

# Evaluation criteria for health apps supporting medication adherence in early-stage technology development – a scoping review

## Bewertungskriterien für Gesundheits-Apps zur Unterstützung der Medikamenteneinnahme in frühen Entwicklungsphasen der Technologie – ein Scoping Review

### Abstract

**Introduction:** Health apps offer an approach to improve the patients' management of their medication. Although the Digital Healthcare Act (DVG) has created a claim in the statutory health insurance (SHI), the large number of health apps available and their varying quality make it difficult for service providers and especially for medical laypersons to select an adequate high-quality medication app. Manufacturers need guidance for the development of high-quality apps right from the start. Various general evaluation concepts for health apps have been available to date. However, the requirements that should be met by healthcare depend largely on the field of application and the type of apps. This article aims to provide an overview of the international evidence on specific criteria for the evaluation of medication apps.

**Methods:** Within the framework of a scoping review, a systematic search was conducted in PubMed and EMBASE on January 29, 2020. The search was limited to publications from 2007 onwards as well as to English and German articles. Additionally, a semi-systematic research of reference lists of the previously included articles as well as a structured search of websites of relevant stakeholders were conducted. Inclusion criteria were the following: the publication deals with health apps that can be used on smartphones and focus on supporting medication intake; the publication does not refer to evaluation criteria for a single app exclusively. The included publications were examined in a qualitative content analysis searching for evaluation criteria and categorizing them according to the framework criteria of the DVG and the Digital Health Applications Ordinance (DiGAV).

**Results:** 2,542 articles were identified in the systematic search (999 in PubMed, 1,543 in EMBASE, 560 duplicates). A total of 16 studies met the inclusion criteria. The semi-systematic research and the structured search identified one further study. A catalog of criteria was developed based on the included 17 studies. This catalog covers the general topics "patient orientation" (data protection and security, consumer protection, user friendliness) and "quality/core functions of medication apps" (reminder, self-monitoring, (drug) information, motivation to change behavior, drug/patient safety, robustness) as well as "interoperability/cooperation". Due to its specific importance for medication apps, the subcategory "motivation for behavioral change" stands out beneath the general topic "quality/core functions of medication apps". This category aims to evaluate the design of individual functions with regard to their potential to actually change the behavior of app users.

**Discussion:** The criteria for the evaluation of health apps mentioned in the DiGAV intersected with the criteria identified in the literature research. However, the area of positive health care effects was hardly covered by the included studies. In the development of the criteria

Anja Niemann<sup>1</sup>  
Theresa Hüer<sup>1</sup>  
Anja Neumann<sup>1</sup>  
Jürgen Wasem<sup>1</sup>  
Petra Schnell-Inderst<sup>2</sup>  
Silke Neusser<sup>1</sup>

1 Institute for Health Care Management and Research, University Duisburg-Essen, Essen, Germany

2 Institute of Public Health, Medical Decision Making and Health Technology Assessment, Department of Public Health, Health Services Research and Health Technology Assessment, UMIT TIROL – University for Health Sciences and Technology, Hall i.T., Austria

catalog, it was not possible to weight the identified criteria. Therefore, the catalog should be understood as a supporting checklist for service providers, manufacturers, and/or users.

**Conclusions:** A large variety of possible evaluation criteria for medication apps could be shown. Future research should focus on the possibilities of weighting these diverse evaluation criteria, using not only clinical studies but also methods to identify preferences.

**Keywords:** mobile application, mHealth, health app, drug therapy, systematic review, assessment, criteria

## Zusammenfassung

**Einführung:** Gesundheits-Apps bieten einen möglichen Lösungsansatz zur Verbesserung des Medikamentenmanagements durch Patienten. Zwar wurde mit dem Digitale-Versorgung-Gesetz (DVG) ein Anspruch in der Gesetzlichen Krankenversicherung (GKV) geschaffen, aber die Vielzahl angebotener Gesundheits-Apps mit schwankender Qualität machen es dem Leistungserbringer und vor allem dem medizinischen Laien schwer, eine bedarfsgerechte und qualitativ hochwertige Medikations-App auszuwählen. Auch Hersteller benötigen für die Entwicklung von qualitativ hochwertigen Apps von Anfang an Orientierungshilfen. Bislang liegen diverse allgemeine Bewertungskonzepte für Gesundheits-Apps vor. Welche Anforderungen Gesundheits-Apps erfüllen sollten, hängt jedoch maßgeblich von der Einsatzsituation beziehungsweise dem Typ der App ab. Daher ist die Zielsetzung des vorliegenden Beitrags, einen Überblick über die internationale Evidenzlage zu spezifischen Kriterien zur Bewertung von Medikations-Apps zu schaffen.

**Methoden:** Im Rahmen eines Scoping Reviews wurde eine systematische Recherche am 29. Januar 2020 in PubMed und EMBASE durchgeführt. Die Suche wurde auf Publikationen ab dem Jahr 2007 und auf englische sowie deutsche Artikel beschränkt. Ergänzt wurde diese durch eine teilsystematische Recherche in den Referenzlisten der zuvor eingeschlossenen Artikel sowie auf Internetseiten von relevanten Stakeholdern. Einschlusskriterien waren: Die Publikation befasst sich mit Gesundheits-Apps, die auf dem Smartphone genutzt werden können und ihren Fokus auf die Unterstützung der Medikamenteneinnahme legen. Zudem durfte die Publikation nicht ausschließlich auf Bewertungskriterien für eine einzelne App bezogen sein. Die identifizierten Publikationen wurden in einer qualitativen Inhaltsanalyse nach Bewertungskriterien untersucht und diese in Anlehnung an die Rahmenkriterien des DVG bzw. der Digitale-Gesundheitsanwendungen-Verordnung (DiGAV) kategorisiert.

**Ergebnisse:** Die systematische Recherche ergab 2.542 Treffer (999 in PubMed, 1.543 in EMBASE, 560 Dubletten). Insgesamt erfüllten 16 Studien die Einschlusskriterien. Die teilsystematische Recherche identifizierte eine weitere Studie. Der auf Basis der 17 Studien entwickelte Kriterienkatalog umfasst die Oberthemen „Patientenorientierung“ (Datenschutz und -sicherheit, Verbraucherschutz, Nutzerfreundlichkeit) und „Qualität/Kernfunktionen von Medikations-Apps“ (Erinnerung, Selbstüberwachung, (Arzneimittel-)Informationen, Motivation zur Verhaltensänderung, Arzneimittel-/Patientensicherheit, Robustheit) sowie den Bereich „Interoperabilität/Zusammenarbeit“. Aufgrund ihrer spezifischen Bedeutung für Medikations-Apps sticht unter dem Oberthema „Qualität/Kernfunktionen der Medikations-Apps“ die Unterkategorie „Motivation zur Verhaltensänderung“ besonders hervor. Diese zielt auf die Bewertung der Ausgestaltung einzelner Funktionen hinsichtlich ihres Potenzials, das Verhalten der App-Nutzer tatsächlich zu verändern, ab.

**Diskussion:** Die in der DiGAV genannten Kriterien zur Bewertung von Gesundheits-Apps wiesen eine große Schnittmenge mit den in der Literaturrecherche identifizierten Kriterien auf. Allerdings wurde der Bereich

der positiven Versorgungseffekte mit den eingeschlossenen Studien kaum erfasst. Im Rahmen der Entwicklung des Kriterienkatalogs konnte keine Gewichtung der identifizierten Kriterien vorgenommen werden. Daher sollte der Katalog von Leistungserbringern, Herstellern und/oder Nutzern als eine unterstützende Checkliste verstanden werden.

**Fazit:** Es konnte eine große Vielfalt möglicher Bewertungskriterien für Medikations-Apps gezeigt werden. Künftige Forschung sollte den Fokus auf Möglichkeiten der Gewichtung dieser vielfältigen Bewertungskriterien legen und dafür neben klinischen Studien auch Methoden der Präferenzhebung einsetzen.

