10 Pragmatic Points to Consider When Performing a Systematic Literature Review of Clinical Evidence on Digital Medical Devices

Márta Péntek, Zsombor Zrubka and László Gulácsi

Health Economics Research Center, University Research and Innovation Center; Doctoral School of Innovation Management, Óbuda University, Bécsi út 96/b, H-1034 Budapest, Hungary, pentek.marta@uni-obuda.hu; zrubka.zsombor@uni-obuda.hu; gulacsi@uni-obuda.hu

Miklós Weszl

Department of Translational Sciences, Semmelweis University, Tűzoltó u. 37-47, H-1094 Budapest, Hungary, miklos.weszl@mdrpros.com

János Tibor Czere

Doctoral School of Innovation Management, Óbuda University, Bécsi út 96/b, H-1034 Budapest, Hungary; PSI CRO Hungary LLC, Dózsa György út 84/B, H-1068 Budapest, Hungary, czere.janos@phd.uni-obuda.hu

Tamás Haidegger

Antal Bejczy Center for Intelligent Robotics, University Research and Innovation Center, Óbuda University, Bécsi út 96/b, H-1034 Budapest, Hungary; haidegger@irob.uni-obuda.hu; Austrian Center for Medical Innovation and Technology, Viktor-Kalpan-str. 2, 2700 Wiener Neustadt, Austria, haidegger@acmit.at

Abstract: Innovative Digital Medical Devices (DMDs) are gaining a growing importance in healthcare, thus the interest regarding their actual clinical value is increasing. Gathering and evaluating all available clinical evidence on a device is needed for the developers, manufacturers, notified bodies (certification), product managers, clinicians and, above all, for the users. Systematic Literature Review (SLR), is a scientific, standardized method, often used to identify and evaluate all available evidence on a

relevant question, hence this method is widely used and accepted to provide high-quality evaluations on clinical evidence. Moreover, the Medical Device Regulation (EU 2017/745) also considers clinical evidence derived from the peer-reviewed scientific literature for the assessment of medical devices being adequate, considering even sustainability aspects. Researchers therefore can actually contribute to the clinical assessment of novel DMDs through the rigorous assessment of the State-of-the-art and the comprehension of the existing levels of evidence. Although the methodology of SLRs has been established in the past decades, some specific questions arise when it comes to DMDs. The objective of this paper was, therefore, to establish a set of practical points to consider in SLRs of clinical evidence on DMDs. We believe that our work helps to speed up DMD development, authorization, management and implementation, thereby eventually, contribute to the health of patients.

Keywords: innovation; digital medical device; systematic literature review; medical device regulation; management; robotic surgical procedures; chatbot

1 Introduction

Digital Medical Devices (DMDs) represent a broad spectrum of digital tools developed for preventive, diagnostic or therapeutic purposes. DMDs include, for instance, surgical or assistive social robots, diagnostic imaging software, mobile phone applications, and wearables to monitor health parameters.

As with any medical intervention, a fundamental question for DMDs is their effectiveness including both, the beneficial effects of an intervention on intended outcomes and the problems [1]. For clinical decisions, we need to know whether new technologies differ or exceed in their efficacy and safety compared to existing methods. These questions need to be answered for the authorization and implementation of innovative DMDs in the clinical practice, thus should be considered throughout their management, from the development to the market introduction. Questions on clinical evidence can be addressed in experimental clinical trials, observational studies or by analyses of real-world data.

The systematic literature review (SLR) is a scientific, standardized method to identify and evaluate all available evidence on a relevant question. SLRs are fundamental in evidence-based medicine and have been widely applied in non-medical fields. The independent Cochrane Collaboration [2] is considered globally as the preeminent leader in the production of SLRs and the development of their methodological standards.

The Medical Device Regulation (MDR) of the European Parliament and of the Council (EU 2017/745) requires clinical evidence from manufacturers and considers clinical study data reported *'in scientific literature, reports published in peer reviewed scientific journals'* and the assessment includes the '*methodology*

for the literature search and *relevant documentation of the literature search*' [3]. Furthermore, several countries published evidence frameworks for the health technology assessment of digital health applications for public financing decisions [4]. The methodological quality of SLRs is a focus point of the regulatory assessment, and the accelerating technological development and institutional frameworks will increase the demand for the generation and synthesis of clinical evidence on DMDs devices, calling for systematic reviews to summarize findings and support decision-making.

Our objective is to help stakeholders in planning, conducting and reporting SLRs in the field of DMDs. While this has been widely applied as a research tool, to assess the maturity of a technology, new aspects have to be considered when SLR is used as an instrument along the regulatory clearance pathways. We aim to establish a set of Points To Consider (PTCs) that arise specifically in the field of DMDs but are not necessarily addressed or emphasized in general SLR tutorials or have been poorly followed in the available DMD-related SLR works. First, we provide a brief introduction to the SLR method in general. Then, we formulate eight general and two device-specific practical points we suggest considering when carrying out a SLR of clinical evidence on DMDs.

2 Background: Standard steps of a SLR

Systematic reviews and evidence syntheses, are considered research on their own [5]. Therefore, the planning, implementation, and reporting of such research require a systematic approach and scientific rigor, akin to empirical studies [6, 7]. The methodology for undertaking SLRs continues to evolve with the growing need for evidence-informed decision-making by healthcare and public health professionals, and policymakers [8]. In the following sections, the standard steps of a SLR are presented briefly, including considerations from the most recent guidelines.

2.1 Planning the SLR

Planning for resources: The implementation of a SLR requires a significant amount of time, methodological savvy and meticulous planning from the very beginning. Setting up a capable review team with clinical and methodological expertise, including librarian, statistical and data management support, is a key factor to conduct the review within a reasonable time. Additionally, formulating a research question aligned with the author team's capabilities ensures effective utilization of the available evidence [8].

Framing a review question: A SLR involves the pre-definition of a focused research question. This often requires a thorough and structured preliminary research of the literature to map the landscape of the available evidence on a specific topic and identify the potential gaps [8, 9]. This initial assessment guides the precise formulation of the research question for the SLR. A review question formulated as a structured, single sentence, serves as the cornerstone of an SLR [8]. The creation of a well-formulated review question often demands a multidisciplinary team and careful consideration. Various approaches exist for this purpose, including the Population, Intervention, Comparator, and Outcomes (PICO) framework [10]; Sample, Phenomenon of Interest, Design, Evaluation, and Research type (SPIDER) criteria [11]; and Setting, Perspective, Intervention, Comparison, and Evaluation (SPICE) [12]. PICO is used in clinical and quantitative research, SPIDER was designed for qualitative or mixed methods research topics (to be used rather on samples than on populations), while SPICE focuses on qualitative research.

