

Bundesgesundheitsbl 2021 · 64:1269–1277  
<https://doi.org/10.1007/s00103-021-03413-x>  
 Eingegangen: 30. April 2021  
 Angenommen: 20. August 2021  
 Online publiziert: 15. September 2021  
 © Der/die Autor(en) 2021



Lars G. Hemkens<sup>1,2,3</sup>

<sup>1</sup> Basel Institute for Clinical Epidemiology and Biostatistics (ceb), Department of Clinical Research, University Hospital Basel, Basel, Schweiz

<sup>2</sup> Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, USA

<sup>3</sup> Meta-Research Innovation Center Berlin (METRIC-B), Berlin Institute of Health, Berlin, Deutschland

# Nutzenbewertung digitaler Gesundheitsanwendungen – Herausforderungen und Möglichkeiten

## Einleitung

Digitale Technologien werden zunehmend zur Prävention, Diagnose oder Behandlung von Erkrankungen eingesetzt. Sie sind das Kernelement digitaler Gesundheitsanwendungen (DiGA), die zum Beispiel als Software auf Smartphones vielfältige Möglichkeiten versprechen, die Gesundheit und medizinische Versorgung von Patienten zu verbessern [1]. Seit 2019 können in Deutschland gesetzliche Krankenversicherungen die Kosten verordneter DiGA erstatten [1]. Wie bei jeder klinischen Entscheidung muss auch hier die Frage nach dem Nutzen und Schaden für Patienten gestellt werden.

Dieser Beitrag soll eine kurze Übersicht zur evidenzbasierten Nutzenbewertung und zu den Herausforderungen an die zugrunde liegende Evidenz als Voraussetzungen für eine optimale Entscheidungsfindung geben. Er beschreibt klassische Studiendesignkonzepte, aktuelle Entwicklungen und innovative Ansätze mit dem Ziel, zukünftige Entwicklungsfelder für innovative Studiendesigns und strategische Evaluationskonzepte für DiGA aufzuzeigen.

Der Beitrag fokussiert auf den gesundheitsbezogenen Nutzen oder Schaden, den Patienten bei der Behandlung von Krankheiten haben können, und daher nicht auf Primärprävention. Auf rechtliche, regulatorische, technologische und ökonomische Aspekte wird nicht

eingegangen, ebenfalls nicht auf Rahmenbedingungen zur Zulassung oder Erstattung in bestimmten Gesundheitssystemen sowie Fragen des Datenschutzes oder der Informationssicherheit.

## Grundlagen der Nutzenbewertung

### Nutzen und Schaden

Wenn Patienten und Ärzte entscheiden müssen, welche Behandlungsoption die beste ist, haben sie als Entscheidungsgrundlage idealerweise optimale Informationen zu den erwartbaren Effekten auf für sie relevante Parameter (Endpunkte). Veränderungen von Lebensqualität, Morbidität und Mortalität sind unmittelbar patientenrelevant und zentral für Behandlungsentscheidungen [2, 3]. Der Nutzen einer medizinischen Handlung liegt in der Verbesserung dieser Endpunkte, ein Schaden in einer Verschlechterung.

Eine DiGA kann sich auf andere Parameter, z. B. die Strukturen und Prozesse der Gesundheitsversorgung auswirken. Hierfür wurde in Deutschland im Kontext von Erstattungsentscheidungen der Begriff „positiver Versorgungseffekt“ in Abgrenzung zu „Nutzen“ geprägt.

## Evidenzbasierte Entscheidungen

Bei der evidenzbasierten Entscheidungsfindung wird die Evidenz zum Nutzen

und Schaden abgewogen unter Einbezug der jeweiligen Situation sowie der Wertvorstellungen und Präferenzen des Patienten [4]. Auf Methoden zur empirischen Bestimmung von Patientenpräferenzen wird an dieser Stelle nicht weiter eingegangen [5, 6].

Die prinzipiellen Anforderungen an evidenzbasierte Entscheidungsgrundlagen sind unabhängig von der Art der Behandlungen (z. B. Medikamente, chirurgische Therapie, DiGA), zwischen denen man sich entscheidet, und sollen eine zuverlässige Abschätzung von Nutzen und Schaden ermöglichen.

Theoretisch würden Entscheidungsgrundlagen Evidenz zur exakten Vorhersage der Folgen einer Handlung bzw. Behandlung und der jeweiligen Alternative bereitstellen und exakt den kausalen Zusammenhang zwischen Handlungsentscheidung und Ergebnis (Endpunkt) beschreiben [7]. In der Praxis verfügbare Evidenz ist hierin jedoch nur mehr oder weniger zuverlässig und wird mit ebenfalls mehr oder weniger Aufwand generiert. Dabei wird zunehmend erkannt, dass eine hohe Zuverlässigkeit nicht mit einem hohen Aufwand einhergehen muss [8].

## Klinisch-epidemiologische Grundlagen zu Evidenz und ihrer Zuverlässigkeit

### Randomisierung

Randomisiert kontrollierte Studien (engl. „randomized controlled trials“, RCTs) liefern die zuverlässigste Evidenz zum Nutzen und Schaden von Handlungsentscheidungen [9]. Denn sie erlauben es, kausale Effekte von Handlungen direkt abzuschätzen [7], und adressieren unmittelbar das kritische Problem von Verzerrungen des Zusammenhangs zwischen Exposition und Gesundheitsergebnis durch eine fremde dritte Variable (Confounding-Bias, s. unten). Die Randomisierung stellt 2 Vergleichsgruppen her, zwischen denen sämtliche Unterschiede (z.B. Risikofaktoren für einen schweren Krankheitsverlauf) zum Zeitpunkt der Entscheidung rein zufällig sind und daher recht einfach statistisch berücksichtigt werden können. In RCTs ist es nicht erforderlich zu wissen, welche Risikofaktoren relevant wären; entsprechende Daten müssen nicht erhoben werden, um eventuelle Dysbalancen zwischen Vergleichsgruppen zu suchen oder zu beschreiben [9]. Natürlich ist es oft hilfreich, Daten zu den Charakteristika der Studienpopulation zu haben, aber grundsätzlich ist es optional. In ihrer Reinform benötigen RCTs aus methodischer Sicht also nur die zufällige Zuteilung und die Messung von 2 Variablen – die Gruppenzugehörigkeit und einen Endpunkt.