**Schlüsselwörter:** mobile Anwendung, mHealth, Gesundheits-App, Arzneimitteltherapie, systematische Übersicht, Bewertung, Kriterien

## Background

Insufficient adherence is a widespread problem in drug therapy. The WHO defines adherence as follows: “[Adherence is] the extent to which a person’s behavior – taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider” [1]. Studies reveal different, but alarming results. These depend on the indication and patient population considered, as well as on the definition used (exclusively non-adherence or additionally partial adherence). The World Health Organization (WHO) [1] reports that about 50% of medications are not taken as prescribed. Schäfer [2] refers to studies according to which about 20% of all patients do not redeem their prescriptions at all, and half of the remaining 80% of prescriptions are not used at all or not used as prescribed. The consequences of inadequate adherence are complex and affect the health of the patient through medical complications, treatment failure and restrictions on quality of life, but also health care costs and the national economy as a whole. These economic consequences mainly result from increased hospitalization and nursing home admissions, additional costs for therapeutic measures or drugs, and work absenteeism [2]. For example, one in four hospital admissions is reported to be directly or indirectly related to incorrect medication use [2].

The WHO [1] describes forgetting to take the correct dosage as the most common reason for non-adherence. However, numerous other factors also play a role in the correct intake of medications. These include the timing of intake or even foods that are ingested in combination and could trigger interactions. In addition, with the aging population, the proportion of multimorbid individuals taking multiple medications simultaneously is also growing [3]. For example, a U.S. study identifies polypharmacy in 39% for the age group over 65 years. According to the most commonly used definition, polypharmacy is defined as the daily use of five or more medications [4]. However, a significant increase in polypharmacy can also be observed in persons 20 to 39 years old [4]. Adherence in drug therapy can thus become a complex challenge for the individual patient.

One possible approach to improving medication management by patients is offered by health apps to support medication intake (hereinafter medication apps). The smartphone is ubiquitous among many people and can thus be reliably integrated into everyday life. Apps for documenting and reminding patients to take their medications are used increasingly. According to a 2017 survey of 18- to 64-year-olds in Germany conducted by the opinion research company Statista [5], 12% have used a health app within the last twelve months. Specific medication apps offer features such as medication reminders, tracking of doses taken or skipped, an overview of current medication, warnings of drug interactions, medication refill reminders, medication information, and sharing of medication data with third parties [6]. There is evidence that medication apps can improve medication adherence among users. A systematic review included a total of 17 comparative studies examining the effect of medication apps on medication adherence. In twelve of these 17 studies, a significant improvement in adherence was shown for at least one adherence measurement method [7]. With the entry into force of the Digital Health Care Act (DVG) [8] on December 19, 2019, the entitlement of insured persons in the Statutory Health Insurance (SHI) to health apps was anchored in the German Social Code (SGB) V. The health apps must either be prescribed by a contract doctor or psychotherapist or be approved by the health insurer (§§ 33a, 139e SGB V). Due to the high attractiveness of the healthcare market and the reimbursement option recently created by the DVG in the SHI system, numerous providers of health apps can be expected to enter the SHI market. A large number of health apps are already available in the app stores, and this number is expected to grow. The quality of these apps varies [9]. This confusing market already makes it difficult for the medical layperson and will also confront the contract physician with the challenge of selecting a suitable application for his or her needs or those of the app user [9]. The DVG addresses this issue by limiting the claim to health apps that will be included in a directory to be maintained by the Federal Institute for Drugs and Medical Devices (BfArM) after comprehensive review. The Digital Health Applications Ordinance (DiGAV) [10] to the DVG, which entered into force on April 21, 2020, specifies

the requirements for digital health apps to be included in the directory according to Section 139e SGB V. A guideline on this procedure prepared by the BfArM [11] includes precise explanations for manufacturers, service providers and users.

Manufacturers need guidance for the development of high-quality apps right from the start [12]. Currently, there is a variety of general evaluation criteria and concepts that manufacturers can use. For example, in October 2019, the Quality Criteria Core Set for Digital Health Apps (DiGA) “AppQ” was published by the Bertelsmann Foundation [13] with the participation of Fraunhofer FOKUS. The development was funded by the German Federal Ministry of Health. This core set consists of 24 criteria organized into the nine topics of *medical quality, positive care effects, data protection, information security, technical quality, consumer protection and fairness, interoperability, user friendliness and motivation, as well as the connection to the healthcare system*. The criteria are applied via a web application to collect self-reported data from providers of DiGA certified as a medical device. The starting point for the development of the “AppQ” core set of quality criteria was the “APPKRI” meta-criteria catalog for the description and evaluation of health apps previously developed by Fraunhofer FOKUS [14], the development of which was also funded by the German Federal Ministry of Health. In addition, there are various international efforts to create transparency in the field of health apps. In some countries, for example, public health institutions offer platforms with a selection of quality-checked health apps. One example is the NHS Digital Apps Library of the National Health Service in the United Kingdom [15], which was published in 2017 and provides users and care providers with a selection of trusted health apps on different topics. A five-step review process assesses requirements in the nine areas of clinical effectiveness (e.g., evidence-based, clarity of purpose and use of the app), regulatory compliance (e.g., declaration as a medical device), clinical safety (e.g., potential adverse effects of the app), privacy and confidentiality (e.g., privacy statement, consent), (IT) security (e.g., data storage, network communication), usability and accessibility, interoperability (e.g., compatibility with other systems), technical stability (e.g., defined plan for product development), and change management (e.g., version control).

However, the requirements that health apps should meet depend largely on the objective of the app (for example, diagnostic support, companionship for various somatic and psychiatric conditions, documentation of health data, support for a healthy lifestyle). For example, medication apps with the goal of promoting adherence focus on changing patient behavior. In order to create a meaningful evaluation basis (in terms of criteria and instruments) for medication apps, systematic as well as supplementary semi-systematic research on the international evidence base has been conducted as part of a scoping review.

## Research questions

The aim of the present study is to examine the extent to which the legal requirements in Germany are consistent with the results on evaluation criteria retrieved from international literature. However, the evaluation criteria to be identified for medication apps should not be considered in isolation, but should always be classified against the background of the interests of users, manufacturers, and healthcare providers. Finally, the extent to which the identified criteria cover relevant areas of the DiGAV will be considered.

## Methodology

This article was based on a two-stage approach. In the first phase, studies were identified that addressed evaluation criteria for medication apps. The second phase involved the development of a criteria catalog and the assignment of the previously identified criteria to it.

In the first phase, studies were sought that focused on the criteria themselves. For this purpose, the scoping review method was used, which aims to provide an overview of the currently available evidence in the literature, without assessment of the quality of the publications or restrictions regarding the type of study. It can also identify gaps in the evidence [16]. First, a systematic literature search was performed in the PubMed and EMBASE databases on January 29, 2020. Search terms on the top-level topics of health apps, medication adherence, and evaluation criteria were linked using the Boolean operator AND. In addition, the search was limited to results from 2007 onwards, the release year of the first iPhone [17], and to English and German articles. Conference transcripts, letters, and notes were excluded. In the next step, duplicates were removed. Subsequent screening of titles and abstracts, as well as subsequent full-text screening was performed independently by two persons. Discrepancies were resolved by discussion. Publications were included if the following criteria were met: the publication addresses health apps that can be used on smartphones and focus on medication adherence support. In addition, the publication had to consider evaluation criteria and tools for medication apps in general and could not be exclusively related to a specific app. Publications were excluded based on the following criteria: abstract or full text were not available; the health apps studied were primarily used for diagnosis, research, or information, aimed at healthcare providers, limited to SMS exchange, or did not represent standalone software. The search terms are detailed in Appendix 1 (Attachment 1) and the inclusion and exclusion criteria are detailed in Appendix 2 (Attachment 1). In addition, a supplementary semi-systematic search was conducted as part of the scoping review. Reference lists of articles included in the systematic search were screened for relevant publications. In addition, websites of stakeholders, asso-

ciations and public institutions were searched for relevant publications (see Attachment 1, Appendix 3).

