

ORIGINAL RESEARCH

# Physicians found an interactive tool displaying structured evidence summaries for multiple comparisons understandable and useful: a qualitative user testing study

Per Olav Løvsletten<sup>a,b</sup>, Birk Stokke Hunskar<sup>b</sup>, Anja Fog Heen<sup>a,c</sup>, Geertruida Bekkering<sup>d,e,f</sup>, Louise Vanden Poel<sup>d</sup>, Dena Zeraatkar<sup>g</sup>, Mieke Vermandere<sup>d</sup>, Bert Aertgeerts<sup>d,e</sup>, Nicolas Delvaux<sup>d</sup>, Frankie Achille<sup>c</sup>, Jason W. Busse<sup>h,i</sup>, Thomas Agoritsas<sup>c,j,k</sup>, Per Olav Vandvik<sup>a,b,c,\*</sup>

<sup>a</sup>Department of Medicine, Lovisenberg Diaconal Hospital, Oslo, Norway

<sup>b</sup>Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway

<sup>c</sup>MAGIC Evidence Ecosystem Foundation, Oslo, Norway

<sup>d</sup>Academic Centre for General Practice, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

<sup>e</sup>Belgian Centre for Evidence Based Medicine (Cebam), Leuven, Belgium

<sup>f</sup>Cochrane Belgium, Leuven, Belgium

<sup>g</sup>Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

<sup>h</sup>Department of Anaesthesia, McMaster University, Hamilton, Canada

<sup>i</sup>Department of Health, Evidence and Impact, McMaster University, Hamilton, Canada

<sup>j</sup>Department of Internal Medicine, University Hospitals of Geneva, Geneva, Switzerland

<sup>k</sup>Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

Accepted 21 May 2024; Published online 27 May 2024

## Abstract

**Objectives:** To evaluate and improve “Making Alternative Treatment Choices Intuitive and Trustworthy” (MATCH-IT)—a digital, interactive decision support tool displaying structured evidence summaries for multiple comparisons—to help physicians interpret and apply evidence from network meta-analysis (NMA) for their clinical decision-making.

**Study Design and Setting:** We conducted a qualitative user testing study, applying principles from user-centered design in an iterative development process. We recruited a convenience sample of practicing physicians in Norway, Belgium, and Canada, and asked them to interpret structured evidence summaries for multiple comparisons—linked to clinical guideline recommendations—displayed in MATCH-IT. User testing included (a) introduction of a clinical scenario, (b) a think-aloud session with participant–tool interaction, and (c) a semistructured interview. We video recorded, transcribed, and analyzed user tests using directed content analysis. The results informed new updates in MATCH-IT.

**Results:** Distributed across 5 development cycles we tested MATCH-IT with 26 physicians. Of these, 24 (94%) reported either no or sparse prior experience with interpretation of NMA. Physicians perceived MATCH-IT as easy to interpret and navigate, and appreciated its ability to provide an overview of the evidence. Visualization of effects in pictograms and inclusion of information on burden of treatment (“practical issues”) were highlighted as potentially useful features in interacting with patients. We also identified problems, including undiscovered functionalities (drag and drop), suboptimal tutorial, and cumbersome navigation of the tool. In addition, physicians wanted definition/explanation of key terms (eg, outcomes and “certainty”), and there were concerns that overwhelming evidence from a large NMA would complicate applicability to clinical practice. This led to several updates with development of a new start page, tutorial, updated user interface for more efficient maneuvering, solutions to display definition of key terms and a “frequently asked questions” section. To facilitate interpretation of large networks, we improved categorization of results using color coding and added filtering functionality. These modifications allowed physicians to focus on interventions of interest and reduce information overload.

**Funding:** This work was supported by a PhD grant from South-Eastern Norway Regional Health Authority (grant number 2017015) and a Medical Student Research Program Grant from The Research Council of Norway (grant number 271555/F20). In addition, MAGIC has contributed with in-kind resources for the web designer and subsequent programming and development of the MATCH-IT prototype.

**Disclaimer:** None of the funding bodies of this study—The South-Eastern Norway Regional Health Authority and The Research Council of Norway—were involved in planning, execution, or reporting of this study.