### Keine Randomisierung

Nichtrandomisierte Studien haben den Vorteil der zufälligen Zuteilung nicht. Daher benötigen sie statistische Verfahren zur Adjustierung von Unterschieden zwischen den Vergleichsgruppen, um das Risiko für Bias durch Confounding bei fehlender prognostischer Balance zu verringern und so zumindest eine gewisse Vertrauenswürdigkeit zu gewährleisten [10]. Dazu muss man nicht nur sämtliche relevante Faktoren bzw. Confounder und ihre Interaktion kennen ([10–12]; was eine sehr starke und oft unrealistische Voraussetzung und auch eine sub-

jektive Annahme ist), sondern man muss all diese Daten auch mit hohem Aufwand und hinreichend präzise sammeln (eine Aufgabe, die bei RCTs entfällt).

Für eine nichtrandomisierte Studie zum Effekt von DiGA auf Lebensqualität z. B. braucht es ein genaues Verständnis darüber, welche Faktoren mit Lebensqualität assoziiert sind und wie sie mit der Verordnung von DiGA zusammenhängen. Psychosoziale oder sozioökonomische Faktoren haben jedoch oft komplexe und unklare Zusammenhänge mit Lebensqualität, Morbidität oder Mortalität [13] und sind oft mit Therapieadhärenz und anderen Faktoren assoziiert [13]. Routinedaten für solche Faktoren sind selten vorhanden, was eine aktive (prospektive) Datensammlung mit erheblichem Mehraufwand (Datenmanagement, Qualitätsmanagement, Monitoring) erfordern würde [14]. Selbst in intensiv beforschten Bereichen herrscht zudem oft kein Konsens dazu, welche Faktoren bzw. Confounder auf welche Weise in Analysemodelle eingehen sollen [15].

Allerdings kann die Wahl des Modells die Ergebnisse massiv beeinflussen und durchaus eine geschätzte Risikoerhöhung in eine Risikosenkung verdrehen [16]. Diese im Vergleich zu RCTs ungleich höhere *Vibration of Effects* hat ein extremes Potenzial für bewusste oder unbewusste Einflussnahme und reduziert die Zuverlässigkeit weiter [12, 17, 18]. Hoch detaillierte Studienprotokolle und vorab festgelegte Analysepläne zum Schutz vor solchen Mechanismen wären essenziell, existieren jedoch nur selten – ganz im Gegensatz zur Situation bei RCTs, obwohl dort die Analysen sogar deutlich weniger komplex sind (da es diese Adjustierungen nicht braucht; [12]). Zahlreiche systematische metaepidemiologische Untersuchungen haben gezeigt, dass die Ergebnisse von RCTs und nichtrandomisierten Studien oft deutlich abweichen [12], auch wenn moderne statistische Verfahren verwendet werden (z. B. Propensity Scores, kausale Modelle; [17, 19]).

Insgesamt erlauben nichtrandomisierte Studien (oder gar Studien ohne Kontrollgruppe) keine verlässlichen Schlussfolgerungen zu Effekten von

Therapieentscheidungen (von seltenen Situationen abgesehen; [9, 12]), während sie gleichzeitig oft einen hohen Zusatzaufwand mit sich bringen.

## Evaluation digitaler Innovationen mit randomisierten Studien außerhalb der Gesundheitsversorgung

Den Vorteil randomisierter Studien haben die erfolgreichsten Unternehmen der digitalen Wirtschaft (z. B. Microsoft, Google, Facebook, Amazon, Netflix) längst erkannt und nutzen sie systematisch zur kontinuierlichen Evaluation und Verbesserung ihrer Produkte, auch wenn ihnen große Datenmengen zur Verfügung stehen, die mitunter als Argument für nichtrandomisierte Studien angeführt werden [9, 20]. Tatsächlich finden die meisten RCTs heutzutage als sog. A/B-Tests in der digitalen Technologierbranche statt [20]. Anstatt Fortschritt zu behindern, beschleunigen sie Innovation und führen zu einer nachhaltigen Verbesserung wichtiger Kennzahlen (z. B. Umsatz oder Benutzerzufriedenheit; [20]). Jede noch so kleine Verbesserung der digitalen Produkte wird trotz der schnellen Innovationszyklen in kürzester Zeit einer systematischen, randomisierten Testung unterworfen [20]. Dies weist darauf hin, dass dies technisch und methodisch auch für DiGA möglich und vielversprechend wäre. Gleichzeitig zeigt es, dass Herausforderungen für schnelle Evaluationen, Anpassungen und Reaktionen nicht primär technisch-methodischer Natur sind, sondern sich z. B. aus regulatorischen Umständen ergeben.

Der nachhaltige Stellenwert für eine strukturelle Verbesserung des Gesamtsystems wird daran erkennbar, dass randomisierte Untersuchungen zahlreiche nutzlose Innovationen ausfiltern konnten, die zu keiner Verbesserung führten. Dies betraf 2 von 3 der vielversprechendsten Ideen, in manchen Bereichen lagen die Ausfälle im Bereich von 80–90 % [20].

## Herausforderungen: andere Studiendesigns oder andere randomisierte Studien?

Bei der Betrachtung der Evidenz zur Nutzenbewertung gilt es zu differenzieren, ob es andere Studiendesigns als RCTs braucht oder vielmehr andere RCTs [12]. Ein Großteil aller RCTs ist in der Tat ungeeignet, die Fragen zu beantworten, die für Entscheidungsträger, Kliniker und Patienten relevant sind, und kann daher in vielerlei Hinsicht als nutzlos angesehen werden [21]. Darauf wiesen Schwartz und Lellouch bereits 1967 hin [22]. Sie entwickelten das Konzept „explanatorischer“ und „pragmatischer“ RCTs, das zunehmend wieder in den Fokus gelangt. Für die Rahmenbedingungen zur Schaffung und Einordnung von Evidenz zur Nutzenbewertung von DiGA sind diese Konzepte essenziell.