The starting point for the second phase was the evaluation criteria for medication apps identified in the first phase, which until then still existed in a loose unsystematic collection. The central added value of this study lies in the development of a catalog of criteria and the assignment of the previously identified individual criteria to this catalog. The starting point for this was the analysis of the DVG [8]. This defines the requirements for digital health applications, which also include health apps, for inclusion in the directory maintained by the BfArM in accordance with § 139e SGB V. These include requirements for safety, functional suitability, quality, data protection and data security, but also the existence of positive effects on care. The details of these requirements are specified in the DiGAV [10]. Two annexes could be taken from this regulation, in which very detailed question catalogs specify the requirements or criteria. These legally defined criteria for the evaluation of health apps were incorporated into the structure of the criteria catalog. The structure derived from the law was adapted and supplemented against the background of the specific medication reference of this study. Only categories for which evaluation criteria could be identified in the literature search were included in the criteria catalog. Categories that were omitted from the criteria catalog with respect to the legally derived structures are explained in more detail in the discussion section. The additions to the structure of the criteria catalog to the categories derived by law resulted from conspicuously frequently mentioned criteria of the included studies that could not be classified in the legal grid. The final criteria catalog was consented by two authors of this paper. In a next step, the included studies were analyzed within the framework of a coding system, and their individual criteria were assigned to the developed criteria catalog. The detailed procedure can be found in Appendix 4 (Attachment 2). This linkage of current legislation with the criteria and categories identified in the scientific literature ascribes relevance to health policy.

## Results

As part of the scoping review, the systematic search identified 999 hits in PubMed and 1,543 hits in EMBASE. After exclusion of 560 duplicates, 1,982 publications entered the title and abstract screening. On this basis, 1,956 publications were excluded because they either did not meet the predefined inclusion criteria and/or met at least one of the exclusion criteria. The inclusion and exclusion criteria can be found in Appendix 2 (Attachment 1) or, in abbreviated form, in the flow chart below (Figure 1). Thus, a total of 26 studies were subjected to full text screening, in which a further ten studies were excluded. One additional study was identified via the supplementary semi-structured search. Thus, a total of 17 studies were included in the content analysis. Since

the aim of this paper is to provide an overview of the specific criteria for the evaluation of medication apps that can be taken from the international evidence base, the general evaluation concepts cited in the background which have no direct reference to medication apps are not included in the analysis. However, the general assessment tools used in the included studies (e.g. MARS) are included in the analysis because they are specifically used to assess medication apps in the context of the respective studies. These assessment tools are explained in more detail below.

An overview of the 17 included studies is presented in Table 1. In addition to the authors, the year of publication and the country, this overview also shows the methodological approach of the studies and the instruments developed and/or tested for the evaluation of medication apps.

These are predominantly very recent studies (only 2/17 studies were published before 2016) conducted in the USA (n=6), Canada (n=6), Singapore (n=2), China (n=1), Ireland (n=1), and Australia (n=1).

The included studies differ in terms of their methodology with respect to three phases of the development of an assessment tool (see also Table 1 for details). In the first phase, criteria for the evaluation of medication apps are identified. A systematic literature review was conducted by two of the included 17 studies [18], [19]. In seven studies the criteria were retrieved from a consensus of the authors and/or semi-systematic searches [6], [20], [21], [22], [23], [24], [25]. Seven other studies adopted previously developed criteria and instruments [26], [27], [28], [29], [30], [31], [32]. One study did not provide information on the development of the criteria list [33]. In the second phase, an assessment instrument is developed based on the identified criteria. In the third phase, the assessment tool developed in each study was applied to evaluate medication apps. Between four [26] and 645 [28] medication apps were included in the different studies (see Table 1 for more details). The evaluation of the respective selected apps was partly based only on the app descriptions and screenshots of these and/or by downloading and testing the apps. The evaluation of the medication apps was carried out by the authors of the respective study or medical specialists (n=17) and, in some cases, by the target group itself (n=2) [26], [29].

The last two columns in Table 1 describe the evaluation instruments developed for the assessment of apps to support medication adherence. These evaluation instruments were either developed independently or adopted from other studies. Through these instruments, the criteria are measured or operationalized in the included studies.

A total of seven of the 17 included studies developed new instruments for the assessment [6], [18], [20], [33], [21], [19], [25]. In the other studies previously developed instruments were used exclusively (n=7) [26], [27], [28], [29], [30], [31], [32] or in addition to self-developed instruments (n=3) [22], [23], [24]. The simplest assessment

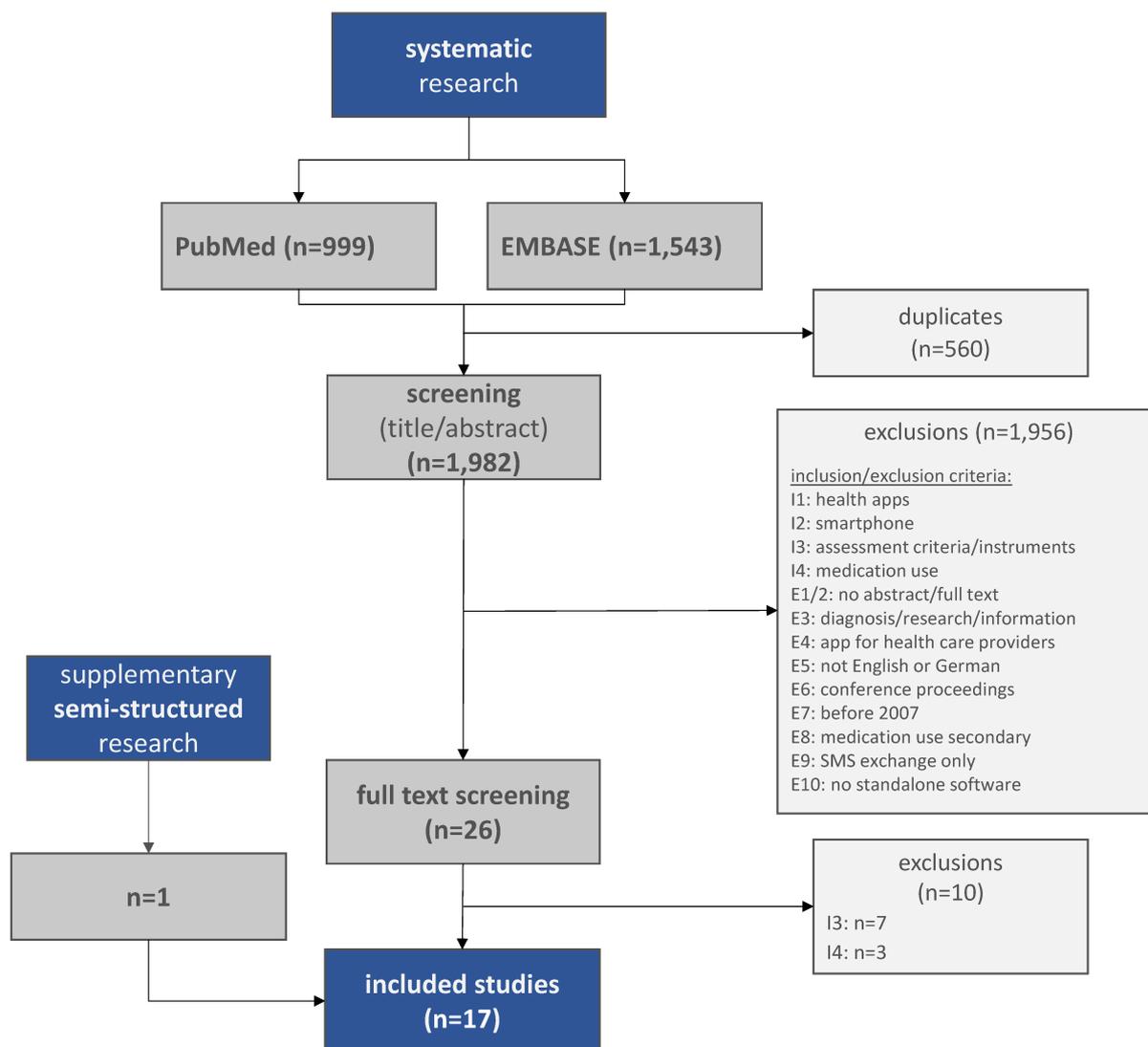


Figure 1: Flowchart of the scoping review

instrument is the nominal recording of a criteria. Accordingly, only the presence of certain functions or characteristics is recorded and coded binarily (yes/no or 0=criterion not present/1=criterion present). This operationalization was chosen by five of the ten studies with own contributions in the development of the assessment instrument [6], [33], [23], [24], [25]. The remaining five studies with self-developed instruments operationalized their criteria with varying differentiations. In these studies, individual criteria were weighted. For example, a subsequent weighting in a dichotomous response format was applied (n=3) [20], [21], [22]. In addition, weighting was done with response options to which fixed points are assigned (n=2) [19], [24] and/or Likert scales, i.e., a multilevel response scale with more or less strong agreement or disagreement were used [21]. In addition, a visual rating with one to five stars is used in one study [21]. Often, however, the evaluation instruments consist of a combination of different response formats.