\* Corresponding author. Department of Medicine, Lovisenberg Diaconal Hospital, Lovisenberggata 17C, 0456 Oslo, Norway.

E-mail address: [perolav@magicvidence.org](mailto:perolav@magicvidence.org) (P.O. Vandvik).

**Conclusion:** This study provides proof of concept that physicians can use MATCH-IT to understand NMA evidence. Key features of MATCH-IT in a clinical context include providing an overview of the evidence, visualization of effects, and the display of information on burden of treatments. However, unfamiliarity with the Grading of Recommendations Assessment, Development and Evaluation concepts, time constraints, and accessibility at the point of care may be challenges for use. To what extent our results are transferable to real-world clinical contexts remains to be explored. © 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Keywords:** Decision support tool; Structured evidence summaries; Network meta-analysis; User testing; Health-care professionals; Multiple treatment options; User-centred design

### Plain language summary

This study looked at MATCH-IT, a digital interactive tool designed to help physicians understand and use research findings from overviews of medical literature (systematic reviews) summarizing results from multiple primary studies that altogether compare 3 or more medical treatments (network meta-analysis).

We tested the tool with 26 physicians from Norway, Belgium, and Canada. The physicians found MATCH-IT easy to use and liked that it gave a clear summary of the research results. However, they sometimes did not discover all features of the tool and were also worried that comparing many treatments could be overwhelming.

Based on their feedback, we made updates like adding a better tutorial, improving how to navigate the tool, and organizing the evidence with filters and color codes which was beneficial for the use and understanding of the tool.

In conclusion, MATCH-IT appears to be helpful for physicians to understand complex research information, but how the tool function in a real clinical context remains to be seen and will be investigated in future studies.

## 1. Introduction

Evidence-based decisions should be informed by the best current research evidence, integrated with clinical expertise and patients' values and preferences [1]. For most clinical conditions, multiple treatment options are available and the use of network meta-analysis (NMA) allow statistical comparisons of all these treatment options in one analysis using a combination of direct and indirect evidence [2–9]. To navigate the volume and complexity of the rapidly evolving body of evidence from NMAs, physicians and patients need trustworthy guidance of clinical practice guidelines (CPGs) and digital decision support tools adhering to these principles [10,11].

Systematic reviews (SRs) with NMA increasingly inform CPGs [2,12–14], and results should be reported in *structured evidence summaries*—displaying the absolute effect differences between interventions across benefits and harms with corresponding certainty ratings according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, and ranking of interventions [2,15–17]. Structured evidence summaries of pairwise comparisons, often reported as GRADE Summary of Findings (SoF) tables, have been shown to facilitate understanding and efficient use of the evidence [10,18–22]. However, for multiple comparisons, the overwhelming amount of results have raised daunting challenges in how to visually present comprehensible structured evidence summaries [2]. Currently, few SRs with NMA report structured evidence

summaries [23], and there is a lack of validated decision support tools for multiple comparisons as well.

Acknowledging these challenges, researchers from the nonprofit organization MAGIC Evidence Ecosystem Foundation (<https://www.magicvidence.org>) have developed a digital and interactive decision support tool, henceforth called “Making Alternative Treatment Choices Intuitive and Trustworthy” (MATCH-IT). The tool is designed to display structured evidence summaries for multiple comparisons from NMAs, and is already published as part of SRs and CPGs [4,12,13,24]. Although the evidence is inconclusive regarding whether interactivity improve comprehension [25–29], the tool draws upon previous related projects, which have underscored the value of presenting evidence in adjustable formats [21,30,31]. In this study we aimed to evaluate and further improve MATCH-IT to help physicians interpret and apply NMA evidence for their clinical decision-making.

## 2. Study design and setting

### 2.1. Project management and study design

We conducted qualitative user testing of MATCH-IT with physicians in 3 countries, in collaboration with researchers at KU Leuven (Belgium) and McMaster University (Canada). We obtained ethical and data privacy approvals in all countries and legal contracts regulating

**What is new?****Key findings**

- Practicing physicians found an interactive tool (“Making Alternative Treatment Choices Intuitive and Trustworthy” (MATCH-IT)) that dynamically displays structured evidence summaries for multiple comparisons understandable and useful.

