Digital health applications from a government-regulated directory of reimbursable health apps in Germany—a systematic review for evidence and bias

Florian Dittrich[1,2]^, Annabelle Mielitz[1]^, Evgenii Pustozerov[1]^, Dennis Lawin[1,3]^, Ute von Jan[4]^, Urs-Vito Albrecht[1]^

[1]Department of Digital Medicine, Medical Faculty OWL, Bielefeld University, Bielefeld, Germany; [2]Joint Centre Bergisch Land, Remscheid, Germany; [3]Department of Cardiology and Intensive Care Medicine, University Hospital OWL of Bielefeld University, Campus Klinikum Bielefeld, Bielefeld, Germany; [4]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany

Contributions: (I) Conception and design: F Dittrich, U von Jan, UV Albrecht; (II) Administrative support: U von Jan, UV Albrecht; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: A. Mielitz, U von Jan, E Pustozerov, UV Albrecht; (V) Data analysis and interpretation: F Dittrich, A Mielitz, U von Jan, UV Albrecht; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. rer. biol. hum. Ute von Jan. Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany. Email: ute.von.jan@plri.de.

Background: The Digital Healthcare Act, passed in November 2019, authorizes healthcare providers in Germany to prescribe digital health applications (DiGA) to patients covered by statutory health insurance. If DiGA meet specific efficacy requirements, they may be listed in a special directory maintained by the German Federal Institute for Drugs and Medical Devices. Due to the lack of well-founded app evaluation tools, the objectives were to assess (I) the evidence quality situation for DiGA in the literature and (II) how DiGA manufacturers deal with this issue, as reflected by the apps available in the aforementioned directory.

Methods: A systematic review of the literature on DiGA using PubMed, Scopus, and Web of Science was started on February 4, 2023. Papers addressing the evidence for applications listed in the directory were included, while duplicates and mere study protocols not reporting on data were removed. The remaining publications were used to assess the quality of the evidence or potential gaps in this regard. Results were aggregated in tabular form.

Results: The review identified fourteen relevant publications. Six studies suggested inadequate scientific evidence, five mentioned shortcomings of tools for validating DiGA-related evidence, and four publications described a high potential for bias, potentially influencing the validity of the results. Concerns about limited external generalizability were also raised.

Conclusions: The literature review found evidence-related gaps that must be addressed with adequate measures. Our findings can serve as a basis for a plea for a more detailed examination of the quality of evidence in the DiGA context.

Keywords: Health apps; digital health applications; register; reimbursement; digital health applications (DiGA)

^ ORCID: Florian Dittrich, 0000-0002-5135-4736; Annabelle Mielitz: 0009-0008-1823-0900; Evgenii Pustozerov, 0000-0001-8109-1319; Dennis Lawin, 0000-0002-1292-7845; Ute von Jan, 0000-0001-9225-593X; Urs-Vito Albrecht, 0000-0002-8989-6696.

Introduction

Background

Using smartphones for communication, information processing, or data collection for personal or professional purposes has become commonplace (1). Mobile Health (mHealth) is increasingly being utilized in the German healthcare system. This is reflected by the political and regulatory efforts to promote digitalization in the medical field (2). After initial pioneering steps, such as the E-Health Law of 2015 (3), recent legislation has gained momentum toward a comprehensive national digitalization strategy. The "Law for Better Care through Digitalization and Innovation" (Digital Health Care Act, DVG), passed by the German Bundestag on November 7, 2019, and published on December 9, 2019, paved the way for the regulation of apps, improved use of web-based video consultations, and increased data security in the transmission of health data (4). This justifies hope for progress toward evidence-based, technology-supported processes in health care (5).

With the rapid development of a fast-paced app industry, the number of apps available in the app stores has exploded over the last decade. The stores' offerings are complex and poorly regulated, making them diffuse and heterogeneous. The market is so dynamic that the quantity and quality of apps can even vary daily. Comprehensive information about app specifications, essential for safe use in the medical context, is only sporadically provided in the app stores. Inadequate store descriptions, which offer little transparency and only insufficient information about the intended use and limitations of the apps (6,7), as well as data protection, make it difficult to identify credible apps. Finding a safe and high-quality health app in the app stores is like "looking for a needle in a haystack" (8).