In addition to these newly designed evaluation instruments, established evaluation instruments were used, which were not specifically designed for the evaluation

of medication apps. Although these general assessment catalogs are not directly aimed at answering the research question on medication-specific criteria, the MARS instrument, published in 2015, will be briefly discussed below due to the widespread use. It is also available in a version for end users (uMARS). MARS comprises a total of 29 questions in the five areas engagement, functionality, aesthetics, information quality, and subjective quality scale. The questions within the first four areas are measured on a five-point Likert scale (from “inadequate” to “excellent”). The subjective quality assessment is based on predetermined answering options. A mean value is calculated for each area, which is included into the overall evaluation of all mean values [34], [35]. In the studies by Carmody et al. [26] and Morrissey et al. [27], apps are assessed by using a Behavior Change Technique (BCT) catalog [36], [37]. Interventions that aim to change behavior are complex. The challenge is to identify the effective components within the intervention. A BCT describes a component of an intervention that is intended to change behavior, such as reinforcement or feedback. To create BCT catalogs, intervention descriptions from

**Table 1: Overview of the studies included in this scoping review**

Author (year)	Country	Method		Evaluation tool		
		1. Identification of criteria	2. Development of an evaluation tool	3. Application/evaluation	Newly developed	Adopted from studies
Ali et al. (2018) [18]	Singapore	Publication by Huckvale et al. (2015) [39] and Lewis (2013) [40] as well as systematic literature research	Authors (pharmacists) arrange and weight criteria according to relevance (MedAd-App)	Evaluation of 52 apps by two authors independently of each other, determination of the interrater correlation	MedAd-AppQ: 24 items (maximum 43 point) in three fields, selection	
Carmody et al. (2019) [26]	USA	Using a BCT catalog (according to Abraham et al. (2008) [37]) and uMARS (Stoyanov et al. (2016) [35])		BCT: evaluation of 75 apps by two authors independently of each other, uMARS: 20 young, chronically ill adults rate user friendliness of four apps	Availability of BCT, 0/1 uMARS (26 items in four fields, Likert scale)	
Dayer et al. (2013) [20]	USA	Authors develop list of desirable functions/features	Weighting by relevance of criteria through authors	First rating 160 apps with app descriptions, then testing of the best apps by two authors	13 weighted functions, yes/no (if yes, then 1 to 3 points), maximum 33 points	
Dayer et al. (2017) [28]	USA	Minor adjustments to existing tool according to Heidenbrand et al. (2016) [21] on the basis of Dayer et al. (2013) [20]		Evaluation by two authors independently of each other (first 645 with app descriptions, then top 100 through testing), determination of the correlation between user reviews (from app store) and own rating	System according to Heidenbrand et al. (2016) [21], with minor adjustments on the basis of Dayer et al. (2013) [20]	
Grindrod et al. (2014) [29]	Canada	Use of the SUS (according to Brooke et al. (1996) [41])		Testing of 22 apps by two authors independently of each other, five different apps for test by 35 user and focus groups	SUS (according to Brooke et al. (1996) [41]) =ten items, Likert scale, visual analogue scale	
Grindrod et al. (2017) [33]	Canada	Information on the development of the criteria list was not given		Evaluation of 184 apps by two authors independently of each other, determination of the inter-rater correlation	Availability of functions, 0/1	
Haase et al. (2017) [6]	USA	On the basis of 30 systematically selected apps, a catalog of the functions is created which showcases the "ideal functions" as evaluation criteria	Application Score Card (compilation about availability/absence of criteria)	Evaluation of the 30 apps with the Application Score Card used in 1. presumably by authors (no exact details)	Application Score Card: functions, 0/1	
Heidenbrand et al. (2016) [21]	USA	Authors (experts in medicine and science) create a list with 27 functions/features	Authors weight these by relevance and arrange them, development of a transfer modality into visual evaluation	Evaluation of 367 apps with app descriptions and testing of 77 apps by two authors independently of each other	27 weighted functions/features in four fields, yes/no (if yes, then 1 to 3 points according to Dayer et al. (2013) [20]) + health literacy (twelve items), Likert scale, transfer into visual evaluation with stars (one to five)	

(Continued)

**Table 1: Overview of the studies included in this scoping review**

Author (year)	Country	Method		3. Application/evaluation		Evaluation tool		Adopted from studies
		1. Identification of criteria	2. Development of an evaluation tool			Newly developed		
Jupp et al. (2018) [30]	Canada	Analogue approach (that is including MARS according to Stoyanov et al. (2015) [34])		Evaluation of 27 apps by at least two (unclear) authors independently of each other, analogue to Santo et al. (2016) [22]				System according to Santo et al. (2016) [22]
Kim et al. (2018) [31]	Canada	Use of MARS (according to Stoyanov et al. (2016) [35])		Download and evaluation of 23 apps by two authors independently of each other				MARS: five fields with 23 items, Likert scale
Liu et al. (2016) [32]	China	Use of the modified Silberg Scale (according to Jeon et al. (2014) [42])		First, analysis on availability of ten functions in Chinese apps, then evaluation of Chinese apps and comparison of each of the ten Chinese and international apps by two (if necessary three) medical specialists				Modified Silberg Scale (mS Scale); seven categories, 20 indicators
Loy et al. (2016) [19]	Singapore	Systematic literature research as well as research on the website iMedicalApps (2020) [43]		Development of a questionnaire with weighted answer options		81 items in four fields, selection (0 up to 4 points)		
Morrissey et al. (2016) [27]	Ireland	Use of a BCT catalog (BCTTv1, according to Michie et al. (2013) [36])		Download and evaluation of 166 apps by two authors independently of each other				BCT, 0/1
Nguyen et al. (2016) [23]	USA	Analysis of selected pediatric adherence studies (Kahana et al. (2008) [44]; Pai et al. (2014) [45]) by "useful" functions		List of "useful" functions and MARS (according to Stoyanov et al. (2015) [34])		Functions/features, 0/1		MARS (Field 1–3), twelve items, Likert scale
Park et al. (2019) [24]	Canada	In accordance with Dayer et al. (2013) [20] supplemented by criterion user friendliness		Evaluation and user test of 20 apps by two authors (pharmacists) independently of each other + content analysis of 1,323 user reviews		Functions without weighting, 0/1; user friendliness, 1 (easy) to 3 (hard)		Functions according to Dayer et al. (2013) [20]
Santo et al. (2016) [22]	Australia	Publications of Dayer et al. (2017) [28]; Stawarz et al. (2014) [46]; Bailey et al. (2014) [47] and discussions in author team		Own tool + use of MARS (according to Stoyanov et al. (2015) [34])		20 weighted functions/features, yes/no (if yes, then 1 to 2 points), maximum 23 points		MARS: five fields with 23 items, Likert scale
Tabi et al. (2019) [25]	Canada	Publications of Bailey et al. (2014) [47]; Shen et al. (2015) [48]; Bender et al. (2013) [49] and discussion. Adjustment of list according to information collected in the evaluation of the apps		Evaluation of 66/328 apps by two authors together, 262/328 apps by one author according to app descriptions/Google search		Functions/features, 0/1		

BCT: Behavior Change Technique. MARS: Mobile Application Rating Scale. uMARS: User Version of the Mobile Application Rating Scale. SUS: System Usability Scale. Source: own presentation

studies that identified techniques associated with effective outcome were used [36]. Medication apps aim to change the user's behavior to promote adherence. If the concept of the respective app is divided into different components, e.g., reminder function, reward system, or medication list, it is possible to identify the effective components of the apps aiming at behavior change by coding the components according to BCT catalogs. This allows for the differentiation of apps. The presence or absence (0/1) of 26 BCT items is coded in Abraham et al. [37] and of 93 BCT items in Michie et al. [36]. For example, Carmody et al. [26] identify techniques in the apps studied such as detailed planning/definition of behavior (setting alarms, defining medication schedule), identification of environmental cues that remind of desired behavior (setting alarms for a specific time), or recording behavior (recording whether medication dose was taken). In addition to the use of known BCT catalogs, the other included studies (8/17) also recorded elements designed to motivate the user to change behavior [6], [26], [21], [30], [22], [31], [23], [25]. These include, for example, the motivational functions or features "protocol", "incentives", and "interaction". Even though the authors of these studies do not explicitly refer to these evaluation criteria as BCT, they also pursue the objective of identifying components of the tested apps that can change the user's behavior by evaluating them based on motivational functions and features.