Writing the protocol: The purpose of a protocol is to plan and implement measures to minimize bias associated with the conduct and presentation of SLR findings [8]. To achieve this goal, the methods for identifying, selecting, and appraising published data of interest, as well as synthesizing and reporting the review findings, need to be elaborated and documented in advance [13, 14]. To maintain the integrity and objectivity of the SLR, deviations from or changes to the methods outlined in the protocol should be transparently communicated in review reports [15].

2.2 Conducting the SLR

Search for Studies: The SLR protocol includes references to databases where published information is searched. The literature search is conducted based on carefully devised search terminology (search terms) to ensure alignment with the concepts encapsulated in the original review question. A comprehensive search may include the so-called 'grey' literature, referring to studies published in non-peer-reviewed journals as well as unpublished data, with the aim of reducing the negative effects of publication bias [16].

Screening the Identified Studies: The protocol includes clearly defined inclusion and exclusion criteria for primary studies that align with the review question. At least two independent reviewers should screen the identified studies to make the final decision on inclusion, while the review team can resolve any differences of opinion. In an SLR, the screening process is documented using the PRISMA flowchart, which should be included in the final report to ensure transparency [17].

Appraisal of Included Studies: Assessing the quality of studies included in a SLR helps to determine the strength of the evidence. A pre-defined, structured

assessment helps to ascertain whether the included studies are reliable enough to impact healthcare or policy decisions. Different tools are available for the critical appraisal of studies with various designs, e.g., Grading of Recommendations Assessment, Development and Evaluation (GRADE) rating the certainty of evidence, Critical Appraisal Skills Program (CASP) for qualitative studies and CADAS for diagnostic accuracy studies [18-20]. Selection of evaluation tool should be adjusted to the type of the studies. When an SLR includes more than one type of study design, it is important to carefully consider the applicability of standard tools.

Data extraction: The protocol includes detailed information regarding the extent and method of data extraction with the aim of standardizing the process and improving the validity of the results. The decision on the quantity of the data to extract from the included studies is based on the objective of the SLR. To reduce the risk of errors and biases, the process should be monitored by more than one member of the reviewer team [8].

Synthesis: A systematic plan for synthesis is outlined in the protocol of a SLR, considering the review question and the anticipated types of studies. The synthesis process entails the combination and summary of the findings from individual studies included in the review. The quantitative synthesis involves statistical methods like meta-analysis, while the qualitative synthesis employs a narrative approach in case the statistical analysis is unsuitable [8, 21]. Qualitative and quantitative approaches may be applied together depending on the study characteristics. Initiating both types of synthesis should begin with the construction of a clear descriptive summary of the included studies. Texts and tables should also be included in a SLR to provide an explanatory overview of methods and findings of the involved studies.

2.3 Communicating the SLR

The primary objective of the dissemination of the findings from SLRs is to provide valuable insights to the healthcare and policy decision-makers. As a result, SLR reports are crafted with a style that is clear and accessible to a wide audience, including users and various stakeholders. SLR reports are expected to follow the recommendations outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) framework, which is specifically designed to ensure that SLR reports are written in a manner that facilitates the reproducibility of the entire review by independent researchers who were not involved in the original study [22, 23]. Transparent and comprehensive reporting are foundational principles, given that the findings of SLRs play a crucial role in shaping clinical guidelines, healthcare programs, and policies. [8].

3 Practical Points to Consider (PTC) in Carrying out a SLR of Digital Medical Devices

3.1 PTC 1: Defining the Objective of the SLR Requires Oreliminary Search of the Literature

To define a well-established study objective and perform a high-quality SLR, a preliminary literature search is needed by which previous reviews, evidence gaps, technological and regulatory developments can be identified and analyzed. This preparatory work can guarantee that the SLR will address a relevant clinical question related to the DMDs and supply patients, health care professionals and decision-makers with valuable evidence.

The number of digital technology-related clinical studies[1] and systematic reviews[2] identified via clinical filters of the PubMed database [24-26] have shown steady increase over the past years (Fig 1), with up to 35-40 new studies and 4 systematic reviews published each day.

[1] PubMed search syntax: (digital AND (((clinical[Title/Abstract] AND trial[Title/Abstract]) OR clinical trials as topic[MeSH Terms] OR clinical trial[Publication Type] OR random*[Title/Abstract] OR random allocation[MeSH Terms] OR therapeutic use[MeSH Subheading]) OR (sensitiv*[Title/Abstract] OR sensitivity and specificity[MeSH Terms] OR diagnose[Title/Abstract] OR diagnosed[Title/Abstract] OR diagnoses[Title/Abstract] OR diagnosing[Title/Abstract] OR diagnosis[Title/Abstract] OR diagnostic[Title/Abstract] OR diagnosis[MeSH:noexp] OR (diagnostic equipment[MeSH:noexp] OR diagnostic errors[MeSH:noexp] OR diagnostic imaging[MeSH:noexp] OR diagnostic services[MeSH:noexp]) OR diagnosis, differential[MeSH:noexp] OR diagnosis[Subheading:noexp]))) NOT ((Biography[pt] OR Case Reports[pt] OR Editorial[pt] OR Introductory Journal Article[pt] OR Lecture[pt] OR Legislation[pt] OR Letter[pt] OR Meta-Analysis[pt] OR News[pt] OR Newspaper Article[pt] OR Review[pt] OR Scientific Integrity Review[pt] OR Systematic Review[pt]))

[2] Pubmed search syntax: digital and (systematic[sb] OR meta-analysis[pt])

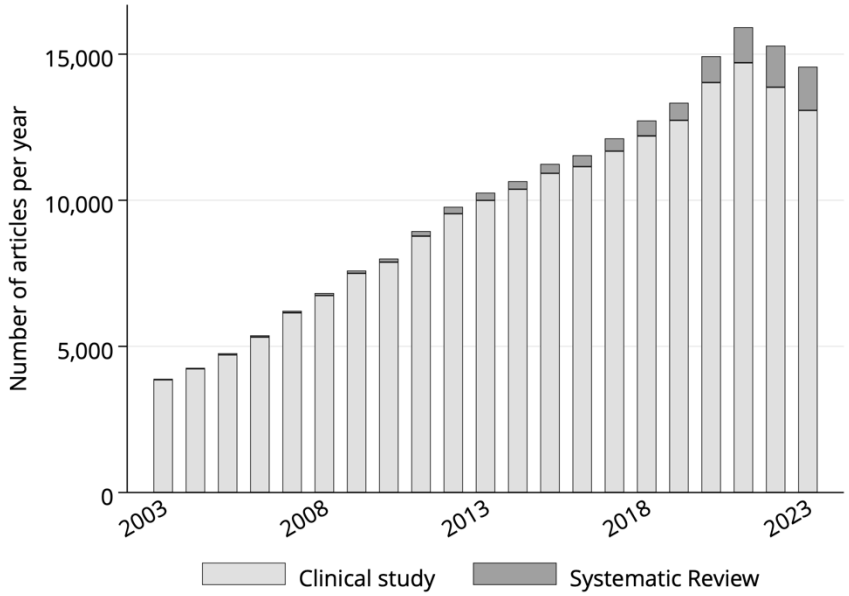

Figure 1
Distribution of clinical evidence publications on digital technologies over time in PubMed database

Although systematic reviews and meta-analyses of randomized controlled trials are considered the highest level of evidence for clinical benefits or harms [27], one needs to be aware that the methodological quality and the certainty of evidence published in SLRs are often meagre [28, 29]. While subsequent studies may change the conclusions of SLRs, which may especially be true for rapidly developing digital technologies, most published systematic reviews are not updated in a timely manner [30].