Apps listed in the "Health and Fitness" and "Medical" categories often differ in terms of their topic as well as regulation and technology. Many of these apps are not based on standardized content, and there is often little or no safety testing. The threshold for proving the safety and, in particular, the efficacy of apps is correspondingly low. However, it was possible to differentiate a group of 47 apps (as of February 2023), which are Digital Health Applications (acronyms: DiGA in German, DiHA in English). There is a world of difference between unregulated health apps and DiGA. The latter must be medical devices, meet special data protection and security requirements, and demonstrate their benefit. This is because the statutory health insurance scheme reimburses these apps. Therefore, they must fully comply with the requirements for reimbursement under social legislation. After examination by the Federal Institute for Drugs and Medical Devices (BfArM), the manufacturer can apply for inclusion in the publicly accessible DiGA directory (DV) (4,9). In addition to the technical requirements, proving a medical benefit is essential. For permanent inclusion in the directory, the manufacturer must demonstrate a "positive impact on care" (PIC). Suppose the manufacturer does not have evidence of this positive effect on care at the time of application. In that case, provisional inclusion in the DV is possible for twelve or, in some instances, up to 24 months. PIC must be demonstrated no later than by the end of this period. However, this provisional inclusion necessitates submitting a "scientific evaluation concept for demonstrating a PIC created by an independent institution" and the medical services required for the trial (4).

Rationale and knowledge gap

Apart from meeting technical requirements, which are easy to assess, evaluating the medical benefits of DiGA remains difficult. This may be attributed to the lack of well-founded tools for evaluating the medical quality of apps, study designs, and other factors. However, despite all the enthusiasm about introducing prescribable apps and the exciting possibilities they provide for health care in Germany, a comprehensive overview of the factors
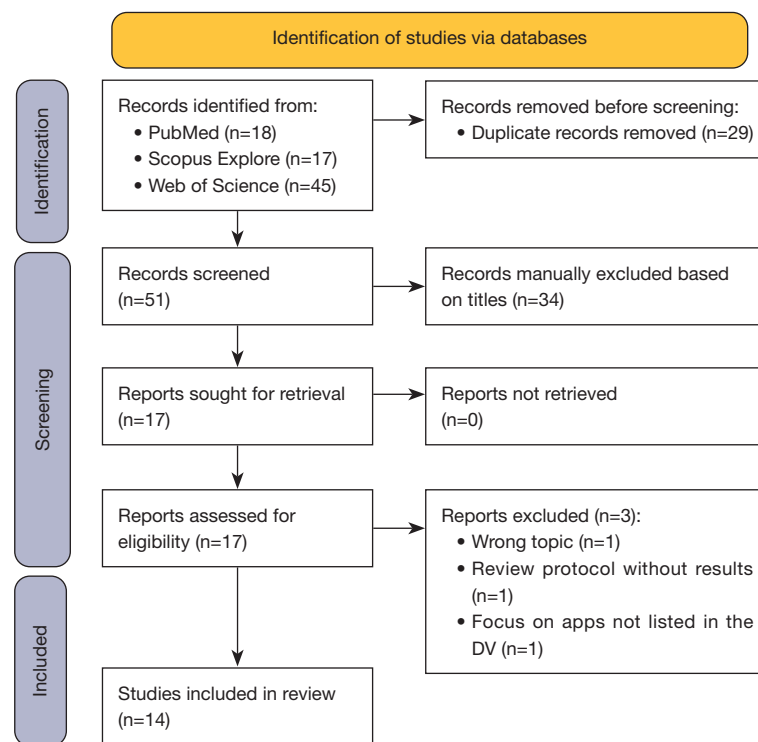

Highlight box

Key findings
- The reviewed literature shows a high potential for bias (high dropout rates, duration of use, inadequate study settings) and limited external generalizability (no/inadequate control groups), potentially affecting the validity of related studies.

What is known and what is new?
- Given the very young field of regulatory studies for digital health applications (DiGA), we conducted a current review of the literature on the evidence quality of these DiGA-related studies. This allowed for the most comprehensive analysis possible regarding a snapshot and correlation with the existing literature results.

What is the implication, and what should change now?
- The all-encompassing guidelines leading to the inclusion of DiGA in the DV need to be more precise and fine-tuned. Appropriate tools for evaluating future digital health applications are urgently needed.

Figure 1 PRISMA 2020 flow diagram (10). DV, DiGA directory.

influencing their relatively slow uptake—as of February 2023, there were only 47 permanently or provisionally listed apps available in the DV—is lacking.

Objective

This paper intends to provide an overview of the current literature on "evidence quality" in the DiGA context. A special focus will be placed on how DiGA manufacturers are dealing with the various factors influencing the generation of the required evidence or on where there are currently gaps that, if adequately addressed, could positively influence future evidence creation. We present this article in accordance with the PRISMA reporting checklist (available at https://mhealth.amegroups.com/article/view/10.21037/mhealth-23-17/rc) (10).