### Explanatorische Studien

Zunehmend wird erkannt, dass die steigenden Herausforderungen bei der Durchführung von RCTs nicht am Studienprinzip liegen, sondern an der konzeptionellen Umsetzung neben gesundheitspolitischen und akademischen Rahmenbedingungen [9, 23]. Diese resultiert aus der Tradition der Arzneimittelforschung und hier vor allem der frühen Entwicklungsphase. Studien dieser Phase gleichen meist hoch kontrollierten Laborexperimenten und haben ein sogenanntes explanatorisches Design [22]. Sie wollen ein besseres Verständnis der pathophysiologischen Wirkmechanismen möglichst ohne Störeinflüsse durch z. B. suboptimale Therapieadhärenz liefern [24]. Placebos werden eingesetzt, um möglichst sauber den Effekt einer hoch standardisierten Verabreichung eines Wirkstoffes zu messen. Sie haben strenge Einschlusskriterien [25, 26] und viele Follow-up-Untersuchungen, um Variablen zu messen, die wichtig sein können, die Arzneimittel zu verbessern oder ihre Mechanismen zu verstehen, aber nicht unbedingt patientenrelevant sind (z. B. Biomarker oder Pharmakovigilanzdaten; [24]).

Bundesgesundheitsbl 2021 · 64:1269–1277 <https://doi.org/10.1007/s00103-021-03413-x>  
© Der/die Autor(en) 2021

L. G. Hemkens

## Nutzenbewertung digitaler Gesundheitsanwendungen – Herausforderungen und Möglichkeiten

### Zusammenfassung

Digitale Gesundheitsanwendungen (DiGA) versprechen, die Gesundheit und medizinische Versorgung von Patienten zu verbessern. Dieser Beitrag gibt eine kurze Übersicht zur evidenzbasierten Nutzenbewertung und den Herausforderungen an die zugrunde liegende Evidenz als Voraussetzungen für eine optimale, patientenorientierte Entscheidungsfindung. Es werden klassische Konzepte des Studiendesigns, aktuelle Entwicklungen und innovative Ansätze beschrieben mit dem Ziel, zukünftige Entwicklungsfelder für innovative Studiendesigns und strategische Evaluationskonzepte für DiGA aufzuzeigen. Ein besonderer Schwerpunkt liegt auf der Anwendung pragmatischer Studiendesigns. Evidenzbasierte Nutzenbewertung hat fundamentale Anforderungen und Maßstäbe unabhängig von der Art der evaluierten Behandlungen. Zuverlässige Evidenz ist unverzichtbar. Eine schnelle, effiziente, zuverlässige und praxisrelevante Evaluation von DiGA gelingt nicht durch eine Hinwendung

zu nichtrandomisierten Studien, sondern vielmehr durch bessere, pragmatische randomisierte Studien. Sie sind machbar und verbinden die Charakteristika von DiGA, klassische methodische Konzepte und neue Ansätze der Studiendurchführung. Routinedaten, kontaktarme Studiendurchführung („virtual trials“, „remote trials“) und digitale Biomarker fördern nützliche randomisierte Real-World-Evidenz als solide Evidenzbasis von DiGA. Eine kontinuierliche, lernende Evaluation im Versorgungsalltag mit in die Routine eingebetteten randomisierten Studiendesigns ist der Schlüssel zur nachhaltigen und effizienten Nutzenbewertung von DiGA und kann entscheidend für eine strategische Verbesserung der Gesundheitsversorgung sein.

### Schlüsselwörter

Digital Health · Mobile Anwendungen · Nutzenbewertung · Pragmatische Studien · Studiendesign

## Benefit assessment of digital health applications—challenges and opportunities

### Abstract

Digital health applications promise to improve patient health and medical care. This analysis provides a brief overview of evidence-based benefit assessment and the challenges to the underlying evidence as prerequisites for optimal patient-oriented decision making. Classical concepts in study design, recent developments, and innovative approaches are described with the aim of highlighting future areas of development in innovative study designs and strategic evaluation concepts for digital health applications. A special focus is on pragmatic study designs. Evidence-based benefit assessment has fundamental requirements and criteria regardless of the type of treatments evaluated. Reliable evidence is essential. Fast, efficient, reliable, and practice-relevant evaluation of digital health applications is not achieved by turning to nonrandomized trials,

but rather by better pragmatic randomized trials. They are feasible and combine the characteristics of digital health applications, classical methodological concepts, and new approaches to study conduct. Routinely collected data, low-contact study conduct (remote trials, virtual trials), and digital biomarkers promote useful randomized real-world evidence as solid evidence base for digital health applications. Continuous learning evaluation with randomized designs embedded in routine care is key to sustainable and efficient benefit assessment of digital health applications and may be crucial for strategic improvement of healthcare.

### Keywords

Digital health · Mobile applications · Benefit assessment · Pragmatic trials · Research design

## Herausforderungen und Limitationen

Die meistgenannten Limitationen von RCTs ergeben sich direkt aus diesen Charakteristika explanatorischer Studien, ebenso die Herausforderungen bei der Durchführung von RCTs. Der hohe Aufwand erzeugt hohe Kosten, umso mehr, je größer und länger die Studie ist. Aus Kostengründen werden dann kleinere und kürzere Studien geplant, die es nicht erlauben, patientenrelevante Endpunkte und Langzeiteffekte zu messen oder relevante Subgruppeneffekte zu finden. Strikte Einschlusskriterien erschweren die Rekrutierung und können die Anwendbarkeit der Ergebnisse einschränken. Placebos erhöhen den Aufwand zudem mitunter beträchtlich (geschätzt 10 % des Gesamtstudienbudgets [27]) und führen zu einem Behandlungsvergleich, der unter Real-World-Bedingungen nicht existiert und reale Therapieentscheidungen nicht abbildet [24].

## Pragmatische Studien: Randomisierte Real-World-Evidenz

Für die Nutzenbewertung sind viele dieser Charakteristika nicht nur unnötig, sondern sogar kontraproduktiv. Studien, die Evidenz für Behandlungsentscheidungen liefern sollen, sollten diese strikten Eigenschaften explanatorischer Studien vermeiden [24]. Entscheidungsorientierte, sogenannte pragmatische RCTs haben die meisten traditionellen Limitationen von (explanatorischen) RCTs nicht. Dabei steht „pragmatisch“ keineswegs für niedrigere Ansprüche an Sorgfalt, Zuverlässigkeit oder für methodische Abkürzungen. Vielmehr steht die direkte Handlungsorientierung im Vordergrund. Pragmatische Studien liefern definitionsgemäß nützliche, patientenorientierte Evidenz, die direkt in Entscheidungen einfließen kann und möglichst wenig Annahmen zur Übertragbarkeit von der Studie auf die Praxis erfordert [24]. Sie finden entsprechend nicht unter artifiziellen Bedingungen mit selektierten Patienten und Placebokontrollen statt.