Altogether, a thorough horizon scanning and preliminary literature search can largely contribute to the ultimate value of the SLR for the patients, the health care professionals and the decision-makers.

3.2 PTC 2: Select Source Databases using a Multidisciplinary Approach

In addition to searching in large biomedical and health science databases, the use of literature databases comprising publications in engineering, informatics and computer sciences is recommended.

The development and testing DMDs require interdisciplinary knowledge from biomedical, engineering and IT fields. The relevant literature of DMDs on their development, clinical trials, post-marketing follow-up studies, and real-world data

measurements is not necessarily published in biomedical and health sciences literature databases. It can be challenging to find the best set of sources in order to identify all relevant publications. Searching a series of databases can be time consuming but ignoring potentially relevant databases may lead to biases [31]. As a rule of thumb, at least two databases should be used for the search and the publication practice of the disciplines related to the specific DMDs should be taken into account. In addition to the two large biomedical databases, MEDLINE that is free to use via PubMed platform, the more comprehensive EMBASE and the commonly searched sources (Web of Science, Scopus, CINAHL, Psychinfo), we suggest considering the IEEE Xplore database, in engineering and technology [32-35].

3.3 PTC 4: Strive to Use Validated Search Terms and Strategies

Primarily apply published and validated search filters when designing a search strategy. In case such search filters are not available, consider combining search strategies of previously published reviews or look for authoritative sources of a comprehensive terminology.

While a comprehensive and replicable search strategy is the hallmark of a systematic review method, sometimes it is challenging to construct search syntaxes that cover relevant concepts of the research question with sufficient sensitivity [8]. This is especially true in computer science and digital technologies, where technological development and the evolution of the regulatory landscape shift the mainstream terminology [36]. Digital health-related definitions proliferated considerably over time, with significant overlap between key terms [37]. Besides vague definitions, the abundance of technical terms represents another challenge in the evidence synthesis of digital technologies. In a SLR of using AI in pediatric diabetes mellitus, we identified two unique techniques per paper on average. When developing the search syntax for this study, a collection of specific methods from the caret library [38] yielded over 5 times more hits when compared to general terms, such as “machine learning” alone [39]. At studying the use of machine learning (ML) for SLR automation, we found only 163 overlapping hits among the 5321 retrieved records (3%) when combining the search criteria of published systematic reviews with similar aims [40].

To overcome these challenges, several validated search filters (hedges) have been developed for the biomedical literature to identify clinical studies on therapy, diagnosis, etiology clinical prediction or prognosis [24], systematic reviews [24], gender, age-groups, or a breadth of study designs, methods or clinical settings in medicine, including health apps or generative AI [41]. Most of these filters apply a collection of terms as well as their extensions in controlled medical vocabularies such as the Medical Subject Headings (MeSH) [10] or the Embase Emtree

Thesaurus [42], which are part of major literature databases popular in engineering science, such as Web of Science (WOS) [43] or the IEEE Xplore [35]. However, IEEE Xplore allows search using IEEE Thesaurus, a controlled vocabulary for engineering terms. A checklist is available to appraise the quality of the search filters, before their use [44].

3.4 PTC 5: Treat Automated and Semi-Automated SLR Technologies with Caution

Automation of some stages of the SLR may be attractive to save time and human resources. The performance of currently available automation methods is rather heterogenous. There are promising results on complementing human work with automation. As SLRs are used to inform decision making in healthcare, it is crucial to check the evidence on the automation method you intend to use.

SLR is a time consuming and human resource-intensive scientific work, while a comprehensive (new or updated) summary of the available clinical evidence can be a matter of urgency. Developments have been made for (semi-)automated solutions for SLRs, using AI-based solutions. Most of the SLR automation studies have been performed for record screening which is otherwise an intense manual work of at least two experienced reviewers [40]. However, the performance of currently available SLR automation methods is rather heterogeneous. For an SLR, the decisive question is whether automation of the process may achieve or even outperform the quality (sensitivity, specificity) of the human method, otherwise its potential time saving benefit will not pay off. Applying supplementary automation to increase the sensitivity of the screening could be one of the first use in practice in the foreseeable future [45].

3.5 PTC 5: Focus on Data Measured by Relevant and Validated Clinical Outcome Measures

Choose outcomes that are meaningful and understandable for both the doctors and patients, the families, the society, the policy makers and the funders. However, only reliable and validated outcome measures should be considered in your SLR.

According to the Food and Drug Administration (FDA), 'Providers and other stakeholders are using digital health technologies in their efforts to: Reduce inefficiencies, Improve access, Reduce costs, Increase quality, and Make medicine more personalized for patients' [46]. These are the goals that we aim to assess, with outcome measurements. The European MDR (2017/745) does not provide any definition and criteria of preferred measures for outcomes, but defines clinical benefit as 'the positive impact of a device on the health of an individual, expressed in terms of a meaningful, measurable, patient-relevant clinical outcome(s),

including outcome(s) related to diagnosis, or a positive impact on patient management or public health' [3]. Hence the choice of outcome measures in trial design and their assessment in SLRs are crucial points and requires a good understanding of outcome measurement. The Cochrane Handbook discusses in detail, how to select, prioritize and group outcomes in the SLR [8]. Therefore, here we highlight some essentials in the topic.

Identification and analysis of clinical and other outcomes is an important goal of SLR, but it can also be quite complex. A good example is an SLR carried out with 10 RCTs, in which 62 different outcomes were used, and selectively measured at 20 different time points [47].