Methods

Online databases (PubMed, Scopus, Web of Science) were systematically searched on February 4, 2023, using English and German keywords "(DiGA OR DiHA) AND Evidence". Studies conducted between 2021 and 2022 were considered. Five reviewers (FD, AM, EP, UVJ, and UVA) independently assessed and agreed on the eligibility of the articles by a systematic screening process (title and abstract) using Ray (Rayyan Systems Inc., Cambridge, USA). For this review, we only considered publications that evaluated the evidence of mHealth concerning the DiGA criteria for apps found in the DV, either positively or negatively. Misidentified studies dealing with a different topic and study protocols not presenting actual data were removed, as were duplicates. For the analysis of the remaining articles (performed by FD, AM, UVJ, and UVA), special attention was given to all mentions of quality criteria in the respective articles, including possible bias and other factors with potential influence on the quality of the reported studies (e.g., related to study design or implementation, the study duration). The articles were then categorized and clustered according to the evidence aspects they covered. All authors participated in the related discussion and agreed on the identified categories. The authors manually compiled the key findings and presented them in tabular and narrative form.

Results

Based on the keyword search, the systematic literature search (*Figure 1*) identified fourteen studies that adequately

Table 1 Included literature and its assignment to the respective cluster

#	Resource	Cluster			
		Adequacy of scientific study designs	Adequacy of evidence validation tools	Potential for bias	External generalizability
01	Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L (15)	√	–	–	–
02	Düvel JA, Gensorowsky D, Hasemann L, Greiner W (16)	√	–	–	–
03	Gerlinger G, Mangiapane N, Sander J (17)	√	–	–	–
04	Gensorowsky D, Witte J, Batram M, Greiner W (18)	√	–	–	–
05	Gensorowsky D, Lampe D, Hasemann L, Düvel J, Greiner W (19)	–	√	–	–
06	Geier AS (20)	–	√	–	–
07	Gregor-Haack J, Busse T, Hagenmeyer EG (21)	–	–	–	√
08	Jeindl R, Wild C (14)	√	√	–	–
09	Kolominsky-Rabas PL, Tauscher M, Gerlach R, Perleth M, Dietzel N (22)	–	–	√	–
10	König IR, Mittermaier M, Sina C, Raspe M, Stais P, Gamstaetter T, *et al.* (23)	√	√	–	–
11	Lantzsch H, Eckhardt H, Campione A, Busse R, Henschke C (24)	–	–	√	√
12	Lubisch B, Hentschel G, Maaß E (25)	√	–	–	–
13	Schliess F, Affini Dicenzo T, Gaus N, Bourez JM, Stegbauer C, Szecsenyi J, *et al.* (26)[†]	–	–	–	–
14	Wolff LL, Rapp M, Mocek A (27)[†]	–	–	–	–

The tick symbol √ indicates that the article in question fits into the respective cluster. [†], the authors did not provide any critical assessment regarding the identified clusters.

addressed the issue of the evidence base for DiGA. Three additional studies were not included because they did not cover the correct topic (11), only described a study protocol but did not provide an associated analysis (12), or focused on apps that were not listed in the DV, and the reasons for their not being listed in the directory (13). Of the fourteen remaining publications, all articles described evidence-related aspects. Four articles evaluated evidence related to specific DiGA listed in the DV. All publications described the general aspects of DiGA, their prerequisites, and related legal aspects. In many cases, other factors relevant when using DiGA, such as risks, opportunities, reimbursement, etc., were also included in the studies. These will, however, only be briefly touched on in this review.

Jeindl and Wild positively noted the need for high-quality, evidence-based mHealth in Germany. They detailed the scientifically based assessment decisions for inclusion in the DV, which could potentially improve healthcare.

On a positive note, the authors emphasized that Germany is pioneering in a situation where numerous apps are available but with little evidence. This relates not only to the curation of possible prescribable mHealth apps that is taking place but also to the requirement for proof of benefit through studies (14). Despite some positive remarks across the articles, criticisms regarding the evidence for DiGA generally prevailed. Overall, we were able to identify four related thematic clusters in the evaluated literature: adequacy of scientific study designs, adequacy of evidence-related validation tools, potential for bias, and external generalizability. The majority of the cited authors agreed on these clusters, and some key findings can be made based on the selected studies (*Table 1*).

Adequacy of scientific study designs

Seven publications (14-18,23,25) suggest that established

concepts and basic evidentiary requirements were abandoned when DiGA were introduced. They refer to the relatively low quality of the DiGA studies submitted by the manufacturers. Specifically, Gerlinger *et al.* and Gensorowsky *et al.* postulated that the quality of studies submitted by manufacturers of DiGA is currently significantly lower than that of studies presented to support the inclusion of other new health services, such as new diagnostic and treatment methods in health insurance coverage (17,18). König *et al.* discussed the methodological study designs using five DiGA as examples. In summary, the scientific data on DiGA is still considered insufficient in the publications, and the studies designed to date are described as addressing only partial aspects of the desired evidence. Studies on the efficacy of DiGA in routine use, e.g., based on billing data, have not yet been published (23). König *et al.* (23) referred to the DiGA report of the Techniker Krankenkasse (28), which supports the results of their study.