## Machbarkeit

Die Machbarkeit hochwertiger pragmatischer RCTs zur kurzfristigen Bewertung neuer Interventionen hat das RECOVERY Platform Trial bei COVID-19 eindrucksvoll belegt. In 2 Tagen geplant, wurde nach 9 Tagen der erste und nach 2 Monaten der 10.000. Patient eingeschlossen [28, 29]. In kürzester Zeit wurden effektive Behandlungen gefunden und zahlreiche Therapien als nicht nutzbringend identifiziert (obwohl sie aufgrund mechanistischer Überlegungen plausibel und teilweise in großen nichtrandomisierten Studien vielversprechend waren; [28, 29]). Die Gründe für den Erfolg waren nicht nur die im Vereinigten Königreich solide etablierten Strukturen für klinische Studien und akademischen Rahmenbedingungen [23]. Entscheidend waren auch das sehr pragmatische Studiendesign, mit einer präzisen Reduktion der Prozesse auf das Wesentliche (z. B. mit lediglich 3 grundlegenden Einschlusskriterien), eine konsequente Einbettung in den Versorgungsalltag und die Nutzung von Routinedaten [28, 29].

## Routinedaten

Routinedaten (z. B. aus Registern, Krankenhausdatenbanken oder Abrechnungsdaten) können die Durchführung von RCTs substanziell verbessern [8, 30, 31]. Statt aufwendig neue Dateninfrastrukturen aufzubauen, werden Daten benutzt, die sowieso erhoben werden. Das übrige Studiendesign eines RCT bleibt unverändert (d. h. eine randomisierte Zuordnung der Intervention), aber statt z. B. aktiv Erkundigungen einzuholen, ob ein Teilnehmer hospitalisiert war, wird z. B. der Versicherer angefragt. RCTs können auch vollständig in vorhandene Datenstrukturen eingebettet werden (z. B. Register oder Kohorten [31]) oder traditionelle Datenerhebung wird verknüpft mit Abfragen bestehender Datenquellen [32]. Mit Routinedaten können nicht nur Endpunkte gemessen, sondern auch Studienteilnehmer effizienter rekrutiert werden [8]. So werden RCTs zu einem Bruchteil der bisher üblichen Kosten möglich (in einer Größenordnung von 50–2000 USD pro Patient; [33]).

Darüber hinaus entfallen künstliche Datenerhebungen und Interaktionen mit Teilnehmern nur für Studienzwecke und so ergibt sich eine größere Nähe zum normalen Versorgungsalltag. Routinedaten mit besonderer Bedeutung für DiGA sind solche, die z. B. von mobilen Geräten direkt gemessen werden (digitale Biomarker; s. unten). Natürlich muss eine hinreichende Qualität der Routinedaten gewährleistet sein, was jedoch zunehmend erwartet werden kann [14].

Insgesamt liefern pragmatische Studien naturgemäß eher die für eine Nutzenbewertung relevante Evidenz und sind ein Grundstein von Comparative Effectiveness Research bzw. Real World Evidence [32, 34, 35]. Sie vereinen hohe interne Validität von RCTs (hier stehen sie nichtpragmatischen RCTs in nichts nach) mit hoher externer Validität (Anwendbarkeit, Übertragbarkeit der Ergebnisse). Für die Evaluation von DiGA sind pragmatische RCTs daher zentral.

## Pragmatische Studien sind zentral für die Nutzenbewertung von DiGA

DiGA haben zahlreiche Charakteristika, die mitunter als Herausforderung für RCTs beschrieben wurden, aber bei näherer Betrachtung im Kontext von pragmatischen Studien keine oder nur eine geringe Rolle spielen. Viele Elemente explanatorischer Studien, die sich aus Parallelen zur frühen Arzneimittelentwicklung ergeben (z. B. Prüfungen der Verträglichkeit und Sicherheit, Dosisfindung), sind bei DiGA kaum relevant. Folglich treffen die Limitationen, die solche Elemente mit sich bringen für DiGA nicht zu. Weitere Elemente sind schon rein konzeptionell bei pragmatischen Studien nachrangig bis irrelevant.

## Ein- und Ausschlusskriterien

Die Studienpopulation entspricht in pragmatischen Studien der Zielpopulation, d. h. bei DiGA idealerweise den Patienten, denen sie verordnet werden würden. Ausschlüsse aus Sicherheitsgründen (z. B. wegen Begleiterkrankungen oder Arzneimittelwechselwirkungen) sind eigentlich nicht relevant. Eine



artifizielle Homogenisierung der Studienpopulation (z. B. Ausschluss multimorbider Patienten) mit dem Ziel geringerer Effektivität und damit geringerer benötigter Fallzahl wäre hier explizit unerwünscht. Das erleichtert die Rekrutierung erheblich, senkt Kosten und Aufwand und erlaubt größere Studien in kürzerer Zeit.

## Adhärenz

Bei pragmatischen Studien steht explizit die Entscheidung für eine Therapieoption, die Intention der Behandlung, im Vordergrund und nicht die Durchführung bzw. Umsetzung [22, 24]. Wenn Teilnehmer trotz bester Intention eine Therapie nicht so befolgen wie angedacht oder gar niemals beginnen, ist dies Bestandteil der Behandlung als Ganzes [24]. Komplexe Wechselwirkungen mit dem Verhalten der Patienten beeinträchtigen die Zuverlässigkeit pragmatischer RCTs zum Einsatz von DiGA nicht, da sie der randomisierten Therapieentscheidung nachfolgen und so Bestandteile eines intendierten Therapiekonzeptes mit DiGA sind (entsprechend folgt die Analyse dem Intention-to-treat (ITT)-Prinzip; als sog. „Treatment Policy Estimand“ [17, 24, 36, 37]).

Bestimmt wird also der kausale Effekt der Intention der Behandlung und nicht ihre plangemäße Umsetzung [17, 24]. Dies ist plausibel, denn es ist z. B. wahrscheinlich nicht nützlich, eine DiGA zu verordnen, die unter Routinebedingungen fast nie benutzt wird, weil sie als zu umständlich wahrgenommen wird. Ebenso wäre die Verordnung eines abscheulich schmeckenden Hustentees, der nie getrunken werden würde, ohne Nutzen. Der Effekt einer perfekten Umsetzung ist für pragmatische Studien nicht relevant, da dies unter Real-World-Bedingungen unrealistisch und maximal artifiziell ist (dies wäre Gegenstand einer explanatorischen Studie).