Some outcomes are (relatively) easy to measure, such as laboratory tests and parameters monitored by DMDs. We can measure changes in blood pressure, but the main goal of the blood pressure therapy is to avoid hypertension-related events such as stroke and myocardial infarction. The level of blood pressure can predict the risk of these major events. Similarly, blood sugar and HbA1c level predict diabetes complications such as retinopathy, cardiovascular events, amputation. Stroke, myocardial infarction, retinopathy, cardiovascular events, amputation are examples of the so-called *hard outcomes (endpoints)* [48]. They are meaningful and well understood by the doctors, as well as other key players of the health care system such as patients, family, society, and funders. Blood pressure, blood sugar and HBA1C level are the so-called *surrogate outcomes (endpoints)*. Their correlation with hard outcomes has been established, hence can be used as input for medical decision making.

In case of DMDs, a lot of data can be and are collected, however, these can be considered as intermediate outcomes in case their relationship with hard outcomes is not supported by firm evidences [49, 50].

Another important aspect is that regulators (EMA, FDA), health policy makers, funders, clinicians, developers and managers are increasingly recognizing the importance of the patients' perspective, globally, in DMDs development [51]. The patients' perspective can be assessed by Patient Reported Outcomes (PRO), including quality of life (QoL) and well-being measures. The growing importance of PROs is demonstrated by their increasing use in Phase III clinical trials [52] and their wide adoption by medical professional guidelines. Principles for selecting, developing and adapting PRO for use in medical device evaluation was published by the FDA [53].

In the first phase of a SLR design, it is suggested to define the outcomes and their types to be identified and reviewed. Choosing a hard outcome (if available) is advisable otherwise analysis can rely on surrogate outcomes. Only validated and reliable outcome measures can provide reliable conclusions of the SLR. To guide your choice, it is suggested to check the respective literature or choose an outcome that can be found in EMA, FDA or the clinical professional guidelines, these are

most likely validated outcomes. Efforts should be made to identify and analyze PROs where available.

3.6 PTC 6: Think about Conducting Meta-Analysis of the Data

Consider combining the data from different clinical studies and perform their joint statistical analysis using meta-analysis methods, in order to increase the power and precision of the evidence. However, it is crucial to carefully follow the methodological requirements of meta-analysis.

Meta-analysis is a statistical technique to quantitatively combine results of primary studies that address the same research question, with the aim of generalizing the findings [54]. The breadth of research questions may vary from specific clinical questions concerning the same clinical population, intervention, comparator, outcome, with same timing (treatment duration and follow-up) and setting (where the study is implemented) (PICOTS) [55, 56] to a broader scope leading to general hypotheses or theories. By representing the highest level of medical evidence [27], meta-analysis is typically performed on randomized controlled trials, but it can be used to synthesize the outcomes from a greater variety of research designs (including individual patient data meta-analysis) applied in the development and appraisal of digital technologies [4], including diagnostic accuracy [57], observational studies in epidemiology [58], single-group studies [59], multi-group comparisons and indirect comparisons [60], or even single-case experimental designs [61]. Usually, the broader the topic or the more heterogenous the intervention is, the more studies should be included, but a meta-analysis can be performed on as few as two included studies. Even the so called 'empty meta-analyses' (when no studies are found in a specific topic) can be informative on research gaps in a given field [54]. Meta-analyses allow the testing of hypotheses that would not be feasible in single-study settings [54].

Our search for digital technology-related systematic reviews with[3] and without[4] the term 'meta-analysis' suggests that an increasing proportion of studies are published without including a meta-analysis of the outcomes. In 2023, 57% of the published systematic reviews in the field missed a meta-analysis (Fig. 2).

[3] PubMed search syntax: digital AND (systematic[sb] or meta-analysis[pt]) AND meta-analysis

[4] PubMed search syntax: digital AND (systematic[sb] or meta-analysis[pt]) NOT meta-analysis

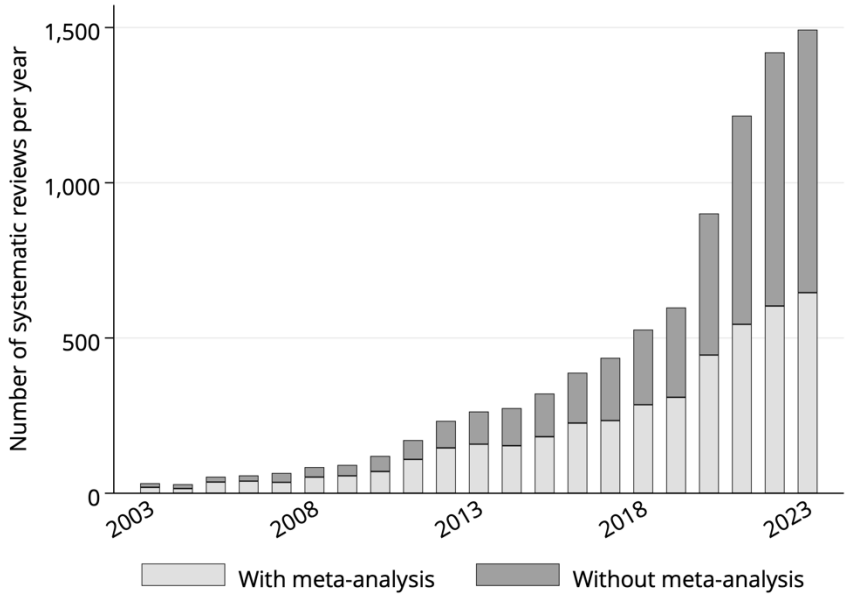

Figure 2
Distribution of systematic reviews with and without meta-analysis in PubMed bibliographic records over time

Reviewers often conclude that due to the heterogeneity of patients, interventions or study designs, conducting a meta-analysis would not be feasible and provide only qualitative summary of the included studies. In such cases we found that meta-analyses were feasible involving two out of eight randomized controlled trials of chatbots [62], or involving the proportions of successful cases in single-arm studies of augmented-reality-navigated spine surgery [63].

Altogether, while being aware the potential biases and pitfalls of meta-analyses [64], we encourage authors to strive to perform meta-analysis of the clinical outcomes of DMDs studies. While the scope of some meta-analyses involving heterogenous technologies or populations may be too broad to address particular clinical questions, quantitative summaries may provide unique insights or benchmarks that may fuel future developments in the field. Furthermore, PICOTS-ComTeC (Population, Intervention, Comparator, Outcome, Timing, Setting, Communication, Technology and Context), is a novel research framework, developed for patient-facing digital health interventions, may facilitate evidence syntheses in the field, by aiding the identification of comparable interventions and the selection of appropriate comparators, that deliver similar effects for patients [65].

3.7 PTC 7: Use the GRADE System for the Assessment of the Quality of a Body of Evidence

In order to draw reliable conclusions from the SLR, to support medical decision making, the quality of original clinical studies providing evidence on a health intervention should be evaluated. The GRADE approach has been developed for this purpose and is suggested for use in SLRs of DMDs.