Düvel *et al.* agreed (16). In their qualitative study, based on focus group interviews that included various stakeholders (e.g., developers of apps or medicinal products in general, as well as individuals active in the regulatory or care context or those working on assessments and evaluations of new solutions in care), they asked the relevant stakeholders about potentials for reform. These ranged from adapting existing structures to the requirements of digital solutions to introducing an original service category for DiGA. Düvel *et al.* found that the newly created DiGA processes contradict their focus groups' call to preserve existing evidence requirements. In some cases, the new requirements conflict with the traditionally required standards, e.g., related to method evaluation or assessing the medical benefit of medical devices (16).

Lubisch *et al.* (25) emphasized that despite all the indications and statements made by various parties active in the legislative process, scientific proof of the efficacy of DiGA targeting specific diseases (as listed in the DV) is not always adequate. This is especially true when the fast-track procedure is applied.

In these cases, patients might be using DiGA whose effects and side effects can neither be recorded nor adequately judged without professional assistance (25). Jeindl and Wild mentioned that there is actually little evidence for most of the available DiGA (14), and this is mirrored by a majority of healthcare professionals (54.9%) who, as shown by Dahlhausen *et al.* (15), perceive the available evidence to be insufficient as well.

Adequacy of evidence validation tools

A thorough evaluation of DiGA requires a risk-benefit assessment and an evaluation of technology-specific aspects (14). Four studies (14,19,20,23) have noted the lack of standardized tools to adequately assess a DiGA's evidence. König *et al.* (23), for example, state the need for an established methodology for assessing the "body of evidence", as also mentioned in the DiGa report (28), such as the GRADE principles that grade the available data with respect to "Recommendations, Assessment, Development and Evaluation". These are also used in other evaluation contexts in Germany for summarizing the quality of a body of evidence for a technology or intervention (28). Other authors, e.g., Jeindl and Wild, mention the use of various other evaluation methodologies (14), including the NICE (29) standards framework, that, as other tools, covers various aspects such as the level of clinical evidence (generated in scientific studies), usability, security, and compliance with regulatory requirements based on potential risk classes. However, it is lamented that many existing tools were designed and validated with eHealth-based solutions in mind (14). As such, they do not necessarily consider specific requirements of mobile health solutions or DiGA into account. So far, no generally agreed tool or framework for assessing evidence in the DiGA context seems to exist.

Against this background, König *et al.* state that DiGA are considered to be fundamentally innovative. As they belong to a low-risk class as defined by the law, their potential adverse effects are assumed to be limited.

As mentioned in the previous section, appropriate study designs are a major challenge, but also an important building block in developing comprehensive assessment instruments that cover all aspects relevant to the DiGA context. Part of the challenge can be attributed to DiGA being complex interventions whose success also depends on user and prescriber factors (23). DiGA require further accompanying research to address the specific challenges of study design and evidence-generation methods for digital solutions (20). When evaluating the benefit of treatments, randomized controlled trials (RCTs) are generally accepted as the highest level of evidence. Still, RCT-based designs often suffer from low flexibility and incur considerable effort and costs, which makes them sometimes less than ideal for the fast-paced development of DiGA (19). As established in the DVG, the fast-track process addresses this criticism by providing comparative retrospective

study designs with lower levels of evidence as the standard for demonstrating positive health outcomes (19). On a positive note, although the performance of an RCT is not mandatory when evaluating DiGA, Schliess *et al.* reported that more than 90% of manufacturers had used this study design when evaluating DiGA (26).

Following a survey of DiGA manufacturers, Geier noted that, as DiGA are a completely new component of care, comparisons with other therapies in healthcare, such as pharmaceuticals, and thus using evaluation tools that have proven their value in evaluating more conventional approaches in healthcare settings, are not always possible or meaningful (20). On a similar note, Jeindl and Wild (14) analyzed technology and risk-benefit assessment tools to evaluate their applicability in the DiGA context. They found that existing assessment tools often do not comprehensively cover the areas required for a thorough health technology assessment. Decision-makers are therefore faced with new challenges in the assessment of these apps. Jeindl and Wild identified six assessment tools that were current and well-developed at the time of their analysis, which they expected to be applicable to DiGA. An analysis of these six tools showed that four of the six, and thus the majority of the tools, mentioned suggestions for study designs for DiGA (14). Although these assessment tools consider the risk of application errors with clinical consequences, only the NICE Evidence Standards assessment tool (29) proposes a precise classification according to the defined risk classes of the DiGA to be assessed (14).