Eingeschränkte Adhärenz, Behandlungs-Cross-over („Kontamination“) und Verhaltensinteraktionen erzeugen keinen Bias der ITT-Effekte in RCTs (s. oben; [24]). Auch entfällt für pragmatische RCTs die Notwendigkeit, detaillierte Daten zur Adhärenz zu erheben,

was größere Studien realisierbarer macht und die Nähe zum Versorgungsalltag fördert. Gleichwohl lässt sich die Nutzung einer DiGA in vielen Fällen vergleichsweise einfach erfassen. So wurde in einer französischen pragmatischen RCT mit 2804 Teilnehmern zum Angebot einer Raucherentwöhnungs-App erkannt, dass sehr viele Teilnehmer die DiGA kaum oder gar nicht benutzt haben [38].

Eine Möglichkeit, den Nutzen bei Personen mit wahrscheinlich hoher Adhärenz zu testen, sind Run-in-Phasen zu Beginn einer Studie [39]. Patienten würde hier die DiGA testweise angeboten und die Studie würde diejenigen einschließen, die sie zumindest eine Zeit lang planmäßig verwenden. Da hier der Kontrollgruppe die DiGA vorenthalten würde (zumindest eine Zeit lang), könnte dies ein artifizielles Setting erzeugen und möglicherweise zu Unzufriedenheit und/oder Behandlungs-Cross-over führen (d. h., die Patienten beschaffen sich die DiGA selbst und wenden sie an, obwohl dies in der Studie nicht vorgesehen ist). Mit kausalen Modellen kann in RCTs versucht werden, die Effekte theoretisch optimaler Adhärenz zu messen (im Gegensatz zum ITT-Effekt; sog. Hypothetical Estimand; [36, 37, 40, 41]). Da hier jedoch die Randomisierung aufgegeben wird [40], bestehen wie bei nichtrandomisierten Studien ein sehr hohes Risiko für Bias, die Notwendigkeit, umfangreiche Daten zu sammeln, und weitere große Herausforderungen [41].

## Verblindung

Eine Verblindung von Arzt oder Patient ist in der üblichen Versorgung unrealistisch und erzeugt ein sehr artifizielles Behandlungssetting, was einem pragmatischen Studiendesign widerspricht. Die Zuverlässigkeit der Effektschätzer wird durch eine fehlende Verblindung nicht zwangsläufig beeinträchtigt, wenn analog den obigen Ausführungen zur Adhärenz argumentiert wird, dass die Kenntnis der Behandlung (mit allen sich daraus ergebenden Konsequenzen) nicht Bias erzeugt, sondern ein inhärenter Teil der Behandlungsentscheidung selbst ist [24]. Davon unbenommen ist die Notwendigkeit verblindeter Endpunk-

terhebung, da sonst ein Risiko für Bias bei der Ergebnismessung besteht. Eine Verblindung des Untersuchers bzw. der Studiendurchführung kann ggf. in gewissem Ausmaß vor bewusst verzerrenden, manipulativen Einwirkungen schützen (z. B. suboptimale Begleitbehandlungen der Kontrollgruppe). Routinedaten z. B. sind in der Regel formal verblindet, da die Erfassung von z. B. Hospitalisierungen durch Versicherer automatisch bzw. in Unkenntnis etwaiger DiGA-Verordnungen stattfindet. Die Verwendung subjektiver Endpunkte oder im Extremfall Patient-reported Outcomes (PROs), deren Messung sich durch die Kenntnis der Behandlung beeinflussen lässt, kann mit einem hohen Risiko für Bias einhergehen.

Somit ist die Verwendung von PRO-Endpunkten, wie z. B. Schmerzskaleten oder Lebensqualität, bei der Evaluation von DiGA kritisch zu betrachten, wenn nicht tatsächlich adäquate Placebo-DiGA verwendet werden. DiGA zu verblinden ist nicht einfach (z. B. mittels Placebo-DiGA). Ob jedoch der im Versorgungsalltag nichtexistierende Vergleich mit Placebo-DiGA hilfreich ist, um Effekte auf z. B. Lebensqualität zu messen, muss diskutiert werden. Weniger artifiziel wäre ein Vergleich von 2 vom Patienten nicht unterscheidbaren DiGA-Versionen (die z. B. unterschiedliche Algorithmen anwenden).

## Besonderheiten bei der Evidenzgenerierung zu DiGA

Während die Prinzipien zur Nutzenbewertung und Evaluation von DiGA und anderen medizinischen Maßnahmen sich nicht unterscheiden und durch eine Hinwendung zu pragmatischen Studien auch traditionelle Hürden bei der Evidenzgenerierung ihre vermeintliche Relevanz verlieren, haben DiGA doch einige Besonderheiten (■ Tab. 1).

Vor allem die oftmals kurzen Innovationszyklen von DiGA werden gelegentlich als hinderlich für die Durchführbarkeit von RCTs angeführt. Die Technologien können sich so schnell verändern, dass neue Versionen bereits verfügbar sind, bevor die Bewertung des Vorgängers abgeschlossen ist. Ein Schlüssel zur Nutzen-

**Tab. 1** Besonderheiten von digitalen Gesundheitsanwendungen (DiGA) und Relevanz für Evidenzgenerierung und Nutzenbewertung