The abbreviation GRADE refers to 'Grading of Recommendation, Assessment, Development and Evaluation', an established method to evaluate and weight the clinical data revealed by the SLR [66]. The Cochrane Collaboration has adopted this method and uses the GRADE approach to assess the evidence on outcomes in clinical studies of interventions [8]. This stage is often missing from published SLRs on DMDs, presumably, because the assessment requires broad understanding of epidemiology, clinical study types, the PICO system, the outcome measurement and further details. These may be beyond the average knowledge of a researcher working in the field of DMDs, hence we suggest the involvement of an experienced researcher.

3.8 PTC 8: Use Reporting Guidelines to Report your SLR

Good quality reporting will make a real use of your results. Reporting guidelines are useful tools to present your SLR in a standardized, transparent manner, as well as to check its accuracy and completeness. The PRISMA 2020 statement is considered as the gold standard reporting guideline for SLRs, its extensions for specialized topics and other checklists also deserve attention.

Poorly reported medical research generates immense waste of resources [67], while the adoption of reporting checklists have been shown to improve reporting quality [68, 69]. The core definition of SLRs includes the use of transparent and reproducible methods for searching source documents and the synthesis of results [70]. Hence, adherence to reporting guidelines is inherent part of the systematic review process.

We note, that while detailed methodological guidelines exist on the implementation of SLRs and meta-analyses [64], reporting guidelines focus on the items that should be included in publication, without providing methodological guidance on how to perform the respective research step. For systematic reviews and meta-analyses, the updated PRISMA 2020 statement is considered as the gold standard reporting guideline [22]. PRISMA has several extensions for specialized systematic reviews [71].

In the context of DMDs, further systematic review reporting checklists may be relevant such as the TRIPOD-SRMA for multivariable prediction models for individual prognosis or diagnosis [72], CHARMS for prediction modelling studies

[73], MOOSE for meta-analyses of observational studies [58], and the TECH framework for health app reviews [74]. Moreover, PRIOR statement guides the reporting of overviews of systematic reviews [75].

New reporting guidelines and checklists are being continuously developed and published, and the number of potentially relevant checklists for a given publication may be overwhelming [39]. To avoid confusion, we suggest that authors consult the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) network's website [71], the target journal's requirements and some benchmark reviews in their field that were published in the leading journals, when selecting the appropriate reporting guideline for their review.

3.9 PTC 9: SLRs on AI-based Chatbots in Healthcare

Validated search strategies for AI-based healthcare chatbots are lacking. Clinical studies are relatively scarce and their reports are often poorly standardized. Hence SLRs focusing on AI-based healthcare chatbots require special caution both in terms of literature search and evaluation of the results.

The literature search for studies on AI-based healthcare chatbots is challenging due to the lack of validated search terms and strategies for 'chatbot' and also for 'AI' [76]. SLRs used different 'chatbot' search terms (e.g., virtual/embodied agent, conversational bot/interface, virtual assistant / coach, nursing avatar, social/smart bot, etc.). Combinations of search terms from relevant SLRs can be a pragmatic solution but might lead to a large number of irrelevant hits. A recent SLR on conversational agents in healthcare identified eight randomized controlled trials [77]. Authors missed to perform meta-analysis although results of two randomized controlled trials on two relevant clinical outcomes were proved to be appropriate for the analysis [62]. Another important issue is the transparent reporting of the chatbot technology used. Risk of bias assessment of the trials revealed a rather low risk in general, however a transparent reporting of the AI method used was often missing. According to our best knowledge, specific reporting guidelines for trials and SLRs on AI-based healthcare chatbots are missing [39]. We suggest, that AI-chatbot researchers, should rely on well-established reporting guidelines on different study types and consider the available checklists for AI-based studies.

3.10 PTC 10: SLRs on Robotic Surgery

Always assess the risk arising from the autonomy of a surgical robot system. Other key points to consider may be identical to general innovative medical electrical equipment, yet AI and ML methods are pushing towards more autonomous systems.

Standardizing DMDs has been often regulated by large organizations, particularly the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). To cover the PTC above, *ISO 13485:2016 - Quality management systems* and the *IEC 60601-1 – Medical electrical equipment,* general standards of the domain, have been amended with the *IEC/TR 60601-4-1: Medical electrical equipment – Part 4-1: Guidance and interpretation – Medical electrical equipment and medical electrical systems employing a degree of autonomy* and the *IEC 80601-2-77: Particular requirements for the basic safety and essential performance of robotically assisted surgical equipment* [78, 79]. These new standards bridge the gap between the traditional approach of treating medical devices (i.e., Medical Electrical Equipment (MEE) and Medical Electrical Systems (MES) in the standard's taxonomy) separate from robots (falling under the machinery directives) [80]. It has been clearly defined that an MEE/MES can be a robot, while still being regulated as a medical device, with a certain degree of autonomy. The development and application of AI-based and ML methods, in robot-assisted surgeries, requires well-defined criteria for validation [81, 82]. In addition, methods that can deal with data heterogeneity, as well as sparsity and real-time capability are needed, as well as taking into consideration the aspects of the sustainability of large-scale computing [83-85]. This requires real-time control and novel communication networks with low latency and high resilience in the OR [86, 87].

Conclusions

We have summarized 10 points to consider for SLRs, in the field of DMDs from an explicitly practical perspective. We have relied on our scientific knowledge in systematic literature reviews and experience in the field of DMDs. We have not aimed for completeness, in contrast, we encourage researchers to develop these points further, expand with new aspects and apply them to DMDs in general and to specific DMD assets. Our suggestions provide a valuable basis for that work. Nonetheless, we believe that this set of PTC is a valuable tool for those (DMD developers, researchers, regulatory agencies, business development and sales managers, policy makers, clinicians, patients), who are interested in providing and obtaining reliable conclusions on the clinical evidence concerning DMDs, thereby, improving the overall health of patients. Our work is intended to help in achieving this goal.

Acknowledgements

This project has been supported by the National Research, Development, and Innovation Fund of Hungary, financed under the TKP2021-NKTA-36 funding scheme, at Óbuda University. T. Haidegger is a Consolidator Researcher, receiving financial support from the Distinguished Researcher program of Óbuda University.

References

[1] B. Haynes, "Can it work? Does it work? Is it worth it? The testing of healthcareinterventions is evolving," *BMJ,* vol. 319, no. 7211, pp. 652-3, Sep 11 1999, doi: 10.1136/bmj.319.7211.652.

[2] "Cochrane." https://www.cochrane.org/ (accessed January 31, 2024).

[3] E. U. a. E. Council. "Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.)." https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745 (accessed 31 January, 2024).

[4] V. Zah, A. Burrell, C. Asche, and Z. Zrubka, "Paying for Digital Health Interventions – What Evidence is Needed?," *Acta Polytechnica Hungarica,* vol. 19, no. 9, pp. 179-199, 2022, doi: 10.12700/aph.19.9.2022.9.10.