Overall, the analysis shows that there is currently no adequate assessment tool for DiGA other than the NICE Evidence Standard assessment tool, part of which can be used to assess DiGA. This highlights the need for agreements on an established methodology for evaluating the "body of evidence".

Potential for bias

Of the included articles, two publications (22,24) found a high potential for bias in studies assessing DiGA, which may affect the validity of the results. This is partly due to some studies' high dropout rates and inadequate control groups.

For example, the potential for bias was evident for Kolominsky-Rabas *et al.* in their study of DiGA assigned to the categories of "nervous system" and "psyche". For the six apps that were permanently listed in these categories of the DV at the time of their evaluation, they found high bias potential. One contributing factor was the high dropout rate described in all of the related studies analyzed by Kolominsky-Rabas *et al.* (22), leading to concerns about the scientific quality of the provided evidence for these DiGA. They state that including DiGA in the DV based on scientific evidence is problematic if the studies supposed to prove the evidence are biased (22).

External generalizability

Wolff *et al.* (27) found in their evaluation of six permanently available psychosocial DiGA (which was based on ten DiGA studies) that all RCTs and the meta-analysis met the criteria for external generalizability. Appropriate outcomes and multidimensional assessments were selected for all evidence studies. However, two of the works on DiGA that are included in our review (21,24) criticize studies evaluating DiGA for their limited external generalizability, more precisely for not sufficiently reflecting the situation of use in routine health care in Germany. Gregor-Haack *et al.* criticized the basis of the eleven DiGA listed in March 2021 because the duration of their use in the respective studies was often shorter than the duration recommended by the manufacturers. A further criticism was that the settings selected in the studies, and using no or inadequate control groups to compare the intervention's effect, only reflect the situation of use in routine health care in Germany to a limited extent (21). Lantzsch *et al.* also identified several shortcomings in DiGA studies that might lead to limited external generalizability. They assumed that the quality of reporting in studies is often inadequate. Patient-centered structural and procedural effects are rarely considered, although they may be important in real-world settings. In addition, there is a lack of transparency about whether and to what extent the prices reflect actual benefits. They stated that policymakers and industry should address these shortcomings to pave the way for evidence-based decision-making (24). In summary, these results show that studies that do not belong to the categories "nervous system" and "psyche" may often be confronted with the problem of limited external generalizability when evaluating DiGA.

Discussion

Germany pioneered the widespread use and reimbursement of mHealth in the healthcare sector. But whether this path will ultimately lead to the goal of fundamentally improving the quality of medical care without losing sight of cost-

effectiveness remains an open question. The success or failure of this ambitious healthcare project will be primarily determined by the question of what and how great the medical benefits of medical apps are. Thus, assessing the medical benefit regarding proof of the PIC is crucial for adequately using DiGA. It is not sufficient to evaluate only compliance with technical requirements. Based on this approach, the BfArM has given its quality concept a proper conceptual framework. However, a precondition for this is submitting a scientific concept for evaluation, which is to be compiled by a "manufacturer-independent institution for the proof of the PIC as well as the medical services required for the test" (4).

In this paper, we presented the results of a systematic literature review to provide an overview of the current literature related to "evidence" available for DiGA. The publications selected in the systematic literature review, which already examined the evidence base and the quality of the studies required for inclusion in the DV, almost consistently showed that the level of evidence in the DiGA studies is considered relatively low. It is worth noting that all the articles studied were critical of the topic. Most of the fourteen studies suggest that established "standards of care" (SOC) and concepts were not followed in the DiGA approval studies. Moreover, the reviewed publications showed a high potential for bias (high dropout rates, shorter duration of use in the studies than the duration recommended by the manufacturers, inadequate study settings) and limited external generalizability (no or inadequate control groups), which may affect the validity of the study results.

Reasons for these shortcomings have also been found in the reviewed articles. The lack of proper tools to adequately assess the evidence for medical apps seems to be one of the major problems. The inherently innovative background of DiGA, on the one hand, and the low-risk class according to the law, on the other hand, suggest that the negative effects are limited. In addition, DiGA require further accompanying research to address the design and evidence-generation methods for digital solutions, which are typically complex interventions whose success also includes inherent factors of users and prescribers.

