



Contents lists available at ScienceDirect

Z. Evid. Fortbild. Qual. Gesundh. wesen (ZEFQ)

journal homepage: <http://www.elsevier.com/locate/zefq>

ZEFQ-Service: Leserbriefe

Leserbrief zu „Wie belastbar sind Studien der aktuell dauerhaft aufgenommenen diwigitalen Gesundheitsanwendungen (DiGA)? Methodische Qualität der Studien zum Nachweis positiver Versorgungseffekte von DiGA“ (ZEFQ 175 (2022) 1-16) von P.L. Kolominsky-Rabas et al.)

Arbeitsgruppe Evidenz des Spitzenverbandes Digitale Gesundheitsversorgung e.V. (SVDGV) Karl-Liebknecht-Straße 1, 10178 Berlin, Deutschland

Das Review der Autoren Kolominsky-Rabas und Kollegen hat sich zum Ziel gesetzt, die methodische Qualität der Zulassungsstudien von sechs digitalen Gesundheitsanwendungen (DiGA) im Bereich „Nervensystem und Psyche“ zu beurteilen. Anhand der Ergebnisse der Studienbewertung durch das „Revised Risk-of-Bias Tool“ (RoB 2) ziehen die Autoren die Schlussfolgerung, dass alle sechs Studien ein hohes Verzerrungspotenzial aufweisen und somit DiGA auf der Grundlage von unzureichender wissenschaftlicher Güte dauerhaft in die Versorgung gelangen.

Das Review listet detailliert die Verzerrungspotentiale der einzelnen Studien auf und bietet eine transparente Bewertung der Studien durch übersichtliche Tabellen. Zudem werden die Hintergründe der DiGA-Studien ausführlich erläutert und Lösungsmöglichkeiten für die Verbesserung der Studienqualität vorgeschlagen. Dennoch gibt es Limitationen, welche die Aussagekraft des Reviews einschränken und welche von den Autoren nicht aufgegriffen wurden. Diese fassen wir wie folgt zusammen:

Unvoreingenommenheit der Autoren: In der Einleitung des Artikels schreiben die Autoren: „Ziel dieser Arbeit ist die es, die Studien zum Nachweis positiver Versorgungseffekte der dauerhaft in das DiGA-Verzeichnis aufgenommenen DiGA aus den Kategorien „Nervensystem“ und „Psyche“ auf ihre methodische Qualität hin zu analysieren und zu bewerten.“ (Seite 3). Dies lässt auf eine nicht-gerichtete Zielsetzung schließen. Dem entgegengesetzt steht im Fazit des Artikels (Seite 14): „Die Befürchtung, dass DiGA auf der Grundlage

von Studien mit unzureichender wissenschaftlicher Güte dauerhaft in die Versorgung gelangen, hat sich somit bestätigt.“ Die Abwesenheit einer Referenz legt nahe, dass dies die Annahme der Autoren war und somit eine Vorannahme über die Studienqualität bestand, welche bestätigt wurde.

Unvollständige Methodenbeschreibung: Entgegen der Empfehlung für das Reporting bei der Verwendung des RoB 2 Tools (MECIR C53; Section 7.3.2 Cochrane Handbook) ist nicht berichtet, wer die Bewertung der Studienqualität durchgeführt hat, wie viele Personen beteiligt waren, ob diese unabhängig voneinander bewertet haben und wie vorgegangen wurde, wenn Uneinigkeit bestand. Dies ist relevant, da die Übereinstimmung von Studienbewertungen bei Verwendung des Tools generell niedrig ist, insbesondere bei den Kategorien Verblindung (Domäne 4) ($\kappa < .60$) und „Bias durch Selektion des berichteten Ereignisses“ ($\kappa < .50$) [1,2], welche im dargestellten Artikel überwiegend zu einem hohen Gesamtrisiko der Studien geführt haben (siehe Tabelle 2). In diesem Kontext ist es auch wichtig zu erwähnen, dass zeitgleich zu dem hier besprochenen Review ein weiteres Review mit ähnlicher Zielsetzung im Journal BMC Health Services Research erschienen ist, welches ebenfalls 8 der 9 Studien mit dem RoB 2 Tool bewertet [3]. Im starken Kontrast zum vorliegenden Review wurde bei keiner der 8 überlappenden Studien Domäne 4, Bias durch Ergebnismessung, mit einem hohen Risiko bewertet. Ein weiterer Punkt ist, dass es Ziel des RoB 2 Tools ist, das Verzerrungsrisiko der Studienergebnisse anhand der

verfügbaren Informationen zu bewerten. Wie von den Autoren erwähnt, können hierfür Informationen verwendet werden, die über die Publikation hinausgehen (z.B. Studienregister; Studienprotokolle). Aus Tabelle 3 geht hervor, dass die Berichterstattung nicht immer vollständig war und somit Informationen für die Studienqualitätsbewertung gefehlt haben. Obwohl dies auf eine fehlerhafte Berichterstattung schließen lässt, welche durchaus verbessert werden sollte, bedeutet dies nicht, dass fehlende Informationen nicht eingeholt werden können. Im Gegenteil, es ist durchaus üblich, die Erstautoren oder Leiter*innen der entsprechenden Studie zu kontaktieren und, wo es angebracht ist, fehlende Informationen einzuholen und dies beim methodischen Vorgehen zu berichten.