[5] K. Feldman, "Using the Work of Others: Some Observations on Reviewing and Integrating," *Sociology of Education,* vol. 44, no. 1, pp. 86-102, 1971.

[6] H. Cooper, L. V. Hedges, and J. C. Valentine, *The Handbook of Research Synthesis and Meta-Analysis*. Russel Sage Foundation, 2019.

[7] K. S. Khan, R. Kunz, J. Kleijnen, and G. Antes, "Five steps to conducting a systematic review," *J R Soc Med,* vol. 96, no. 3, pp. 118-21, Mar 2003, doi: 10.1177/014107680309600304.

[8] J. Higgins *et al.*, *Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023)*. Available from: https://www.training.cochrane.org/handbook: Cochrane, 2023.

[9] E. Aromataris and Z. Munn, *JBI Manual for Evidence Synthesis*. JBI, 2020.

[10] M. B. Eriksen and T. F. Frandsen, "The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review," *J Med Libr Assoc,* vol. 106, no. 4, pp. 420-431, Oct 2018, doi: 10.5195/jmla.2018.345.

[11] A. Cooke, D. Smith, and A. Booth, "Beyond PICO: the SPIDER tool for qualitative evidence synthesis," *Qual Health Res,* vol. 22, no. 10, pp. 1435-43, Oct 2012, doi: 10.1177/1049732312452938.

[12] N. Jahan, S. Naveed, M. Zeshan, and M. A. Tahir, "How to Conduct a Systematic Review: A Narrative Literature Review," *Cureus,* vol. 8, no. 11, p. e864, Nov 4 2016, doi: 10.7759/cureus.864.

[13] L. Shamseer *et al.*, "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation," *BMJ,* vol. 350, p. g7647, Jan 2 2015, doi: 10.1136/bmj.g7647.

[14] S. Straus and D. Moher, "Registering systematic reviews," *CMAJ,* vol. 182, no. 1, pp. 13-4, Jan 12 2010, doi: 10.1503/cmaj.081849.

[15] A. M. Drucker, P. Fleming, and A. W. Chan, "Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews," *J Invest Dermatol,* vol. 136, no. 11, pp. e109-e114, Nov 2016, doi: 10.1016/j.jid.2016.08.021.

[16] A. Paez, "Grey literature: An important resource in systematic reviews," *J Evid Based Med,* Dec 21 2017, doi: 10.1111/jebm.12265.

[17] S. Waffenschmidt, M. Knelangen, W. Sieben, S. Buhn, and D. Pieper, "Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review," *BMC Med Res Methodol,* vol. 19, no. 1, p. 132, Jun 28 2019, doi: 10.1186/s12874-019-0782-0.

[18] S. E. Brennan and R. V. Johnston, "Research Note: Interpreting findings of a systematic review using GRADE methods," *J Physiother,* vol. 69, no. 3, pp. 198-202, Jul 2023, doi: 10.1016/j.jphys.2023.05.012.

[19] H. Long, D. French, and J. Brooks, "Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis.," *Research Methods in Medicine & Health Sciences,* vol. 1, no. 1, pp. 31-42, 2020, doi: 10.1177/2632084320947559

[20] P. F. Whiting *et al.*, "QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies," *Ann Intern Med,* vol. 155, no. 8, pp. 529-36, Oct 18 2011, doi: 10.7326/0003-4819-155-8-201110180-00009.

[21] J. Thomas and A. Harden, "Methods for the thematic synthesis of qualitative research in systematic reviews," *BMC Med Res Methodol,* vol. 8, p. 45, Jul 10 2008, doi: 10.1186/1471-2288-8-45.

[22] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *Syst Rev,* vol. 10, no. 1, p. 89, Mar 29 2021, doi: 10.1186/s13643-021-01626-4.

[23] I. o. M. U. C. o. S. f. S. R. o. C. E. Research, "Standards for Reporting Systematic Reviews," in *Finding What Works in Health Care: Standards for Systematic Reviews*, J. Eden, L. Levit, and A. Berg Eds.: National Academy Press (US), 2011.

[24] National Library of Medicine. "Search Strategy Used to Create the PubMed Systematic Reviews Filter." National Library of Medicine https://www.nlm.nih.gov/bsd/pubmed_subsets/sysreviews_strategy.html (accessed January 27, 2024.

[25] R. B. Haynes, K. A. McKibbon, N. L. Wilczynski, S. D. Walter, S. R. Werre, and T. Hedges, "Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey,"

BMJ, vol. 330, no. 7501, p. 1179, May 21 2005, doi: 10.1136/bmj.38446.498542.8F.

[26] R. B. Haynes and N. L. Wilczynski, "Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey," *BMJ,* vol. 328, no. 7447, p. 1040, May 1 2004, doi: 10.1136/bmj.38068.557998.EE.

[27] J. Howick *et al.*, "The Oxford Levels of Evidence 2," Oxford, 2011. Accessed: January 27, 2024. [Online]. Available: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence

[28] K. T. McDermott *et al.*, "Digital interventions to moderate physical inactivity and/or nutrition in young people: a Cancer Prevention Europe overview of systematic reviews," *Front Digit Health,* vol. 5, p. 1185586, 2023, doi: 10.3389/fdgth.2023.1185586.

[29] H. Motahari-Nezhad *et al.*, "Digital Biomarker-Based Interventions: Systematic Review of Systematic Reviews," *J Med Internet Res,* vol. 24, no. 12, p. e41042, Dec 21 2022, doi: 10.2196/41042.

[30] D. Moher *et al.*, "When and how to update systematic reviews," *Cochrane Database Syst Rev,* vol. 2008, no. 1, p. MR000023, Jan 23 2008, doi: 10.1002/14651858.MR000023.pub3.

[31] W. M. Bramer, M. L. Rethlefsen, J. Kleijnen, and O. H. Franco, "Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study," *Syst Rev,* vol. 6, no. 1, p. 245, Dec 6 2017, doi: 10.1186/s13643-017-0644-y.

[32] N. L. o. Medicine. "PubMed Overview." https://pubmed.ncbi.nlm.nih.gov/about/ (accessed 31 January, 2024).

[33] Elsevier. "EMBASE." https://www.elsevier.com/products/embase (accessed 31 January, 2024).

[34] T. F. Frandsen, M. B. Eriksen, D. M. G. Hammer, J. B. Christensen, and J. A. Wallin, "Using Embase as a supplement to PubMed in Cochrane reviews differed across fields," *J Clin Epidemiol,* vol. 133, pp. 24-31, May 2021, doi: 10.1016/j.jclinepi.2020.12.022.