Assessment of the findings in context with a recent DV evaluation

The opinions voiced by the various authors in our literature review were overwhelmingly critical. Due to the timespan selected in the inclusion criteria, this analysis may, however, not adequately reflect the current situation in the DV. Therefore, we compared the findings to our recent work based on data collected from the German DV (30). For the fifteen apps permanently included in the German DV at the time of the data acquisition in December 2022, we could identify an Agency for Healthcare Research and Quality (AHRQ) evidence level (31,32) of at least Ib ("At least one sufficiently large, methodologically high-quality RCT"), in one case even Ia ("At least one meta-analysis based on methodologically high-quality randomized controlled trials"). The NICE category (29) was 3b (i.e., the apps had therapeutic purposes and provided "... treatment for a diagnosed condition or guides treatment decisions") in all cases, which requires high-quality RCTs conducted in a setting relevant to the UK health and social care sector. However, similar to various authors in our literature review, several methodological limitations of the RCTs included in the DV were identified, which may reduce the quality of the evidence.

First, the number of participants in the DV-RCTs was often lower than the number of people typically enrolled in clinical trials [median: 215; interquartile range (IQR): 141], mirroring the results of the literature review (22,24) and indicating potentially questionable validity. Dropout rates should also be considered with caution. Often, the rates in the intervention groups (median: 26%; IQR: 10%) exceeded those in the control groups (median: 13%; IQR: 15%), which may point towards problems with the internal validity of the trials. Possible systematic errors in study design, patient recruitment, or study management may explain the significant dropout rates. Again, these results align with Kolominsky-Rabas *et al.*'s findings (22). The definition of the standards of care (SOC) used by the investigators was often unclear, making it difficult to assess the benefits of an intervention concerning SOC adequately. This aspect was also mentioned by Lubisch *et al.*, who were critical of the use of medical devices for up to 24 months without sound evidence of their medical benefit or even adequately identifying their risks (25).

Overall, the endpoints of the trials were often not well aligned with the goals of the intervention. There was frequently no information on whether or not those randomized to the control groups had received prior treatment for their condition. In the literature review, Gregor-Haack *et al.* also mentioned this problem (21). Another source of bias was the lack of blinding of the intervention: in only three cases could we identify positive information about blinding (30).

The reported study and follow-up periods we found in (30) were also problematic: it was not possible to adequately assess the external validity of studies evaluating only short-term use (median: 3 months; IQR: 1 month) with short follow-up periods (median: 6 months; IQR: 6 months). Again, this shortcoming was mentioned by Gregor-Haack *et al.* as well (21). Longer-term evaluations, conducted under everyday conditions, e.g., in the form of postmarket studies, might be one possibility to address this problem (30).

Finally, information on actual peer reviews was not consistently found. In three cases, the studies mentioned in the DV were only listed on the manufacturer's homepage, which raises the question of quality control for these studies (30).

Overall, the literature review results are consistent with our analysis of studies related to apps currently listed in the DiGA registry. No significant discrepancies were found between the literature findings and our work regarding study quality or possible improvements in the quality of evidence. Therefore, the analysis of the DV-RCTs confirms the critical stance of the publications in the literature review.

Strengths and limitations

Limitations of our review are that due to the extremely young research field in the area of evidence from DiGA registry studies, only a limited number of fourteen references were identified in the literature review, with much of the information in the included articles being collected relatively soon after the introduction of DiGA in late 2019/early 2020 (4). To confirm and strengthen the results of the review and to address this unavoidable limitation, the previous section compared the findings from the review to a comprehensive and more recent analysis of pivotal studies for the DiGA permanently listed in the DV (30), confirming that the concerns raised in the review are still valid.

However, because only the most important points in the literature review were analyzed, a risk of possible bias could be that a possibly insufficient selection of search terms may have resulted in certain publications not being included in our analysis, causing us to overlook potentially important aspects. Statistical analyses were not performed at this point. Additionally, the included publications followed different approaches in their analysis, which can be interpreted both as a strength, allowing for a more comprehensive overall picture of the field of DiGA-related evidence, as well

as a limitation, due to only a limited number of authors reporting on certain aspects in this context.

Implications and actions needed

From the literature reviewed, we can conclude that the innovative process leading to the inclusion of DiGA in the DV could be revised, fine-tuned, and improved. Shortcomings in the study designs lead to questionable external generalizability. These could be addressed, for example, by evaluating and considering longer-term use in post-market studies in real-world settings after the initial introduction of the respective apps. It is in the interest of patients that DiGAs whose effects have been measured under study conditions also demonstrate their added value in routine care. Therefore, attention should be paid to creating realistic and ideally even standardized study conditions to make it easier to determine the benefits of DiGA *vs.* SOC in the future. Similarly, the overarching guidelines for DiGA should be clarified. It is critical that the current process results in reimbursement and clinical use of products for which available evidence is often not of the highest quality. Including patient perspectives and experiences in the evaluation process of DiGA is crucial. Collecting patient-reported outcomes can provide valuable insights into the usability, acceptability, and effectiveness of these DiGA in real-world scenarios.