Fehlende Diskussion der Verwendung des RoB 2 Tools: Die Verwendung des RoB 2 Tools und dessen Limitationen im angewendeten Kontext werden nicht hinterfragt, es wird lediglich erwähnt, dass es auch andere Instrumente gibt und nicht ausgeschlossen werden kann, dass andere Aspekte der Qualitätsbewertung nicht berücksichtigt wurden (Seite 14). Die Autoren haben sich auf DiGAs im Bereich „Nervensystem“ und „Psyche“ fokussiert. Aus den Beschreibungen der Anwendungen (Seite 3-6) geht hervor, dass es sich um Interventionen mit psychotherapeutischen Inhalten handelt, welche, in diesem Fall, mittels randomisiert-kontrollierter Studien auf Wirksamkeit geprüft wurden. Das RoB 2 Tool wurde ursprünglich für den medizinischen Bereich entwickelt und stößt bei psychotherapeutischen Studien auf Probleme, welche bei der

<https://doi.org/10.1016/j.zefq.2023.02.001>

Interpretation der Ergebnisse nicht außer Acht gelassen werden sollten [4]. Bei Verwendung des RoB 2 Tools führt ein hohes Verzerrungsrisiko in einer der 5 Domänen zu einem hohen Gesamtrisiko. Bei 6 der 9 Studien des Reviews geht dieses Gesamtrisiko auf das hoch eingeschätzte Risiko der Domäne 4 zurück (siehe Tabelle 2), welche laut Autoren im Wesentlichen aufgrund fehlender Verblindung und selbstberichteter Endpunkte mit hohem Risiko eingestuft wurde. Zusätzlich zu der oben genannten Übereinstimmungsproblematik bei diesen Items, wird auch deren Bedeutsamkeit bei psychotherapeutischer Forschung diskutiert. Denn, 1.) Im Gegensatz zu pharmakologischen Studien ist die Verblindung von Patient*innen und Therapeut*innen bei Psychotherapie nicht möglich [5] und wird daher, aufgrund der konzeptuellen Schwierigkeiten und der Unumgänglichkeit, von >40% der Meta-Analysen zur Psychotherapie nicht berücksichtigt [4]; und 2.) Ist es gängige Praxis der psychotherapeutischen Forschung, selbstberichtete Endpunkte als äquivalent zu klinischen Beobachtungen zu betrachten, obwohl Patient*innen nicht verblindet sind [6,7]. Die Annahme, welche auch von den Autoren getroffen wird (Seite 7 & 14), dass selbstberichtete Endpunkte bei psychotherapeutischer Forschung zu einer Überschätzung der Behandlungsergebnisse führen, konnte in Metaanalysen nicht bestätigt werden. Im Gegenteil, es wurde gezeigt, dass die Bewertung durch verblindete Betrachtung größere Effekte als Selbstbewertung zeigte [8]. Diese Punkte, obwohl stark in der psychotherapeutischen Forschung diskutiert, werden weder bei der Bewertung der Studienqualität noch bei der Interpretation der Ergebnisse berücksichtigt; tatsächlich wird empfohlen, den Endpunkt in Form einer Fremdbeurteilung zu erheben (Seite 14).

Problematische Lösungsansätze: Zusätzlich zu der oben genannten Empfehlung gehen auch die weiteren Lösungsansätze der Autoren auf Domäne 4 des RoB 2 Tools zurück. Unter dem Titel „Verblindung auch bei DiGA möglich“ wird vorgeschlagen, mögliche Placebo-Effekte durch eine „Sham-Intervention“ zu kontrollieren und somit eine Verblindung zu gewährleisten. Auch dies ist im Kontext psychotherapeutischer Forschung fehlerhaft. 1.)