[35] IEEE. "IEEE Xplore." https://ieeexplore.ieee.org/Xplorehelp/overview-of-ieee-xplore/about-ieee-xplore (accessed 31 January, 2024).

[36] J. Torous, A. D. Stern, and F. T. Bourgeois, "Regulatory considerations to keep pace with innovation in digital health products," *NPJ Digit Med,* vol. 5, no. 1, p. 121, Aug 19 2022, doi: 10.1038/s41746-022-00668-9.

[37] A. Burrell *et al.*, "How Useful Are Digital Health Terms for Outcomes Research? An ISPOR Special Interest Group Report," *Value Health,* vol. 25, no. 9, pp. 1469-1479, Sep 2022, doi: 10.1016/j.jval.2022.04.1730.

[38] M. Kuhn. "caret: classification and regression training. The Comprehensive R Archive Network." http://cran.nexr.com/web/packages/caret/index.html (accessed February 10, 2023.

[39] Z. Zrubka *et al.*, "The Reporting Quality of Machine Learning Studies on Pediatric Diabetes Mellitus: Systematic Review," *J Med Internet Res,* vol. 26, p. e47430, Jan 19 2024, doi: 10.2196/47430.

[40] B. Tóth, L. Berek, L. Gulácsi, M. Péntek, and Z. Zrubka, "Automation of systematic reviews of biomedical literature: a systematic review of studies indexed in PubMed," Preprint, October 10, 2023, 2023.

[41] I. I. S. Sub-Group. "The ISSG Search Filter Resource." https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home (accessed January 30, 2024.

[42] "Embase Emtree thesaurus." Embase. https://www.elsevier.com/products/embase/emtree (accessed January 30, 2024.

[43] "Web of Science." Clarivate. https://www.webofscience.com/wos/woscc/basic-search (accessed January 30, 2024.

[44] J. Glanville *et al.*, "So many filters, so little time: the development of a search filter appraisal checklist," *J Med Libr Assoc,* vol. 96, no. 4, pp. 356-61, Oct 2008, doi: 10.3163/1536-5050.96.4.011.

[45] V. V. Odintsova *et al.*, "Genomics of human aggression: current state of genome-wide studies and an automated systematic review tool," *Psychiatr Genet,* vol. 29, no. 5, pp. 170-190, Oct 2019, doi: 10.1097/YPG.0000000000000239.

[46] U. S. F. Drug. "What is digital health?" https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health (accessed 31 January, 2024).

[47] L. Gulacsi *et al.*, "Long-Term Efficacy of Tumor Necrosis Factor Inhibitors for the Treatment of Methotrexate-Naive Rheumatoid Arthritis: Systematic Literature Review and Meta-Analysis," *Adv Ther,* vol. 36, no. 3, pp. 721-745, Mar 2019, doi: 10.1007/s12325-018-0869-8.

[48] O. Ciani *et al.*, "A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials," *EClinicalMedicine,* vol. 65, p. 102283, Nov 2023, doi: 10.1016/j.eclinm.2023.102283.

[49] D. L. DeMets, B. M. Psaty, and T. R. Fleming, "When Can Intermediate Outcomes Be Used as Surrogate Outcomes?," *JAMA,* vol. 323, no. 12, pp. 1184-1185, Mar 24 2020, doi: 10.1001/jama.2020.1176.

[50] T.-A. Tran *et al.*, " Tran, Tuan-Anh ; Péntek, Márta ; Motahari-Nezhad, Hossein ; Abonyi, János ; Kovács, Levente ; Gulácsi, László ; Eigner,

György ; Zrubka, Zsombor ; Ruppert, Tamás " *IEEE Transactions on Systems, Man and Cyberbetics: Systems,* no. 53, pp. 6685-6692, 2023, doi: 10.1109/TSMC.2023.3282141.

[51] U. S. F. D. Administration. "Clinical Outcome Assessments (COAs) in Medical Device Decision Making." https://www.fda.gov/about-fda/division-patient-centered-development/clinical-outcome-assessments-coas-medical-device-decision-making (accessed 31 January, 2024).

[52] A. Gnanasakthy, A. Barrett, E. Evans, D. D'Alessio, and C. D. Romano, "A Review of Patient-Reported Outcomes Labeling for Oncology Drugs Approved by the FDA and the EMA (2012-2016)," *Value Health,* vol. 22, no. 2, pp. 203-209, Feb 2019, doi: 10.1016/j.jval.2018.09.2842.

[53] O. o. S. P. a. T. I. Patient Science and Engagement Program. "Principles for Selecting, Developing, Modifying, and Adapting Patient-Reported Outcome Instruments for Use in Medical Device Evaluation." https://www.fda.gov/media/141565/download (accessed 31 January, 2024).

[54] H. Cooper, L. V. Hedges, and J. C. Valentine, *The Handbook of Research Synthesis and Meta-Analysis, 3rd Edition*. New York: Russell Sage Foundation, 2019.

[55] E. P. Whitlock, S. A. Lopez, S. Chang, M. Helfand, M. Eder, and N. Floyd, "AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program," *J Clin Epidemiol,* vol. 63, no. 5, pp. 491-501, May 2010, doi: 10.1016/j.jclinepi.2009.03.008.

[56] C. Counsell, "Formulating questions and locating primary studies for inclusion in systematic reviews," *Ann Intern Med,* vol. 127, no. 5, pp. 380-7, Sep 1 1997, doi: 10.7326/0003-4819-127-5-199709010-00008.

[57] G. Biondi-Zoccai, *Diagnostic Meta-Analysis*. Cham: Springer Cham, 2018.

[58] D. F. Stroup *et al.*, "Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group," *JAMA,* vol. 283, no. 15, pp. 2008-12, Apr 19 2000, doi: 10.1001/jama.283.15.2008.

[59] T. H. Barker *et al.*, "Conducting proportional meta-analysis in different types of systematic reviews: a guide for synthesisers of evidence," *BMC Med Res Methodol,* vol. 21, no. 1, p. 189, Sep 20 2021, doi: 10.1186/s12874-021-01381-z.

[60] D. C. Hoaglin *et al.*, "Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2," *Value Health,* vol. 14, no. 4, pp. 429-37, Jun 2011, doi: 10.1016/j.jval.2011.01.011.

[61] E. Baek, W. Luo, and K. H. Lam, "Meta-Analysis of Single-Case Experimental Design using Multilevel Modeling," *Behav Modif,* vol. 47, no. 6, pp. 1546-1573, Nov 2023, doi: 10.1177/01454455221144034.