Reasons for the low evidence base may be the lack of adequate evaluation tools that respect the specificities of DiGA and evaluate them as a unit of technology, content, and basic concept. However, there is a need for evaluation methods that can fully assess all facets of such comprehensive health technology. For a holistic approach, quality principles should be considered that do not only include medical evidence (33,34).

The economic pressure and the requirement of the BfArM to prove the PIC within one year might be one reason why DiGA manufacturers have to use study designs that hardly reflect reality. Balancing regulatory imperatives and the need for consistent and reliable evidence generation is important. This is reflected in the study weaknesses mentioned above and could also be defined as a BfArM-induced bias. However, a potential risk for patients, namely letting them use apps that are not medically safe, must be prevented.

Conclusions

The literature review showed that the quality of most of

the available studies for inclusion in the DV is considered relatively low. Further studies of higher quality are needed to improve the evidence base concerning the benefits of prescribable apps. Evidence gaps should be urgently addressed. This applies to Germany and other countries that would like to implement similar policies for health apps in the future or have already done so (35).

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at https://mhealth.amegroups.com/article/view/10.21037/mhealth-23-17/rc

Peer Review File: Available at https://mhealth.amegroups.com/article/view/10.21037/mhealth-23-17/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://mhealth.amegroups.com/article/view/10.21037/mhealth-23-17/coif). UVA serves as an unpaid editorial board member of *mHealth* from August 2018 to February 2025. DL received research funding from Qompium Inc., Hasselt, Belgium, for another study unrelated to this manuscript. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015;3:e101.

2. Brönneke JB, Debatin JF, Hagen J, et al. DiGA VADEMECUM: Was man zu Digitalen Gesundheitsanwendungen wissen muss. Berlin: MWV Medizinisch Wissenschaftliche Verlagsgesellschaft; 2020: 195.

3. Gröhe H. E-Health-Gesetz verabschiedet. Bundesgesundheitsministerium. 2015 [cited 2023 Mar 26]. Available online: https://www.bundesgesundheitsministerium.de/ministerium/meldungen/2015/dezember-2015/e-health.html

4. Bundestag D. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz - DVG). Bundesgesetzblatt 2019;Teil 1:2562-84. Available online: http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl119s2562.pdf

5. Kuhn B, Amelung V. Kapitel 4. Gesundheits-Apps und besondere Herausforderungen. In: Albrecht UV, editor. Chancen und Risiken von Gesundheits-Apps (CHARISMHA), english Chances and Risks of Mobile Health Apps (CHARISMHA). Hannover: Medizinische Hochschule Hannover; 2016: 100-14. Available online: http://nbn-resolving.de/urn:nbn:de:gbv:084-16040811263

6. Albrecht UV, Malinka C, Long S, et al. Quality Principles of App Description Texts and Their Significance in Deciding to Use Health Apps as Assessed by Medical Students: Survey Study. JMIR Mhealth Uhealth 2019;7:e13375.

7. Albrecht UV, Framke T, von Jan U. Quality Awareness and Its Influence on the Evaluation of App Meta-Information by Physicians: Validation Study. JMIR mHealth and uHealth 2019;7:e16442. Available online: https://mhealth.jmir.org/2019/11/e16442

8. Dittrich F, Beck S, Harren AK, et al. Analysis of Secure Apps for Daily Clinical Use by German Orthopedic Surgeons: Searching for the "Needle in a Haystack". JMIR Mhealth Uhealth 2020;8:e17085. Erratum in: JMIR Mhealth Uhealth 2020;8:e21600.

9. BfArM. Diga Leitfaden. Das Fast Track Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V, S. 98 und 130. Bundesinstitut für Arzneimittel und Medizinprodukte; 2022. Available online: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.pdf

10. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

11. de Araújo CDSR, Machado BA, Fior T, et al. Association of positive direct antiglobulin test (DAT) with nonreactive eluate and drug-induced immune hemolytic anemia (DIHA). Transfus Apher Sci 2021;60:103015.

12. Giebel GD, Speckemeier C, Abels C, et al. Problems and Barriers Related to the Use of Digital Health Applications: Protocol for a Scoping Review. JMIR Res Protoc 2022;11:e32702.

13. Lauer W, Löbker W, Höfgen B. Digital health applications (DiGA): assessment of reimbursability by means of the "DiGA Fast Track" procedure at the Federal Institute for Drugs and Medical Devices (BfArM). Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021;64:1232-40.

14. Jeindl R, Wild C. Technology assessment of digital health applications for reimbursement decisions. Wien Med Wochenschr 2021. Available online: http://dx.doi.org/10.1007/s10354-021-00881-3

15. Dahlhausen F, Zinner M, Bieske L, et al. Physicians' Attitudes Toward Prescribable mHealth Apps and Implications for Adoption in Germany: Mixed Methods Study. JMIR Mhealth Uhealth 2021;9:e33012.