**What this adds to what was known?**

- Interactive Summary of Findings tables for multiple comparisons may enhance physicians understanding of complex bodies of evidence.
- In a clinical setting, evidence from network meta-analyses should be categorized and prefiltered to reduce information overload.

**What is the implication and what should change now?**

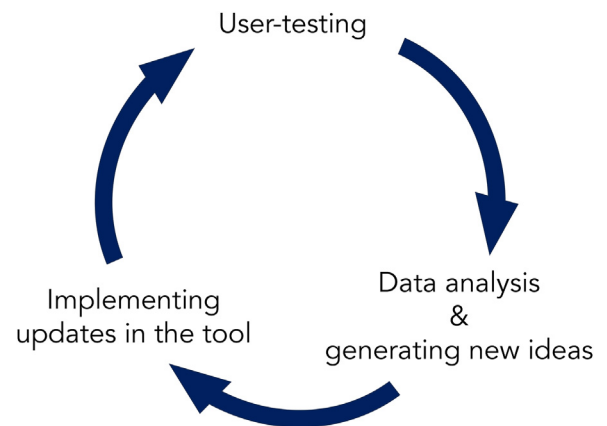
- Researchers are encouraged to include MATCH-IT in network meta-analyses and clinical practice guidelines informed by multiple comparisons.
- Further research is needed to explore the functioning of MATCH-IT at the point of care.

data sharing were signed by all centers. Informed by previous research and innovation projects within digital health [21,32–36], we applied a user-centered design in an iterative development process (Fig 1). We adhered to *the Standards for Reporting Qualitative Research* [37].

### 2.2. The decision support tool

MATCH-IT builds on previously developed decision aids for pairwise comparisons [21], and is an interactive SoF (iSoF) table for multiple comparisons with a section providing information about burden of treatments (“practical issues”) [36]. Results can be viewed according to predefined level of baseline risk. MATCH-IT allows switching between a full overview of the evidence and a more detailed focus on pairwise comparisons with visualization of effects in pictograms. The iSoF table is built by selecting outcomes of interest and can be dynamically modified by changing the comparator, rearranging, or removing interventions and/or outcomes with drag-and-drop functionality. Video 1 showed the initial version of MATCH-IT, while supplement 1 includes a list (with links) of all versions of MATCH-IT used in this study.

In this study, we used MATCH-IT to present data from 3 SRs informing CPGs: medications for type II diabetes [12,38]; cholesterol-lowering medications [13,39]; and management of acute musculoskeletal injuries [14,40].



**Figure 1.** Illustration of the iterative user testing process. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

[Video 1](#) demonstration of the first version of MATCH-IT applied in user testing in this study.

### 2.3. Participants and sampling strategy

Participants in our study were practicing physicians acquired through a convenience snowballing approach. Specifically, with help from colleagues, we recruited physicians, some of whom in turn contributed to recruit other physicians. We aimed for gender balance and variation in age, clinical experience and expertise, and academic credentials.

### 2.4. User testing sessions

Due to the COVID-19 pandemic, most user tests were conducted through video conferences. All user tests were video recorded. A user testing session included introduction, user testing, and a semistructured interview.

We began each user testing session by informing participants about the general concepts of the session, collecting background information, and introducing physicians to a pre-selected relevant clinical scenario and associated guideline ([Supplement 1](#)). Participating physicians were then instructed to explore MATCH-IT freely and uninterrupted while thinking aloud to familiarize themselves with the evidence informing the guideline recommendations. When finished, we conducted a semistructured interview to further explore their user experiences with the tool. All centers followed a standardized interview guide as foundation ([Supplement 2](#)), but with the freedom to adapt/adjust at the interviewer’s discretion.

Before initiation of the study, we piloted the interview guide, performing user tests with 2 physicians. No significant modifications to the guide were needed. We also arranged preparatory meetings for all researchers responsible for conducting interviews. Debriefs between researchers, evaluating the user tests, and interview technique was conducted regularly throughout the process.