Eine Verblindung findet auch unter diesen Umständen nicht statt, da die Patienten sich weiterhin bewusst sind, welche Intervention sie machen [9]. Zudem ist auch eine „Sham-Intervention“ eine Intervention und kann somit kein „real placebo“ sein. Wir stimmen den Autoren zu, dass eine „Sham-Intervention“ oder eine aktive Kontrolle die Verzerrung durch die Erwartungshaltung reduzieren können [4], gleichzeitig ist es aber auch wichtig zu bedenken, dass die Erwartungshaltung ein aktives unspezifisches Therapieelement ist, ebenso wie das systematische Beobachten der Symptome über die Zeit [9,10]. Vergleicht man also eine DiGA mit einer „Sham-Intervention“, so testet man die Intervention, je nachdem wie komplex sie gestaltet ist, gegen ihre eigenen unspezifischen oder spezifischen Therapiekomponenten [11]. Ein solches Studiendesign hat definitiv Stärken, eignet sich jedoch nicht, um, wie von den Autoren korrekt in der Methodik beschrieben, zu zeigen, „dass die Anwendung der DiGA besser ist als die Nichtanwendung“ (Seite 3).

Unvollständige Berücksichtigung der Literatur bei Annahmen: Dass Literatur an Stellen einseitig beschrieben wurde, zeigt sich unter anderem auch in der Annahme auf Seite 14: „Grundsätzlich haben vulnerable Gruppen, also beispielsweise ältere Menschen, eine geringere Adhärenz bei digitalen Anwendungen als nicht vulnerable Gruppen“. Dieses Argument wird von einer Quelle gestützt, welche eine narrative Übersichtsarbeit darstellt, die sich zum Ziel gesetzt hat „das Konzept der Adhärenz bei digitalen Interventionen näher zu beleuchten und von verwandten Konzepten abzugrenzen“ (Quelle 26). In dieser Übersichtsarbeit findet keine systematische Datenauswertung zum Thema Adhärenz bei digitalen Anwendungen statt. Tatsächlich zeigen systematische Übersichtsarbeiten z.B., dass von 20 Studien, welche Alter in Bezug auf online Interventionen mit psychologischen Inhalten untersucht haben, 50% keinen Zusammenhang zwischen Alter und Adhärenz gefunden haben. Alle übrigen berichteten inkonsistente Ergebnisse; 5 Studien zeigten, dass ein höheres Alter mit besserer Adhärenz assoziiert war und 4 Studien zeigten, dass ein jüngeres Alter mit besserer Adhärenz verbunden war. Eine weitere

Studie zeigte gemischte Ergebnisse innerhalb der Analysen [12]. Die Annahme der Autoren ist daher in ihrer Aussage nicht korrekt und irreführend.

Literatur

- [1] Armijo-Olivo S, Ospina M, da Costa BR, Egger M, Saltaji H, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9(5): e96920. <https://doi.org/10.1371/journal.pone.0096920>.
- [2] Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review. *J Clin Epidemiol* 2017;81:72–6. <https://doi.org/10.1016/j.jclinepi.2016.08.012>.
- [3] Lantzsch H, Eckhardt H, Campione A, et al. Digital health applications and the fast-track pathway to public health coverage in Germany: challenges and opportunities based on first results. *BMC Health Serv Res* 2022;22:1182. <https://doi.org/10.1186/s12913-022-08500-6>.
- [4] Munder T, Barth J. Cochrane's risk of bias tool in the context of psychotherapy outcome research. *Psychother Res* 2018;28(3):347–55. <https://doi.org/10.1080/10503307.2017.1411628>.
- [5] Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: a meta-analysis of structural equivalence of placebo controls. *J Consult Clin Psychol* 2003;71(6):973. <https://doi.org/10.1037/0022-006X.71.6.973>.
- [6] Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol Med* 2010;40(2):211–23. <https://doi.org/10.1017/S0033291709006114>.
- [7] Leichsenring F, Rabung S. Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis. *J Am Med Assoc* 2008;300(13):1551–65. <https://doi.org/10.1001/jama.300.13.1551>.
- [8] Cuijpers P, Li J, Hofmann SG, Andersson G. Self-reported versus clinician-rated symptoms of depression as outcome measures in psychotherapy research on depression: a meta-analysis. *Clin Psychol Rev* 2010;30(6):768–78. <https://doi.org/10.1016/j.cpr.2010.06.001>.
- [9] Enck P, Zipfel S. Placebo effects in psychotherapy: a framework. *Front Psych* 2019;10:456. <https://doi.org/10.3389/fpsy.2019.00456>.
- [10] Cuijpers P et al. The role of common factors in psychotherapy outcomes. *Annu Rev Clin Psychol* 2019;15:207–31. <https://doi.org/10.1146/annurev-clinpsy-050718-095424>.
- [11] Gold SM, Enck P, Hasselmann H, Friede T, Hegerl U, Mohr DC, Otte C. Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. *Lancet Psychiatry* 2017;4(9):725–32.
- [12] Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. *Int J Behav Med* 2016;23(6):776–94. <https://doi.org/10.1007/s12529-016-9556-9>.