[62] J. T. Czere and M. Péntek, "Exploring the Feasibility of the Meta-Analysis of Randomized Controlled Trials on Artificial Intelligence Chatbots for Use in Healthcare Based on a Published Systematic Review," in *2022 IEEE 20th Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*, 15-17 Sept. 2022 2022, pp. 000053-000058, doi: 10.1109/SISY56759.2022.10036294. [Online]. Available: https://ieeexplore.ieee.org/document/10036294/

[63] K. Moga, A. Holgyesi, Z. Zrubka, M. Pentek, and T. Haidegger, "Augmented or Mixed Reality Enhanced Head-Mounted Display Navigation for In Vivo Spine Surgery: A Systematic Review of Clinical Outcomes," *J Clin Med,* vol. 12, no. 11, May 31 2023, doi: 10.3390/jcm12113788.

[64] D. JJ, H. JPT, and A. DG, "Chapter 10: Analysing data and undertaking meta-analyses," in *Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023)*, H. JPT *et al.* Eds. Available from: https://www.training.cochrane.org/handbook: Cochrane, 2023.

[65] Z. Zrubka *et al.*, "The PICOTS-ComTeC Framework for Defining Digital Health Interventions: An ISPOR Special Interest Group Report," *Value in Health,* vol. (In press), 2024.

[66] H. Schünemann, J. Brożek, G. Guyatt, and A. Oxman. "GRADE Handbook." https://gdt.gradepro.org/app/handbook/handbook.html#h.xivvyiu1pr3v (accessed 31 January, 2024).

[67] I. Chalmers and P. Glasziou, "Avoidable waste in the production and reporting of research evidence," *Lancet,* vol. 374, no. 9683, pp. 86-9, Jul 4 2009, doi: 10.1016/S0140-6736(09)60329-9.

[68] A. C. Plint *et al.*, "Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review," *Med J Aust,* vol. 185, no. 5, pp. 263-7, Sep 4 2006, doi: 10.5694/j.1326-5377.2006.tb00557.x.

[69] R. A. Agha *et al.*, "Impact of the mandatory implementation of reporting guidelines on reporting quality in a surgical journal: A before and after study," *Int J Surg,* vol. 30, pp. 169-72, Jun 2016, doi: 10.1016/j.ijsu.2016.04.032.

[70] M. Krnic Martinic, D. Pieper, A. Glatt, and L. Puljak, "Definition of a systematic review used in overviews of systematic reviews, meta-

epidemiological studies and textbooks," *BMC Med Res Methodol,* vol. 19, no. 1, p. 203, Nov 4 2019, doi: 10.1186/s12874-019-0855-0.

[71] E. Network. "Enhancing the QUAlity and Transparency Of health Research." UK EQUATOR Centre. (accessed January 24, 2024.

[72] K. I. E. Snell *et al.*, "Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA)," *BMJ,* vol. 381, p. e073538, May 3 2023, doi: 10.1136/bmj-2022-073538.

[73] K. G. Moons *et al.*, "Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist," *PLoS Med,* vol. 11, no. 10, p. e1001744, Oct 2014, doi: 10.1371/journal.pmed.1001744.

[74] N. Gasteiger *et al.*, "Conducting a systematic review and evaluation of commercially available mobile applications (apps) on a health-related topic: the TECH approach and a step-by-step methodological guide," *BMJ Open,* vol. 13, no. 6, p. e073283, Jun 12 2023, doi: 10.1136/bmjopen-2023-073283.

[75] M. Gates *et al.*, "Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement," *BMJ,* vol. 378, p. e070849, Aug 9 2022, doi: 10.1136/bmj-2022-070849.

[76] J. Czere, L. Gulácsi, Z. Zrubka, and M. Péntek, "Quality assessment of clinical trials with artificial intelligence based chatbots in healthcare: points to consider in the protocol development for a systematic literature review," presented at the 2022 IEEE 16th International Symposium on Applied Computational Intelligence and Informatics (SACI), 25-28 May 2022, 2022.

[77] M. Milne-Ives *et al.*, "The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review," *J Med Internet Res,* vol. 22, no. 10, p. e20346, Oct 22 2020, doi: 10.2196/20346.

[78] G. Fichtinger, J. Troccaz, and T. Haidegger, "Image-Guided Interventional Robotics: Lost in translation?," *Proceedings of the IEEE,* vol. 110, no. 7, pp. 932-950, 2022, doi: 10.1109/JPROC.2022.3166253.

[79] E. Prestes *et al.*, "The First Global Ontological Standard for Ethically Driven Robotics and Automation Systems [Standards]," *IEEE Robotics & Automation Magazine,* vol. 28, no. 4, pp. 120-124, 2021, doi: 10.1109/MRA.2021.3117414.

[80] T. Haidegger, "Taxonomy and Standards in Robotics," in *Encyclopedia of Robotics* B. Siciliano, O. Khatib, M. H. Ang, and Eds. Dordrecht, Netherlands: Springer Netherlands, 2022, ch. 190-1, pp. 1-10.

[81] T. Nagy and T. Haidegger, "Performance and Capability Assessment in Surgical Subtask Automation," *Sensors,* vol. 22, no. 7, p. 2501, 2022, doi: https://doi.org/10.3390/s22072501.

[82] M. A. Houghtaling *et al.*, "Standardizing an Ontology for Ethically Aligned Robotic and Autonomous Systems," *IEEE Transactions on Systems Mand and Cybernetics: Systems,* vol. 54, no. 3, pp. 1791-1804, 2024, doi: 10.1109/TSMC.2023.3330981.

[83] R. Nagyné Elek and T. Haidegger, "Non-Technical Skill Assessment and Mental Load Evaluation in Robot-Assisted Minimally Invasive Surgery " vol. 21, no. 8, p. 2666, 2021, doi: 10.3390/s21082666.

[84] R. Nagyné Elek and T. Haidegger, "Next in Surgical Data Science: Autonomous Non-Technical Skill Assessment in Minimally Invasive Surgery Training," *Journal of Clinical Medicine,* vol. 11, no. 24, p. 7533, 2022, doi: 10.3390/jcm11247533.

[85] T. Haidegger *et al.*, "Robotics: Enabler and inhibitor of the Sustainable Development Goals," *Sustainable Production and Consumption,* vol. 43, pp. 422-434, 2023, doi: https://doi.org/10.1016/j.spc.2023.11.011.

[86] T. Haidegger, S. Speidel, D. Stoyanov, and R. M. Satava, "Robot-Assisted Minimally Invasive Surgery—Surgical Robotics in the Data Age," *Proceedings of the IEEE,* vol. 110, no. 7, pp. 835-846, 2022, doi: https://doi.org/10.1109/JPROC.2022.3180350.

[87] R. C. Anderson *et al.*, "Next frontier in planetary geological reconnaissance: Low-latency telepresence," *Icarus,* vol. 368, p. 114558, 2021, doi: https://doi.org/10.1016/j.icarus.2021.114558.