The criteria catalog for medication apps developed as part of this study includes the legally defined criteria and combines them with the criteria identified in the literature search. It comprises the overarching topics "patient orientation" and "quality/core functions of medication apps available" as well as the area of "interoperability/collaboration". The categories "data protection and security", "consumer protection", and "user friendliness" are subsumed under the "patient orientation" area. The categories of the topic area "quality/core functions of medication apps available" include "reminder", "self-monitoring", "(medication) information", "motivation for behavior change", "medication/patient safety", and "robustness". In particular, the medication/adherence-specific criteria go beyond the structure of the DiGAV. The criteria identified in the literature review could first be assigned to the categories of the developed criteria catalog and then summarized as a total of 59 evaluation criteria (see also Attachment 2, Appendix 4). This assignment can be seen in Figure 2. The size of the bubbles indicates the number of mentions. For example, the criterion "recording of the medication use" with 11/17 mentions is shown significantly larger in Figure 2 than the criterion "statistics" with 2/17 mentions.

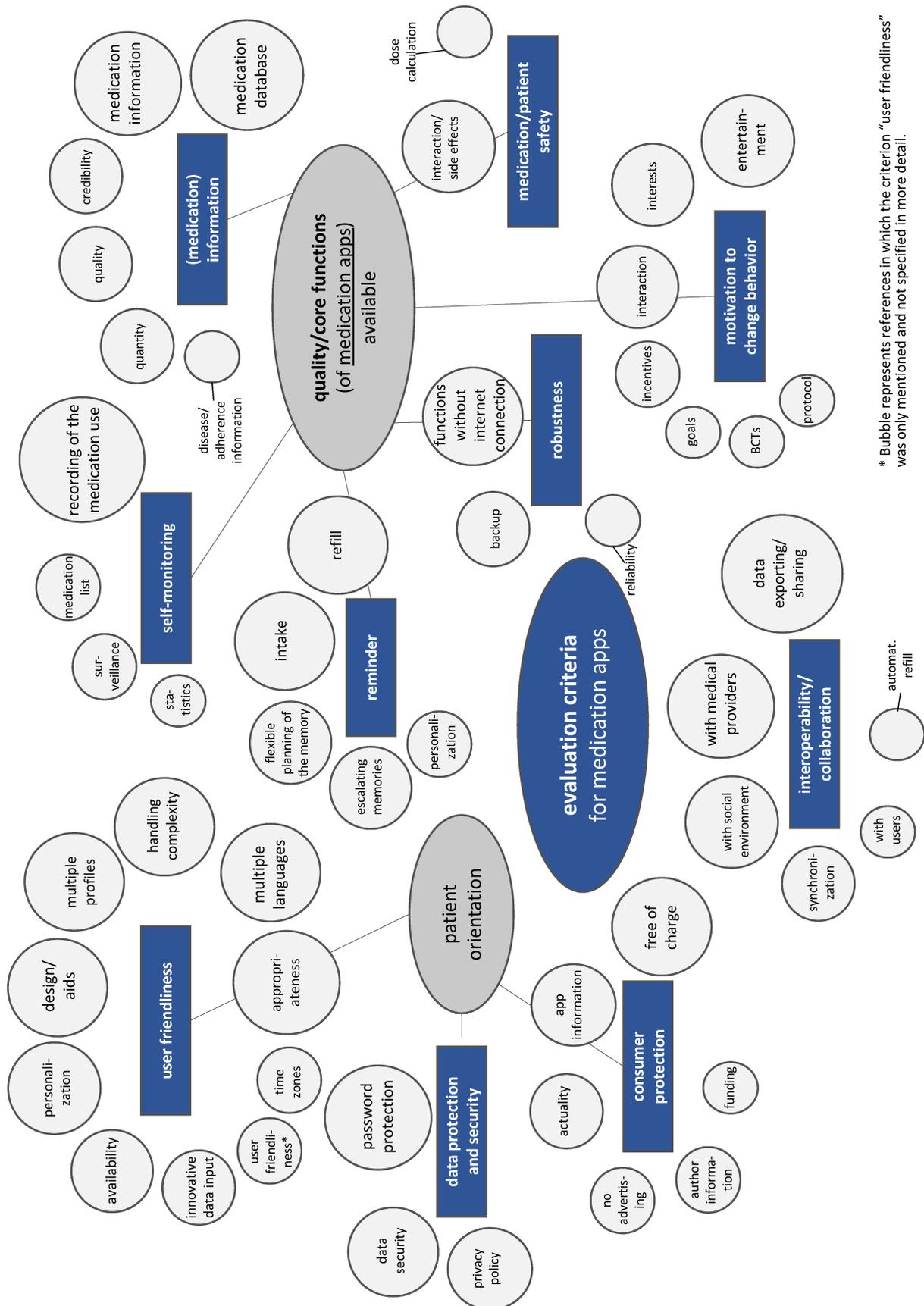
The wide range of mentions of particular evaluation criteria in the respective studies is striking. Seven individual criteria are each listed in only one of the studies included in this scoping review. They are not shown in Figure 2, but are listed in the legend of Figure 2. In contrast, the criterion "recording of the medication use" is mentioned in 11/17 publications. The category "user friendliness"

shows a wide variety of evaluation aspects with eleven subcategories, whereas the categories "robustness" and "data protection and security" show a narrow range of criteria with only three criteria each. The included studies place a large focus on the area of core functions of medication apps. Among the wide range of criteria, the presence of functions for "recording of the medication use" (11/17) and "medication database" (9/17) are mentioned particularly frequently. Other frequently mentioned criteria are "data exporting/sharing" (10/17) and "collaboration with medical providers" (8/17) in the "interoperability/collaboration" category, "password protection" (8/17) in the "data protection and security" category, and "appropriateness" (8/17) in the "user friendliness" category. Furthermore, some aspects are mentioned in the studies to support the app user's motivation to change behavior. Two BCT catalogs have been incorporated, which in turn describe a plethora of behavior change techniques as a stand-alone tool. For example, Michie et al. [36] group 93 BCTs into the following categories: planned consequences, reward and threat, repeat and replace, environment, associations, covert learning, natural consequences, feedback and monitoring, goals and planning, social support, comparing behavior, confidence, comparing outcomes, identity, shaping knowledge, and regulation.

## Discussion

Through the conducted literature search, 59 different criteria for the evaluation of apps for medication adherence support could be identified. In the included studies, 10/17 studies developed assessment tools based on systematic literature review, author consensus, semi-systematic literature search, or in one case without specifying the basis, and/or 10/17 studies used existing generic tools (e.g., MARS or BCT). The 59 identified criteria were assigned to the categories of the criteria catalog developed in this study. This criteria catalog includes the headings "patient orientation" and "quality/core functions of medication apps available" as well as the area "interoperability/collaboration". The three categories "data protection and security", "consumer protection" and "user friendliness" are subsumed under the "patient orientation" area. Categories in the topic area "quality/core functions of medication apps available" include six categories of "reminder", "self-monitoring", "(medication) information", "motivation to change behavior", "medication/patient safety", and "robustness". While some of the categories specifically target medication apps, others are also relevant in the evaluation of general health apps.

