich geförderter Forschung und unter Einbezug von Real-World-Daten entwickelt und evaluiert werden. Eine solche Orientierung würde nicht nur die wissenschaftliche Qualität der Evaluationen verbessern, sondern auch dazu beitragen, Vertrauen bei Patient:innen, Leistungserbringenden und Kostenträgern zu stärken.

Insgesamt gilt: Digitale Prävention ist kein Selbstzweck. Ihr Mehrwert bemisst sich an nachweisbaren Wirkungen – für die Gesundheit der Versicherten ebenso wie für die Effektivität des Versorgungssystems. Nur durch methodisch fundierte und praxisnahe Forschung kann dieses Potenzial langfristig gehoben werden.

## Literaturverzeichnis:

Albrecht, U.-V., & von Jan, U. (2018). Apps in der digitalen Prävention und Gesundheitsförderung. In R. Haring (Hrsg.), Springer Reference Pflege – Therapie – Gesundheit. Gesundheitswissenschaften (S. 433–441). Springer. [https://doi.org/10.1007/978-3-662-58314-2\\_40](https://doi.org/10.1007/978-3-662-58314-2_40)

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). (2025). DiGA-Verzeichnis. <https://diga.bfarm.de/de/verzeichnis>

Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). (2023). DiGA-Leitfaden (Stand: 28. Dezember 2023, Version 3.5). [https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga\\_leitfaden.pdf?\\_\\_blob=publicationFile](https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.pdf?__blob=publicationFile)

Bundesministerium der Justiz. (2024). Fünftes Buch Sozialgesetzbuch – Gesetzliche Krankenversicherung (SGB V). [https://www.gesetze-im-internet.de/sgb\\_5/](https://www.gesetze-im-internet.de/sgb_5/)

Busse, R., Blümel, M., & Spranger, A. (2017). Das deutsche Gesundheitssystem: Akteure, Daten, Analysen. Medizinisch Wissenschaftliche Verlagsgesellschaft.

Dittrich, F., Mielitz, A., Pustozarov, E., Lawin, D., von Jan, U., & Albrecht, U.-V. (2024). Digital health applications from a government-regulated directory of reimbursable health apps in Germany—a systematic review for evidence and bias. *mHealth*, 10, 3. <https://doi.org/10.21037/mhealth-23-17>

Giebel, G. D., Speckemeier, C., Abels, C., & Borchers, K. (2023). Problems and barriers related to the use of mobile health applications: a scoping review. *JMIR mHealth and uHealth*, 11(1), e43808. <https://doi.org/10.2196/43808>

Groene, N., & Schneck, L. (2023). Covering digital health applications in the public insurance system: How to foster innovation in patient care while mitigating financial risks—evidence from Germany. *Frontiers in Digital Health*, 5, 1217479. <https://doi.org/10.3389/fdgth.2023.1217479>

Hernández Rodríguez, J. C., García Muñoz, C., Ortiz Álvarez, J., Saigí Rubió, F., Conejo Mir, J., & Pereyra Rodríguez, J. J. (2022). Dropout rate in digital health interventions for the prevention of skin cancer: Systematic review, meta analysis, and metaregression. *Journal of Medical Internet Research*, 24(12), e42397. <https://doi.org/10.2196/42397>

Imbens, G. W., & Rubin, D. B. (2015). Causal inference for statistics, social, and biomedical sciences: An introduction. Cambridge University Press.

Lauer, W., Löbker, W., & Höfgen, B. (2021). Digitale Gesundheitsanwendungen (DiGA): Bewertung der Erstattungsfähigkeit mittels DiGA-Fast-Track-Verfahrens im Bundesinstitut für Arzneimittel und

Medizinprodukte (BfArM). Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz, 64(10), 1232–1240. <https://doi.org/10.1007/s00103-021-03409-7>

Mäder, M., Timpel, P., Schönfelder, T., Häckl, D., Militzer-Horstmann, C., Scheibe, S., & Heinrich, R. (2023). Evidence requirements of permanently listed digital health applications (DiGA) and their implementation in the German DiGA directory: An analysis. *BMC Health Services Research*, 23, 369. <https://doi.org/10.1186/s12913-023-09287-w>

Mayer, T., Koska, D., Harsch, A.-K., & Maiwald, C. (2024). Prospektive randomisierte kontrollierte klinische Studie zur Wirksamkeit eines Eigentrainings mit der Digitalen Gesundheitsanwendung „Mawendo“ bei der Behandlung von Krankheiten der Patella (ICD-10-GM M22). *Qucosa*. <https://doi.org/10.60687/ubcq-2498>

Meyerowitz-Katz, G., Ravi, S., Arnolda, L., Feng, X., Maberly, G., & Astell-Burt, T. (2020). Rates of attrition and dropout in app-based interventions for chronic disease: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 22(9), e20283. <https://doi.org/10.2196/20283>

Nagel, J., Wegener, F., Grim, C., & Hoppe, M. W. (2024). Effects of digital physical health exercises on musculoskeletal diseases: Systematic review with best evidence synthesis. *JMIR mHealth and uHealth*, 12, e50616. <https://doi.org/10.2196/50616>

Nwosu, A., Boardman, S., Husain, M. M., & Doraiswamy, P. M. (2022). Digital therapeutics for mental health: Is attrition the Achilles heel? *Frontiers in Psychiatry*, 13, 900615. <https://doi.org/10.3389/fpsy.2022.900615>

Scherenberg, V. (2022). Digitalisierung in Prävention und Gesundheitsförderung. In Bundeszentrale für gesundheitliche Aufklärung (Hrsg.), *Leitbegriffe der Gesundheitsförderung und Prävention (Glossar)*. <https://doi.org/10.17623/BZGA:Q4-i130-1.0>

Wasem, J. (Hrsg.), Matusiewicz, D., Neumann, A., & Noweski, M. (Hrsg.). (2019). *Medizinmanagement: Grundlagen und Praxis des Managements in Gesundheitssystem und Versorgung* (2. Aufl.). Medizinisch Wissenschaftliche Verlagsgesellschaft.