## 2.5. Data analysis

All user tests were transcribed manually with a slightly modified verbatim approach. Dutch transcripts and condensed meaning units from Norwegian transcripts were translated to English. We used directed content analysis with all transcripts analyzed by at least 2 researchers [41,42]. We arranged meetings for every iteration to secure alignment in coding between researchers.

We searched the transcripts for meaning units—quotations and operations in the tool—and added codes derived from a modified model of *Moreville's Honeycomb of user-experiences* (Fig S1A) [43] and a model for *the quality of experience* (Fig S1B) [21]. In addition, we added inductively derived codes along 2 lines: “geographical tags”, to pinpoint MATCH-IT strengths and weaknesses, and codes to describe the content more precisely. After completion of the initial coding process for all user tests, we systematically searched for links between the inductive and deductive derived codes, synthesizing layers of subcategories in a code tree. We used Atlas.ti Web (<https://atlasti.com>) for the initial coding process and Microsoft Excel for development of the code tree.

## 2.6. Development

For every cycle in development, we presented and discussed results in interdisciplinary core study team meetings. Sketches and wireframes were actively used to facilitate brainstorming and discussion about potential modifications to MATCH-IT to address limitations

identified through user testing. Final decisions on which updates to prioritize were made by the MAGIC team.

## 3. Results

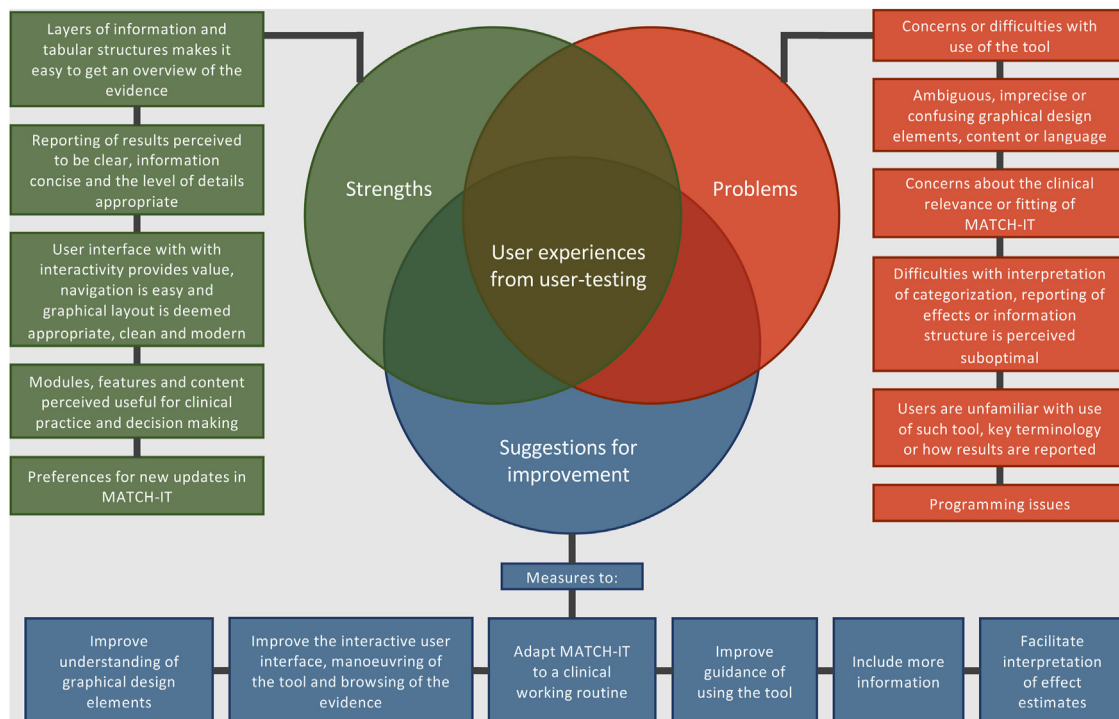
### 3.1. Sample characteristics and iterations of development

The 26 physicians who agreed to participate in our study included 14 men (54%) and 12 women (46%), with a median age of 34 (IQR 28–41) and median clinical experience of 6 years (IQR 3–11). Twelve worked in general practice (46%) and 14 in a hospital setting (54%); physicians were recruited from both medical and surgical departments. Three physicians (12%) had completed a PhD, and 5 (19%) were PhD students. Twenty-four physicians (92%) reported either no or very brief prior experience with interpretation of NMA data. None of the researchers conducting user testing had any formal or informal relationships with the physicians participating in this study.