The category "motivation to change behavior" stands out in the developed criteria catalog due to its specific importance for medication apps and its special design and will be briefly discussed below. This category aims to evaluate the design of individual functions in terms of their potential to actually change the behavior of app users. In two



\* Bubble represents references in which the criterion "user friendliness" was only mentioned and not specified in more detail.

Figure 2: Summary of the evaluation criteria and categorization. Criteria mentioned in only one study are not included in the figure. These include "technical support" (user friendliness), "data transfer to the ePA" (interoperability/collaboration), "recording progress", "problem solving regarding adherence barriers" (motivation to change behavior), "risk assessment", "emergency management" (medication/patient safety), and "reminder of physician appointments" (reminder).

studies [26], [27], detailed question catalogs with evidence-based BCT were used for this purpose. In addition, individual criteria were addressed in eight of the 17 included studies in this area, such as incentives or goal setting. The two included papers using the BCT catalogs assigned the components of the tested apps to specific behavior change techniques (BCT). Via this mapping, it is possible to identify the components of an app that are effective in terms of behavior change. Assuming that motivation for behavior change is one of the main goals of medication apps, a qualitative evaluation of individual components and their specific design can be carried out that way. When designing a medication app, the manufacturer can examine components of the app concerning the criteria of motivation to change behavior and thus focus on particularly effective components and their design.

Since the DiGAV and the BfArM guideline were important starting points for the development of the criteria catalog presented in this study, a brief comparison of the central legal requirements and the included studies will be conducted below. Overall, the appendices of the DiGAV contain comprehensive questionnaires with a higher level of detail than the included studies. These questionnaires serve as a tool for part of the evaluation categories defined by the DiGAV. These include data privacy and security (Appendix 1 of the DiGAV, [10]) as well as interoperability, robustness, consumer protection, user friendliness and accessibility, provider support, medical content quality, and patient safety (Appendix 2 of the DiGAV, [10]). Via yes-no statements as well as the statement “not applicable” the manufacturer confirms the queried criteria. A “not applicable” answer that is not given for selection requires a written explanation of why the criterion has not been fulfilled yet [10]. The category of positive health care effects is explained in the text of the regulation and ranks highly in the DiGAV [10] for inclusion in the Digital Health Applications Directory [10]. According to Section 10 of the DiGAV [10], [11], the manufacturer must demonstrate positive care effects with quantitative comparative studies. These can be clinical or epidemiological studies, but methods from other scientific fields such as health services research, social research, or behavioral research are also permissible, as long as the chosen method fits the subject of the study. Not only prospective, but also retrospective studies (including data from digital patient records, routine data, and registries) are licit. For example, case-control studies, retrospective cohort studies, or even intraindividual comparisons are applicable [10]. In addition to the positive health care effects mentioned in the regulation, the criteria catalog developed in this paper lists individual criteria that correspond to the DiGAV [10] categories of the coordination of treatment processes (e.g., criterion “collaboration with medical providers”), patient safety (e.g., criterion “interaction/side effects”), and health literacy (e.g., “medication information”). For the important categories of prescription adherence and medical benefit (shortening of disease duration, prolon-

gation of survival, improvement of quality of life), however, the criteria catalog does not provide any criterion that directly measures these dimensions. Thus, in the included studies, the assessment of medical effects by criteria plays no or a rather subordinate role. This will be explained in more detail in the following sections. Furthermore, the DiGAV mentions the CE conformity marking as the only aspect for assessing safety and functional suitability. This EU marking is not addressed in the included studies, which were primarily conducted in non-EU countries; however, comparable, general markings are not mentioned either. The category of interoperability and collaboration, which is highlighted in the DiGAV, is also found in the included studies. This involves, for example, resetting an app to its initial state, plausibility checks when entering data, transferring data to the electronic patient record, or sharing data with care providers and the social environment. Overall, it is clear that the structure and content of the criteria catalog developed in this paper based on the included studies show a large overlap with the requirements of the DiGAV. Only the area of positive care effects was hardly covered by the included studies.

For quality assurance purposes, the model developed by Arah et al. [38] in 2006 for the comparative assessment of health care systems in OECD member countries (OECD=Organisation for Economic Co-operation and Development) will be applied to the list of criteria developed in this study. In this model, criteria for assessing the performance of a health care system are grouped into specific dimensions in a condensed list. The comparison shows that the aspects of the OECD model safety of care (esp. via categories “data protection and security”, “consumer protection”, “medication/patient safety”), patient-centeredness (esp. via categories “user friendliness” and “motivation to change behavior”), access to care (esp. category “user friendliness”), and expenditures/costs (esp. criterion “free of charge” in category “consumer protection”) are addressed by the catalog of criteria developed in this study. However, the main segment of effectiveness described in the OECD model cannot be filled with criteria.

The lack of evaluation criteria regarding the effectiveness or positive effects of medication apps on care is due to the fact that the evaluation criteria in the included studies primarily focus on the range of functions and technical features of the apps. This in turn is due to the methodology of the included studies. The evaluation of the selected apps in the third phase of the included studies was partly based only on the app descriptions and screenshots of these and/or by downloading and testing the apps over a very short period of time by the authors and not by users themselves (the latter in only 2/17 studies). Hence, only dimensions that are measurable in the early development phase of the technology and thus can be assessed in the short term can be considered according to the objective of this scoping review. As to the identified criteria, it could not be determined whether and to what extent there are correlations between a particular evaluation criterion and

long-term medical benefits. The evaluation of criteria concerning long-term effectiveness would have to be implemented in clinical trials of a comparative nature in which patients or app users are included. Clinical evaluation of care effects could be measured with outcome parameters related to adherence or drug safety/interactions. At the early stages of technology development, systematic reviews of clinical trials on medication apps might help to identify possibly effective app components that might contribute to the improvement of the quality of care. In addition, manufacturers could evaluate medication app components according to the BCT catalogs described earlier in order to gain an impression of the app's potential in terms of user behavior change.

Furthermore, it should be emphasized that the frequency of criteria mentions does not yet indicate their relevance. It was not possible to weight the identified criteria as part of the development of the criteria catalog. Thus, from this scoping review app developers cannot derive which aspects picking up on the evaluation criteria should be given special emphasis in the design of the app in order to generate the greatest possible medical benefit for the app user. To weight the evaluation criteria identified in this scoping review, further research is essential, particularly in the form of comparative clinical studies that establish and quantify a relationship between the fulfillment of certain criteria and the occurrence of medical effects. This may confirm criteria that are associated with increased patient benefit in the long term. In addition, the relevance of criteria could differ greatly with respect to the perspective taken. Therefore, preference surveys, such as choice-based conjoint analyses, could be a useful addition as a basis for weighting criteria. In these, alternatives could be evaluated as a bundle of their characteristics (i.e., specific evaluation criteria) from different perspectives (service provider, user, payer) in an intuitive decision-making process. For example, from the perspective of the treating physician, criteria that aim at an efficient (data) exchange with the patient and evaluate the effectiveness of the medication app could be important. Users, on the other hand, might consider criteria that evaluate the presentation of medical content and the handling of the app to be more important. A mapping of preferences to the criteria identified in this study with regard to different target groups could improve the use as a targeted basis for decision-making.

## Limitations

In the included studies, the mentioned criteria are presented in different forms, for example as a sentence or only as a short bullet point, and different formulations for criteria are chosen. The summary of the criteria of the included studies into unified evaluation criteria can be differentiated to varying degrees. With this in mind, a balance between detailed differentiation and clear summary was sought in this study and developed by consensus as understood by the authors of this paper. This

also applies to the chosen structure of the criteria catalog, which is based on the results found in the studies in addition to the structure of the DiGAV. Moreover, in the context of a scoping review, the evidence from the literature is reported regardless of the quality of the included studies, so that all criteria found were equally mentioned and summed up. Furthermore, the selected inclusion and exclusion criteria of the literature search were used to identify studies that addressed the design of evaluation criteria/instruments for medication apps in order to apply them to the evaluation of a set of different apps. Studies that only addressed the evaluation of a specific app were excluded. It is possible that evaluation criteria could have been indirectly derived from these publications as well.