Charakteristikum	Relevanz für Evidenzgenerierung	Relevanz für Nutzenbewertung	Anmerkungen
<b>DiGA-spezifisch</b>			
<i>Sicherheitsprofil</i> (geringes Risiko für schwere Nebenwirkung, keine strikte Indikationsstellung wegen Wechselwirkungen/ Begleiterkrankungen)	Kaum Ausschlussgründe	Höhere Anwendbarkeit/ Übertragbarkeit	Solide Festlegung des Sicherheitsprofils erfordert ebenfalls adäquate Evidenz
	Leichtere und schnellere Rekrutierung	Größere Studien (präzise Effektschätzer; ggf. Subgruppeninformationen)	
	Kontaktärmere Studie möglich <sup>a</sup>		
	Weniger Kosten und Aufwand		
	Größere Studie möglich		
Leichtere Einbettung in Routineversorgung			
<i>Verfügbarkeit und Distribution</i> (Leichter Zugang, un-mittelbarer Roll-out)	Einfachere Logistik	Größere Studien (präzise Effektschätzer; ggf. Subgruppeninformationen)	Skalierbarkeit oft leichter als bei Arzneimitteln (d. h. Kosten für DiGA oft weniger abhängig von Benutzerzahl)
	Kontaktärmere Studie möglich		
	Weniger Kosten und Aufwand		
	Größere Studie möglich		
Leichtere Einbettung in Routineversorgung			
<i>Kurze Innovationszyklen</i> (Neue Versionen, sich verändernde Algorithmen)	DiGA kann sich während der Evaluation ändern: detaillierte Planung und Präspezifizierung nötig	Schnelle Bewertung notwendig	Reproduzierung von Studien mit obsoleten Versionen ggf. schwierig
	Plattform-Trials/adaptive Designs	Definition/Abgrenzung neuer Versionen notwendig	Follow-up ggf. eingeschränkt
<i>Digitale Biomarker</i> (Endpunkte durch digitale Geräte erhoben, neuartige Endpunkte)	Kontaktärmere Studie möglich <sup>a</sup>	Endpunktvalidierung notwendig Klärung der Patientenrelevanz	Endpunktmessung nur valide, wenn bei allen Studienteilnehmern gleich und unabhängig von der Gruppenzuteilung
	Erfordert digitale Kontrollintervention		
<i>Verblindung (Placebo-DiGA)</i> (Arzt/Therapeut, Patient)	Wenn möglich zu vermeiden	Ggf. hohes Risiko für Bias bei subjektiven Endpunkten (z. B. PRO, LQ) und digitalen Biomarkern	Einfluss auf Bias sorgfältig diskutieren
	Unrealistisches Setting		
	Hoher Aufwand in der Umsetzung		
	Zusatzkosten durch Placebo		
Nötig für digitale Biomarker und PRO			
<i>Verblindung (DiGA Version 1 vs. Version 2)</i> (Arzt/Therapeut, Patient)	Ggf. kein unrealistisches Setting	Bewertung subjektiver Endpunkte und digitaler Biomarker ggf. möglich	./
	Ggf. kein hoher Aufwand	Kontinuierliche Evaluation	
	Ggf. kaum Zusatzkosten		
	Verwendung digitaler Biomarker und PRO ggf. gut möglich		
Zentral für lernende Evaluation			
<b>Pragmatische Studien generell</b>			
<i>Adhärenz</i> (inkl. Cross-over, Kontamination)	Nicht relevant (ITT-Analyse)	Kein Risiko für Bias in pragmatischen Studien	Daten zur Adhärenz könnten helfen, Wirkweisen der DiGA-Behandlung zu verstehen
	Keine aufwendige Datensammlung nötig		
	Keine spezielle Motivation (außerhalb der Routine) nötig und sinnvoll		Run-in-Phasen sind ggf. hilfreich
	Kontaktärmere Studie möglich		
	Weniger Kosten und Aufwand		
Größere Studie möglich			
<i>Verblindung (Endpunkterhebung)</i>	Immer anzustreben	./	./

ITT Intention to Treat, LQ Lebensqualität, PRO Patient-reported Outcome

<sup>a</sup>Remote Trials/Virtual Trials

bewertung kann hier in einer kontinuierlichen, lernenden Evaluation sich kontinuierlich verändernder Varianten derselben DiGA liegen, die ständig (randomisiert) miteinander verglichen werden, ähnlich wie es heutzutage schon im nicht-medizinischen Bereich digitaler Technologie üblich ist [20].

Möglich wird eine solche kurzfristige Evaluation durch die in der Regel vorhandene schnelle Verfügbarkeit und Distribution neuer DiGA-Versionen, eine angemessene Studienplattform, eine hinreichende Dateninfrastruktur und ggf. weitere neue Ansätze zur Studiendurchführung.

## Verfügbarkeit und Distribution

Der logistische Aufwand bei der Distribution einer DiGA ist weniger komplex als bei Arzneimittelprüfungen, da DiGA online zum Anwender übertragen werden können und klassische Anforderungen an Transport, Lagerung, Bilanzierung des Verbrauchs etc. entfallen und Zeit, Kosten und Aufwand auch für Monitoring gespart werden. Gleichwohl ergeben sich andere Herausforderungen (Datensicherheit etc.). Entscheidend ist, dass langwierige logistische Planungen vermieden und sehr kurzfristig neuere Versionen einer DiGA zur Evaluation eingesetzt werden können.

## Lernende Evaluation im Versorgungsaltag

Eine zeitnahe und effiziente Evaluation verschiedener Versionen einer DiGA erfordert eine bereits vorhandene Studienplattform (Plattform-Trials bzw. Masterprotokolle [42, 43]) mit festgelegten grundsätzlichen Rahmenbedingungen der Evaluation (z. B. Einschlusskriterien, Endpunkte, statistische Analyse). Im detaillierten Studienprotokoll würde prädefiniert, ab wann kleinere Verbesserungen, Bugfixes und größere Updates tatsächlich einer neuen Version entsprechen [44]. Diese Versionen können dann in einer kontinuierlichen (adaptierten) Evaluation systematisch als hinzukommende Interventionen in zusätzlichen Studienarmen vergleichend untersucht werden, während andere

Studienarme nach abgeschlossener Evaluation wegfallen. Eine kontinuierliche Weiterentwicklung im Rahmen der Anwendung kann erfolgen. Das ist eine im Vergleich zu herkömmlichen medizinischen Interventionen grundsätzliche und vielversprechende Neuerung. Sofern Unterschiede der Versionen für Nutzer nicht erkennbar sind (z. B. unterschiedliche Algorithmen), würde diese formale Verblindung eine Analyse subjektiver Endpunkte möglich machen.

**Fixe oder verändernde DiGA.** Für die Nutzenbewertung muss klar festgelegt werden, was evaluiert wird – eine definierte DiGA mit einem fixen Algorithmus oder eine DiGA mit einem flexiblen, sich verändernden Algorithmus, der sich weiterentwickelt und sich möglicherweise auch dem Anwender individuell anpasst. Die Anpassungsfähigkeit eines Algorithmus könnte durchaus seine Stärke sein, weshalb eine Evaluation, die diese Charakteristika außer Acht lässt oder künstlich einschränkt, nicht hilfreich wäre. Therapiestrategien mit Anpassungen zu evaluieren ist in der klinischen Forschung keineswegs unüblich (z. B. kann sich die Anwendung eines chirurgischen Verfahrens auch über den Studienverlauf weiterentwickeln, wenn die Operateure lernen und besser mit einem neuen Verfahren vertraut werden [45]). Diese Erfahrung kann auf DiGA übertragen werden und besondere Methoden (inklusive Randomisierung) wurden vorgestellt [45, 46].