16. Düvel JA, Gensorowsky D, Hasemann L, et al. Digital Health Applications: A Qualitative Study of Approaches to Improve Access to Statutory Health Insurance. Gesundheitswesen 2022;84:64-74.

17. Gerlinger G, Mangiapane N, Sander J. Digital health applications (DiGA) in medical and psychotherapeutic care. Opportunities and challenges from the perspective of the healthcare providers. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021;64:1213-9.

18. Gensorowsky D, Witte J, Batram M, et al. Market access and value-based pricing of digital health applications in Germany. Cost Eff Resour Alloc 2022;20:25.

19. Gensorowsky D, Lampe D, Hasemann L, et al. "Alternative study designs" for the evaluation of digital health applications - a real alternative? Z Evid Fortbild Qual Gesundhwes 2021;161:33-41.

20. Geier AS. Digital health applications (DiGA) on the road to success-the perspective of the German Digital Healthcare Association. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021;64:1228-31.

21. Gregor-Haack J, Busse T, Hagenmeyer EG. The new approval process for the reimbursement of digital health applications (DiGA) from the perspective of the German statutory health insurance. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021;64:1220-7.

22. Kolominsky-Rabas PL, Tauscher M, Gerlach R, et al. How robust are studies of currently permanently included digital health applications (DiGA)? Methodological quality of studies demonstrating positive health care effects of DiGA. Z Evid Fortbild Qual Gesundhwes 2022;175:1-16. Erratum in: Z Evid Fortbild Qual Gesundhwes 2023;176:97.

23. König IR, Mittermaier M, Sina C, et al. Evidence of positive care effects by digital health apps-methodological challenges and approaches. Inn Med (Heidelb) 2022;63:1298-306.

24. 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.

25. Lubisch B, Hentschel G, Maaß E. Digitale Gesundheitsanwendungen (DiGA) – Neue psychotherapeutische Wunderwelt? Psychotherapie aktuell 2020;(4):7-9. Available online: https://www.dptv.de/fileadmin/Redaktion/Bilder_und_Dokumente/Wissensdatenbank_oeffentlich/Psychotherapie_Aktuell/2020/DiGA_Psychotherapie_Aktuell_4.2020.pdf

26. Schliess F, Affini Dicenzo T, Gaus N, et al. The German Fast Track Toward Reimbursement of Digital Health Applications (DiGA): Opportunities and Challenges for Manufacturers, Healthcare Providers, and People With Diabetes. J Diabetes Sci Technol 2022. [Epub ahead of print]. doi: 10.1177/19322968221121660.

27. Wolff LL, Rapp M, Mocek A. Critical Evaluation of Permanently Listed Psychosocial Digital Health Applications into the Directory for Reimbursable Digital Health Applications of the BfArM. Psychiatr Prax. 2022 Sep 28. Available online: http://dx.doi.org/10.1055/a-1875-3635

28. Die Techniker. DiGA-Report 2022. 2022. Available online: https://www.tk.de/resource/blob/2125136/dd3d3dbafcfaef0984dcf8576b1d7713/tk-diga-report-2022-data.pdf

29. National Institute for Health and Care Excellence. Evidence Standards Framework for Digital Health Technologies. NHS England; 2019. Available online: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf

30. Albrecht UV, von Jan U, Lawin D, et al. Evidence of Digital Health Applications from a State-Regulated Repository for Reimbursable Health Applications in Germany. Stud Health Technol Inform 2023;302:423-7.

31. Steiner S, Lauterbach KW. Evidenzbasierte Methodik in der Leitlinienentwicklung. Integration von externer

Evidenz und klinischer Expertise [Development of evidence-based clinical practice guidelines; a model project integrating external evidence and clinical expertise]. Med Klin (Munich) 1999;94:643-7.

32. Berkman ND, Lohr KN, Ansari M, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. 2013 Nov 18. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.

33. Albrecht UV. Einheitlicher Kriterienkatalog zur Selbstdeklaration der Qualität von Gesundheits-Apps.

eHealth Suisse; 2019. Available online: http://dx.doi.org/10.26068/mhhrpm/20190416-004

34. Albrecht UV. Selbsteinschätzung der Wichtigkeit von Qualitätsprinzipen und Recherchehilfe zum Auffinden relevanter Information zu Gesundheits-Apps als Grundlage für eine fundierte Nutzungsentscheidung. Medizinische Hochschule Hannover: Bibliothek; 2020. Available online: https://mhh-publikationsserver.gbv.de/receive/mhh_mods_00001103

35. Essén A, Stern AD, Haase CB, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022;5:31.