## Answering the research questions

The criteria for evaluating health apps identified in the DiGAV had a high degree of intersection with the criteria identified in the literature review. However, the area of positive care effects was poorly captured with the included studies. Quality assurance of the criteria catalog via the OECD model according to Arah et al. [38] revealed that the aspects of safety of care, patient-centeredness, access to care, and expenditures/costs were covered via the developed criteria catalog. The main segment of effectiveness described in the OECD model, however, could also not be fulfilled due to the short-term dimension or the rather functional orientation of the identified evaluation criteria. In addition, the frequency of mentions of the criteria alone does not indicate the relevance of a criterion. For these reasons, it was not possible to weight the identified criteria during the development of the criteria catalog. Therefore, the criteria catalog developed should not be understood by service providers, manufacturers and/or users as a comprehensive evaluation tool with a clear result, but rather as a supporting checklist. It is thus only a first step in the development of decision support tools.

Important potentials are associated with the criteria catalog developed in this article. In order to be included in the list of reimbursable digital health applications maintained by the BfArM in accordance with § 139e SGB V, health apps must fulfill the required conditions of the DiGAV [10], which are explained by the BfArM by means of examples in the DiGAV guide. In addition to the manufacturer's specifications, the list also includes information on the features and performance of the digital health apps listed [11]. If the healthcare provider or a patient himself [11] is faced with the challenge of selecting the most suitable registered application, an evaluation that goes beyond the register and is more specific to the area of medication management plays an important role. The evaluation criteria identified in the present study can serve as a basis for a qualitative evaluation of medication apps and thus provide a supplementary decision-making aid for healthcare providers and also for potential app users for the needs-based selection of medication apps

that have been included in the register of reimbursable digital health apps.

## Conclusion

In view of the ever scarcer resources in the health care system, the efficient use of drugs is essential. Schäfer [2] states that the most expensive drug is the one that the patient does not take. Therefore, digital approaches promoting adherence should be focused on. However, medical care providers, manufacturers and users need more specific guidance than the general concepts available to date. In the present scoping review, an overview of the international evidence base on criteria for the evaluation of medication apps could be created and presented in a structured way in a criteria catalog. In addition to the studies included in the literature search, the legal requirements in Germany (DVG, DiGAV) in particular served as the basis for the development of this criteria catalog. Future research should focus on ways to weight these criteria, using preference survey methods in addition to clinical studies. Furthermore, dimensions of effectiveness or positive effects of care should be added to the criteria catalog.

## Abbreviations

- BCT: Behavior Change Technique
- BfArM: Bundesinstitut für Arzneimittel und Medizinprodukte (Federal Institute for Drugs and Medical Devices)
- DiGA: Digitale Gesundheitsanwendung (digital health application)
- DiGAV: Digitale-Gesundheitsanwendungen-Verordnung (Digital Health Applications Ordinance)
- DVG: Digitale-Versorgung-Gesetz (Digital Healthcare Act)
- ePA: Elektronische Patientenakte (electronic patient file)
- SHI: Statutory health insurance
- MARS: Mobile App Rating Scale
- Medikations-App: Gesundheitsapps zur Unterstützung der Medikamenteneinnahme (health app to support intake of medications)
- NHS: National Health Service (Great Britain)
- OECD: Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (Organisation for Economic Co-operation and Development)
- SGB: Sozialgesetzbuch (social security code)
- uMARS: User Version of the Mobile Application Rating Scale
- WHO: World Health Organization

## Notes

### Funding

The project was supported by the German Institute of Medical Documentation (DIMDI).

### Competing interests

The authors declare that they have no competing interests.

## Attachments

Available from <https://doi.org/10.3205/000317>

1. Attachment 1\_gms000317.pdf (108 KB)  
Appendix 1–3
2. Attachment 2\_gms000317.pdf (119 KB)  
Appendix 4

## References

1. World Health Organization. Adherence to long-term therapies. [last accessed 2020 Mar 11]. Available from: <https://apps.who.int/iris/handle/10665/42682?locale-attribute=de&>
2. Schäfer C. Patient compliance. Adhärenz als Schlüssel für den Therapieerfolg im Versorgungsalltag. 2nd ed. Wiesbaden: Springer; 2017.
3. Masnoon N, Shakib S, Kalisch-Ellett L, Caughy GE. What is polypharmacy? A systematic review of definitions. *BMC Geriatr*. 2017 Oct;17(1):230. DOI: 10.1186/s12877-017-0621-2
4. Kantor ED, Rehm CD, Haas JS, Chan AT, Giovannucci EL. Trends in Prescription Drug Use Among Adults in the United States From 1999-2012. *JAMA*. 2015 Nov;314(17):1818-31. DOI: 10.1001/jama.2015.13766
5. Kunst A. Welche dieser Gesundheits-Apps und smarten Gesundheits-Geräte (mit dem Internet vernetzt) haben Sie in den letzten 12 Monaten genutzt? In: Umfrage in Deutschland zu Health-Apps und smarten Gesundheitsgeräten 2017. Statista; 2017 [last accessed 2020 Mar 11]. Available from: <https://de.statista.com/prognosen/810056/umfrage-in-deutschland-zu-health-apps-und-smarten-gesundheitsgeraeten>
6. Haase J, Farris KB, Dorsch MP. Mobile Applications to Improve Medication Adherence. *Telemed J E Health*. 2017 Feb;23(2):75-9. DOI: 10.1089/tmj.2015.0227
7. Ng R, Carter SR, El-Den S. The impact of mobile applications on medication adherence: a systematic review. *Transl Behav Med*. 2020 Dec;10(6):1419-35. DOI: 10.1093/tbm/ibz125
8. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz – DVG). *Bundesgesetzblatt*. 2019 Dec 18; 2019 Part I(49): 2562.
9. Albrecht UV. Orientierung für Nutzer von Gesundheits-Apps. In: Albrecht UV, editor. Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Medizinische Hochschule Hannover; 2016. p. 282-300.