From February 2021 to August 2022, we completed 5 development cycles and 26 user tests with physicians from Norway ( $n = 16$ , 62%), Belgium ( $n = 5$ , 19%), and Canada ( $n = 5$ , 19%).

### 3.2. Results from user tests

Figure 2 provides an overview of results from user testing of MATCH-IT, with a detailed coding tree available in Supplement 3.



**Figure 2.** Overview of user testing results and considerations for improving MATCH-IT. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

We identified and coded 1399 meaning units; 1205 (86%) were distributed among 3 of 8 coding categories in Moreville—understandability, usability, and usefulness. Thus, we chose to use the overarching themes strengths, problems, and suggestions for improvement, all derived from the quality of the experience model [21], as the root for our coding tree.

### 3.2.1. Strengths

We observed that physicians were able to use and maneuver MATCH-IT in their first interaction. Physicians' feedback was that they could easily use and understand MATCH-IT if they were to use it again (Table 1).

Participants reported that MATCH-IT provided rapid insight into the body of evidence. The tabular structure of displaying results provided a clear overview and made it easy to compare interventions across multiple outcomes (Table 1).

The reporting of effects in absolute measures was well-understood by most. The visualization of effects in pictograms was believed to be useful for interpretation and for communicating with patients (Table 1).

Building the iSoF table step-by-step by selecting outcomes was intuitive for most physicians and was appreciated. It allowed them to focus in on areas of particular interest and the amount of information felt less overwhelming compared to the immediate display of a full SoF table. The ability to adjust the SoF table, remove interventions or change the comparator was also perceived useful (Table 1).

Physicians appreciated the inclusion of the practical issues framework. Specifically, this information provided additional context to treatment effect estimates, covering aspects that they, in their experience, often found to be relevant in a shared decision-making process (Table 1).

### 3.2.2. Problems

While building the SoF table was easy, discovery of the drag-and-drop functionality (eg, changing comparator or removing interventions/outcomes) was not. Still, the physicians were able to easily use these functions once they became aware they were present. The tutorial—a looped video showing the various functions—was perceived to be suboptimal in providing instructions, lacking subtitling or any other form of text support (Table 2).

In the initial version of MATCH-IT, there were some navigation issues between layers. There was no clear route to go back to change baseline risk, nor was there any clear guidance on how to exit/close pop-up windows. This hampered the natural workflow of use (Table 2).

We also received feedback that maneuvering of the tool was “click-heavy” with physicians wanting options for more rapid navigation. Regarding the practical issues framework feedback indicated, in addition to cumbersome