**Remote Trials/Virtual Trials.** Kontaktarme Studien (sog. Remote Trials oder Virtual Trials) werden zunehmend erfolgreich durchgeführt [47–49]. Studien ohne Studienzentren sind bei DiGA eher möglich als bei anderen Maßnahmen, auch aufgrund der digitalen Distributionsmöglichkeiten. Breite Einschlusskriterien und nicht notwendige studienspezifische Untersuchungen zur Sicherstellung der Teilnehmersicherheit, die bei DiGA oft eher möglich sind als bei experimentellen Arzneimitteln, fördern zudem eine effiziente und zügige Rekrutierung auch ohne Studienzentren. Die Interaktion mit Teilnehmern erfolgt weitgehend virtuell, ohne regelmäßige

Visiten der Studienzentren [49]. So gelingt eine weitreichende, ortsungebundene Rekrutierung. Die Datenerfassung kann direkt online mit sicheren Systemen durch die Teilnehmer erfolgen, aufwendiges Monitoring würde entfallen. Der oben genannte RCT zur Raucherentwöhnung z. B. rekrutierte Teilnehmer landesweit online über die Website der französischen gesetzlichen Krankenkasse. Endpunkte wurden über Onlinefragebögen erfasst. Jedoch bestehen durchaus Herausforderungen, die sorgfältig untersucht werden müssen (z. B. wenn besondere regulatorische Anforderungen oder manche Untersuchungen direkten Kontakt mit Studienpersonal erfordern; [47–49]). Insgesamt kann dieser Ansatz wegbereitend sein für ein lernendes Evaluationsmodell von DiGA.

**Digitale Biomarker.** Smartphones (und andere mobile Geräte) erlauben eine schnelle und kontinuierliche Messung von Informationen, die Gesundheit reflektieren und Endpunkte von Studien zur Nutzenbewertung sein können. Diese Informationen können weit komplexer und granularer sein als traditionelle Endpunkte, z. B. durch häufigere und zeitlich präzisere Messungen [50]. Sie sind als digitale Biomarker ein vielversprechendes Zukunftsfeld, z. B. um Krampfanfälle oder kardiale Arrhythmien zu erkennen [51], aber könnten möglicherweise auch helfen, Patienten zu identifizieren, die von einer Therapie besonders profitieren. Eine valide Verwendung setzt allerdings eine für alle Studienteilnehmer gleiche Messung (und gleiche Wahrscheinlichkeit für fehlende Werte) voraus. Sie darf in keinem Zusammenhang mit der DiGA stehen. Große Herausforderungen bestehen bei Kontrollen ohne DiGA, während digitale Biomarker für einen direkten Vergleich zweier DiGA-Versionen im Rahmen einer lernenden Evaluation sehr großes Potenzial haben.

## Fazit

Evidenzbasierte Nutzenbewertung hat fundamentale Anforderungen und Maßstäbe unabhängig von der Art der evaluierten Behandlungen. Zuverlässige

Evidenz ist unverzichtbar. Eine schnelle, effiziente, zuverlässige und praxisrelevante Evaluation von DiGA gelingt nicht durch eine Hinwendung zu nichtrandomisierten Studien, sondern vielmehr durch bessere, pragmatische RCTs. Sie sind machbar und verbinden die Charakteristika von DiGA, klassische methodische Konzepte und neue Ansätze in der Studiendurchführung. Routinedaten, kontaktarme Studiendurchführung und digitale Biomarker fördern nützliche randomisierte Real-World-Evidenz als solide Evidenzbasis von DiGA. Besondere Herausforderungen ergeben sich bei der verlässlichen Messung der Endpunkte in oft unverblindeten Studien und der Datenqualität. Insgesamt ist eine kontinuierliche, lernende Evaluation im Versorgungsalltag mit in die Routine eingebetteten, randomisierten Studiendesigns der Schlüssel zu einer nachhaltigen und effizienten Nutzenbewertung von DiGA und kann entscheidend für eine strategische Verbesserung der Gesundheitsversorgung sein.

### Korrespondenzadresse

**PD Dr. med. Lars G. Hemkens, MPH**  
Basel Institute for Clinical Epidemiology and Biostatistics (ceb), Department of Clinical Research, University Hospital Basel  
Spitalstrasse 12, 4031 Basel, Schweiz  
lars.hemkens@usb.ch

**Funding.** Open access funding provided by University of Basel

### Einhaltung ethischer Richtlinien

**Interessenkonflikt.** L.G. Hemkens gibt an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autoren keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