10. Verordnung über das Verfahren und die Anforderungen zur Prüfung der Erstattungsfähigkeit digitaler Gesundheitsanwendungen in der gesetzlichen Krankenversicherung (Digitale Gesundheitsanwendungen-Verordnung – DiGAV). Bundesgesetzblatt. 2020 Apr 8; 2020 Part I(18): 768.
11. Bundesinstitut für Arzneimittel und Medizinprodukte. Das Fast-Track-Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V. Ein Leitfaden für Hersteller, Leistungserbringer und Anwender. Version 3.1. [last accessed 2023 Feb 02]. Available from: [https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga\\_leitfaden.pdf?\\_\\_blob=publicationFile](https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.pdf?__blob=publicationFile)
12. Brönnner M, Meister S, Breil B, Albrecht UV. Kapitel 15. Orientierung für Hersteller von Gesundheits-Apps. In: Albrecht UV, editor. Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Medizinische Hochschule Hannover, 2016. p. 320-40.
13. Bertelsmann Stiftung. AppQ. Gütekriterien-Kernset für mehr Qualitätstransparenz bei digitalen Gesundheitsanwendungen. [last accessed 2020 Feb 15]. Available from: [https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Studienbericht\\_AppQ\\_191028.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Studienbericht_AppQ_191028.pdf)
14. Fraunhofer-Institut für Offene Kommunikationssysteme. APPKRI Kriterien für Gesundheits-Apps. [last accessed 2020 Feb 15]. Available from: <https://ehealth-services.fokus.fraunhofer.de/BMG-APPS/>
15. NHS England. NHS Apps Library. Guidance for health app developers, commissioners and assessors to get trusted digital tools assessed and published on the NHS Apps Library. [last accessed 2020 Feb 15]. Available from: <https://digital.nhs.uk/services/nhs-apps-library>
16. von Elm E, Schreiber G, Haupt CC. Methodische Anleitung für Scoping Reviews (JBI-Methodologie). Z Evid Fortbild Qual Gesundhwes. 2019 Jun;143:1-7. DOI: 10.1016/j.zefq.2019.05.004
17. Laugesen J, Yuan Y. What Factors Contributed to the Success of Apple's iPhone? In: Ninth International Conference on Mobile Business and Ninth Global Mobility Roundtable (ICMB-GMR); 2010 June 13-15; Athens. p. 91-9.
18. Ali EE, Teo AKS, Goh SXL, Chew L, Yap KY. MedAd-AppQ: A quality assessment tool for medication adherence apps on iOS and android platforms. Res Social Adm Pharm. 2018 Dec;14(12): 1125-33. DOI: 10.1016/j.sapharm.2018.01.006
19. Loy JS, Ali EE, Yap KY. Quality Assessment of Medical Apps that Target Medication-Related Problems. J Manag Care Spec Pharm. 2016 Oct;22(10):1124-40. DOI: 10.18553/jmcp.2016.22.10.1124
20. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers. J Am Pharm Assoc (2003). 2013 Mar-Apr;53(2):172-81. DOI: 10.1331/JAPhA.2013.12202
21. Heldenbrand S, Martin BC, Gubbins PO, Hadden K, Renna C, Shilling R, Dayer L. Assessment of medication adherence app features, functionality, and health literacy level and the creation of a searchable Web-based adherence app resource for health care professionals and patients. J Am Pharm Assoc (2003). 2016 May-Jun;56(3):293-302. DOI: 10.1016/j.japh.2015.12.014
22. Santo K, Richtering SS, Chalmers J, Thiagalangam A, Chow CK, Redfern J. Mobile Phone Apps to Improve Medication Adherence: A Systematic Stepwise Process to Identify High-Quality Apps. JMIR Mhealth Uhealth. 2016 Dec;4(4):e132. DOI: 10.2196/mhealth.6742
23. Nguyen E, Bugno L, Kandah C, Plevinsky J, Pouloupoulos N, Wojtowicz A, Schneider KL, Greenley RN. Is There a Good App for That? Evaluating m-Health Apps for Strategies That Promote Pediatric Medication Adherence. Telemed J E Health. 2016 11;22(11):929-37. DOI: 10.1089/tmj.2015.0211
24. Park JYE, Li J, Howren A, Tsao NW, De Vera M. Mobile Phone Apps Targeting Medication Adherence: Quality Assessment and Content Analysis of User Reviews. JMIR Mhealth Uhealth. 2019 Jan;7(1):e11919. DOI: 10.2196/11919
25. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, Nikoo M, Vigo D, Jang K, Demlova R, Krausz M. Mobile Apps for Medication Management: Review and Analysis. JMIR Mhealth Uhealth. 2019 Sep;7(9):e13608. DOI: 10.2196/13608
26. Carmody JK, Denson LA, Hommel KA. Content and Usability Evaluation of Medication Adherence Mobile Applications for Use in Pediatrics. J Pediatr Psychol. 2019 Apr;44(3):333-42. DOI: 10.1093/jpepsy/jsy086
27. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior Change Techniques in Apps for Medication Adherence: A Content Analysis. Am J Prev Med. 2016 May;50(5):e143-6. DOI: 10.1016/j.amepre.2015.09.034
28. Dayer LE, Shilling R, Van Valkenburg M, Martin BC, Gubbins PO, Hadden K, Heldenbrand S. Assessing the Medication Adherence App Marketplace From the Health Professional and Consumer Vantage Points. JMIR Mhealth Uhealth. 2017 Apr;5(4):e45. DOI: 10.2196/mhealth.6582
29. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth. 2014 Mar;2(1):e11. DOI: 10.2196/mhealth.3048
30. Jupp JCY, Sultani H, Cooper CA, Peterson KA, Truong TH. Evaluation of mobile phone applications to support medication adherence and symptom management in oncology patients. Pediatr Blood Cancer. 2018 Nov;65(11):e27278. DOI: 10.1002/pbc.27278
31. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer Mobile Apps for Potential Drug-Drug Interaction Check: Systematic Review and Content Analysis Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth. 2018 Mar;6(3):e74. DOI: 10.2196/mhealth.8613
32. Liu Y, Chang P, Li N, Lamb KV, Wang J, Huang X, Wang Y, Wu Y. Quality of Chinese CVD Medication Apps. Stud Health Technol Inform. 2016;225:1036-7.
33. Grindrod K, Boersema J, Waked K, Smith V, Yang J, Gebotys C. Locking it down: The privacy and security of mobile medication apps. Can Pharm J (Ott). 2017 Jan-Feb;150(1):60-6. DOI: 10.1177/1715163516680226
34. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth. 2015 Mar;3(1):e27. DOI: 10.2196/mhealth.3422
35. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth. 2016 Jun;4(2):e72. DOI: 10.2196/mhealth.5849
36. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, Eccles MP, Cane J, Wood CE. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med. 2013 Aug;46(1):81-95. DOI: 10.1007/s12160-013-9486-6
37. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol. 2008 May;27(3):379-87. DOI: 10.1037/0278-6133.27.3.379

38. Arah OA, Westert GP, Hurst J, Klazinga NS. A conceptual framework for the OECD Health Care Quality Indicators Project. *Int J Qual Health Care*. 2006 Sep;18 Suppl 1:5-13. DOI: 10.1093/intqhc/mzi024
39. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. *BMC Med*. 2015 May;13:106. DOI: 10.1186/s12916-015-0314-7
40. Lewis TL. A systematic self-certification model for mobile medical apps. *J Med Internet Res*. 2013 Apr;15(4):e89. DOI: 10.2196/jmir.2446
41. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. *Usability evaluation in industry*. London: Taylor and Francis; 1996.
42. Jeon E, Park HA, Min YH, Kim HY. Analysis of the information quality of Korean obesity-management smartphone applications. *Healthc Inform Res*. 2014 Jan;20(1):23-9. DOI: 10.4258/hir.2014.20.1.23
43. iMedicalApps. iMedicalApps Medpage Today. [last accessed 2020 Mar 03]. Available from: <https://www.imedicalapps.com/>
44. Kahana S, Drotar D, Frazier T. Meta-analysis of psychological interventions to promote adherence to treatment in pediatric chronic health conditions. *J Pediatr Psychol*. 2008 Jul;33(6):590-611. DOI: 10.1093/jpepsy/jsm128
45. Pai AL, McGrady M. Systematic review and meta-analysis of psychological interventions to promote treatment adherence in children, adolescents, and young adults with chronic illness. *J Pediatr Psychol*. 2014 Sep;39(8):918-31. DOI: 10.1093/jpepsy/jsu038
46. Stawarz K, Cox A, Blandford A. Don't forget your pill! Designing effective medication reminder apps that support users' daily routines. In: 32nd annual Association for Computing Machinery (ACM) conference on Human factors in computing systems; 2014 Apr 26 – May 01; Toronto, Ontario, Canada; 2014. p. 2269-78.
47. Bailey SC, Belter LT, Pandit AU, Carpenter DM, Carlos E, Wolf MS. The availability, functionality, and quality of mobile applications supporting medication self-management. *J Am Med Inform Assoc*. 2014 May-Jun;21(3):542-6. DOI: 10.1136/amiajnl-2013-002232
48. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, Wiljer D. Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR Mhealth Uhealth*. 2015 Feb;3(1):e16. DOI: 10.2196/mhealth.3713
49. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. *J Med Internet Res*. 2013 Dec;15(12):e287. DOI: 10.2196/jmir.2661

**Corresponding author:**

Dr. Silke Neusser

Institute for Health Care Management and Research,  
University Duisburg-Essen, Thea-Leymann-Str. 9, 45127  
Essen, Germany

silke.neusser@medman.uni-due.de

**Please cite as**Niemann A, Hüer T, Neumann A, Wasem J, Schnell-Inderst P, Neusser S. *Evaluation criteria for health apps supporting medication adherence in early-stage technology development – a scoping review*. *GMS Ger Med Sci*. 2023;21:Doc03.

DOI: 10.3205/000317, URN: urn:nbn:de:0183-0003171

**This article is freely available from**<https://doi.org/10.3205/000317>**Received:** 2022-04-28**Published:** 2023-04-04**Copyright**©2023 Niemann et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.