**Open Access.** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten

Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

### Literatur

1. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) (2021) Digitale Gesundheitsanwendungen (DiGA). [https://www.bfarm.de/DE/Medizinprodukte/DVG/\\_node.html](https://www.bfarm.de/DE/Medizinprodukte/DVG/_node.html). Zugegriffen: 20. Apr. 2021
2. EUnetHTA Joint Action 2 WP (2016) HTA Core Model<sup>®</sup> version 3.0
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2020) Allgemeine Methoden Version 6.0 vom 05.11.2020. [https://www.iqwig.de/methoden/allgemeine-methoden\\_version-6-0.pdf](https://www.iqwig.de/methoden/allgemeine-methoden_version-6-0.pdf). Zugegriffen: 20. Apr. 2021
4. Montori VM, Elwyn G, Devereaux PJ, Straus SE, Haynes RB, Guyatt G (2015) Decision making and the patient. In: Guyatt G, Rennie D, Meade MO, Cook DJ (Hrsg) Users' guides to the medical literature: a manual for evidence-based clinical practice, 3. Aufl. McGraw-Hill, New York, NY
5. Danner M, Hummel JM, Volz F et al (2011) Integrating patients' views into health technology assessment: analytic hierarchy process (AHP) as a method to elicit patient preferences. *Int J Technol Assess Health Care* 27:369–375
6. Mühlbacher AC, Juhnke C (2013) Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making? *Appl Health Econ Health Policy* 11:163–180
7. Hernán MA, Robins J (2020) Causal inference: what if. Chapman & Hall/CRC, Boca Raton
8. McCord KA, Al-Shahi Salman R, Treweek S et al (2018) Routinely collected data for randomized trials: promises, barriers, and implications. *Trials* 19:29
9. Collins R, Bowman L, Landray M, Peto R (2020) The magic of randomization versus the myth of real-world evidence. *N Engl J Med* 382:674–678
10. Kyriacou DN, Lewis RJ (2016) Confounding by indication in clinical research. *JAMA* 316:1818–1819
11. Hemkens LG, Ewald H, Naudet F et al (2018) Interpretation of epidemiologic studies very often lacked adequate consideration of confounding. *J Clin Epidemiol* 93:94–102
12. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA (2016) Routinely collected data and comparative effectiveness evidence: promises and limitations. *Cmaj* 188:E158–E164
13. Munkholm K, Faurholt-Jepsen M, Ioannidis JPA, Hemkens LG (2020) Consideration of confounding was suboptimal in the reporting of observational studies in psychiatry: a meta-epidemiological study. *J Clin Epidemiol* 119:75–84
14. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2020) IQWiG-Berichte – Nr. 863: Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGBV
15. Wallach JD, Serghiou S, Chu L et al (2020) Evaluation of confounding in epidemiologic studies assessing alcohol consumption on the risk of ischemic heart disease. *BMC Med Res Methodol* 20:64
16. Chu L, Ioannidis JPA, Egilman AC, Vasiliou V, Ross JS, Wallach JD (2020) Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk. *Int J Epidemiol* 49:608–618
17. Ewald H, Ioannidis JPA, Ladanie A, McCord K, Bucher HC, Hemkens LG (2020) Nonrandomized studies using causal-modeling may give different answers than RCTs: a meta-epidemiological study. *J Clin Epidemiol* 118:29–41
18. Patel CJ, Burford B, Ioannidis JP (2015) Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J Clin Epidemiol* 68:1046–1058
19. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP (2016) Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ* 352:i493
20. Kohavi R, Tang D, Xu Y, Hemkens LG, Ioannidis JPA (2020) Online randomized controlled experiments at scale: lessons and extensions to medicine. *Trials* 21:150
21. Ioannidis JP (2016) Why most clinical research is not useful. *PLoS Med* 13:e1002049
22. Schwartz D, Lellouch J (1967) Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis* 20:637–648
23. Angus DC, Gordon AC, Bauchner H (2021) Emerging lessons from COVID-19 for the US clinical research enterprise. *JAMA* 325:1159–1161
24. Zwarenstein M (2017) 'Pragmatic' and 'explanatory' attitudes to randomised trials. *J R Soc Med* 110:208–218
25. Van Spall HGC, Toren A, Kiss A, Fowler RA (2007) Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA* 297:1233–1240
26. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J (2015) A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials* 16:495
27. Speich B, Logullo P, Deuster S et al (2021) A meta-research study revealed several challenges in obtaining placebos for investigator-initiated drug trials. *J Clin Epidemiol* 131:70–78
28. RECOVERY Trial (2021) Homepage. <https://www.recoverytrial.net/>. Zugegriffen: 28. Apr. 2021
29. Wise J, Coombes R (2020) Covid-19: the inside story of the RECOVERY trial. *BMJ* 370:m2670
30. Hemkens LG (2020) Commentary on Bertagnoli et al: clinical trial designs with routinely collected real-world data-issues of data quality and beyond. *Clin Trials* 17:247–250
31. Kwakkenbos L, Imran M, McCall SJ et al (2021) CONSORT extension for the reporting of randomised controlled trials conducted using cohorts and routinely collected data (CONSORT-ROUTINE): checklist with explanation and elaboration. *BMJ* 373:n857
32. Hemkens LG (2018) How routinely collected data for randomized trials provide long-term randomized real-world evidence. *JAMA Netw Open* 1:e186014–e186014
33. McCord KA, Ewald H, Ladanie A et al (2019) Current use and costs of electronic health records for clinical trial research: a descriptive study. *CMAJ Open* 7:E23–E32



34. Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M (2012) The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clin Trials* 9:436–446
35. Zuidgeest MGP, Goetz I, Groenwold RHH, Irving E, van Thiel G, Grobbee DE (2017) Series: pragmatic trials and real world evidence: paper 1. Introduction. *J Clin Epidemiol* 88:7–13
36. Food and Drug Administration (2021) Statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials guidance for industry. <https://www.fda.gov/media/148473/download>. Zugegriffen: 17. Juni 2021
37. Mitroiu M, Oude Rengerink K, Teerenstra S, Pétavy F, Roes KCB (2020) A narrative review of estimands in drug development and regulatory evaluation: old wine in new barrels? *Trials* 21:671
38. Affret A, Luc A, Baumann C et al (2020) Effectiveness of the e-Tabac info service application for smoking cessation: a pragmatic randomised controlled trial. *BMJ Open* 10:e39515
39. Collister D, Rodrigues JC, Mbuagbaw L et al (2020) Prerandomization run-in periods in randomized controlled trials of chronic diseases: a methodological study. *J Clin Epidemiol* 128:148–156
40. Hernán MA, Hernández-Díaz S (2012) Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 9:48–55
41. Ewald H, Speich B, Ladanie A, Bucher HC, Ioannidis JPA, Hemkens LG (2019) Marginal structural models and other analyses allow multiple estimates of treatment effects in randomized clinical trials: meta-epidemiological analysis. *J Clin Epidemiol* 107:12–26
42. Woodcock J, LaVange LM (2017) Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med* 377:62–70
43. Saville BR, Berry SM (2016) Efficiencies of platform clinical trials: a vision of the future. *Clin Trials* 13:358–366
44. Angelescu K, Sauerland S (2019) Mobile health applications: what evidence is needed? *Dtsch Arztebl* 116:A1057–A1062
45. Ergina PL, Cook JA, Blazeby JM et al (2009) Challenges in evaluating surgical innovation. *Lancet* 374:1097–1104
46. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N (2013) Continuous evaluation of evolving behavioral intervention technologies. *Am J Prev Med* 45:517–523
47. Dolgin E (2018) Industry embraces virtual trial platforms. *Nat Rev Drug Discov* 17:305–306
48. McDermott MM, Newman AB (2021) Remote research and clinical trial integrity during and after the coronavirus pandemic. *JAMA* 325(19):1935–1936. <https://doi.org/10.1001/jama.2021.4609>
49. Cummings SR (2021) Clinical trials without clinical sites. *JAMA Intern Med* 181(5):680–684. <https://doi.org/10.1001/jamainternmed.2020.9223>
50. Califf RM (2018) Biomarker definitions and their applications. *Exp Biol Med* (Maywood) 243:213–221
51. Coravos A, Khozin S, Mandl KD (2019) Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *NPJ Digit Med* 2:14