**Table 1. Quotations addressing strengths of MATCH-IT**

a. MATCH-IT was easy to use	b. Reporting and visualization of effects
<p><i>I think that for a first time opening too, I think it was actually very easy.</i> <b>57 years old female, Internist</b></p>	<p><b>Reporting of results in absolute effects</b> <i>It's very clear, I guess, what you're looking at, and what the numbers represent. So, I think, even for someone that had no exposure or prior knowledge to this they would be able to easily recognize what the numbers mean and kind of make a hierarchical approach to what their therapy should be.</i></p>
<p><i>I think it's very easy to use.</i> <b>33 years old female, Resident Orthopedics</b></p>	<p><b>27 years old male, Resident Internal Medicine</b></p>
<p><i>Once you get the hang of it, it works pretty quickly.</i> <b>25 years old female, Resident General Practice</b></p>	<p><b>Visualizing effects in a pictogram</b> <i>I think I'm a very visual person, more than text to be honest. So, I really liked—it's very intuitive, it's very easy to follow, it represents visually well in a graph, you know, the difference.</i></p>
c. Building and adjustment of the SoF table	<p><b>27 years old male, Resident Internal Medicine</b></p>
<p><i>So, I think it's a great way to simplify things without just tossing a whole bunch of information, because, you know, I think the user gets to select what they care about and what's relevant to the case at hand.</i> <b>28 years old male, Resident Internal Medicine</b></p>	<p><b>57 years old female, Internist</b></p>
<p><i>It's all going to show up, but what I find interesting is, it's cool that I get to choose what I am looking for.</i> <b>35 years old male, Intensivist</b></p>	<p><i>Yes, for the pictures here, they can be nice to show to patients.</i></p>
<p><i>But I'm going to put that topical NSAID in the first column one more time for comparison. I do find that very helpful, because then you can really see it visually, fewer and more.</i> <b>25 years old female, Resident General Practice (II)</b></p>	<p><b>50 years old male, General Practitioner</b></p>
e. Overview of the evidence	d. Practical issues—relevant in clinical decision making
<p><i>Amazing, I would have used it straight away, I think I would have remembered better, argued better for treatments, had more confidence, and implemented what the research shows.</i> <b>39 years old male, Resident Internal Medicine</b></p>	<p><i>All those boxes have relevant information, so it's nice to have, it's nice if there's something new you haven't heard of or to prepare a bit before the consultation.</i> <b>44 years old male, General Practitioner</b></p>
<p><i>For yourself I think it's also a good reminder of what exactly does this medication</i></p>	<p><i>So, if there's one thing I really liked, there's the practical issues. As I mentioned, it sometimes—we, as clinicians, we can get bogged down by all these statistics, these clinical trials, these guidelines, but yeah. At the end of the day, if these practical issues are not</i></p>

(Continued)

Table 1. Continued

<p><i>do, what is the amount of evidence and how big is the effect that you can expect? (...) I think it's well-organized and shows you what the added value or non-added value of that medication is.</i></p> <p><b>28 years old female, Resident General Practice</b></p>	<p><i>dealt with—and I guess family physicians know this better than anyone else—is if these are not really dealt with, I think we won't really see the benefit of the therapy that we're initiating. So, um, I think that's a very important portion.</i></p> <p><b>28 years old male, Resident Internal Medicine</b></p>
	<p><i>Those boxes—practical issues—give answers to relevant questions that keep popping up, and it's really user-friendly when the patient is sitting there. If you have forgotten to talk about something.</i></p> <p><b>44 years old male, General Practitioner</b></p>

BMJ, *British Medical Journal*; DPP4 inhibitor, dipeptidyl peptidase 4 inhibitor; EBM, evidence-based medicine; MATCH-IT, Making Alternative Treatment Choices Intuitive and Trustworthy; SoF, Summary of Findings; NSAID, non-steroidal anti-inflammatory drug.

maneuvering, that the information was in risk of being too fragmented when split in 15 categories (Table 2).

Physicians requested full transparency with definitions of key terms, if not this could be a potential showstopper for use in clinical practice. The need for references to sources of information and the opportunity to view more information on request was also highlighted. Further, we registered that familiarity with GRADE concepts varied among participants and, for most, their knowledge was limited (Table 2).

Finally, several physicians expressed a sense of information overload when presented with very large networks, which in this study included comparisons of 27 interventions across 8 outcomes [40]. The most important issue raised was that when the evidence was presented in full across all outcomes and interventions, it was not sufficiently tailored to an actual clinical context, as the recommended interventions were not sufficiently distinguished from the rest [14]. Consequently, participants were concerned that it would take too much time to orientate and interpret such a tool in a busy clinical working routine (Table 2).

### 3.2.3. Updates of the tool

To help physicians discover the drag-and-drop functionality, but also to provide more context to terminology used, we developed a new start page with a stepwise tutorial and

Table 2. Quotations addressing identified limitations of MATCH-IT

<p>a. Drag-and-drop functionality</p> <p><i>It wasn't intuitive to me that I could remove or swap them or anything like that.</i></p> <p><b>40 years old male, General Practitioner</b></p> <p><i>That functionality? I don't think I would have taken that, intuitively. Because everything else has been clicking.</i></p> <p><b>34 years old male, Resident Orthopedics</b></p> <p><i>I think without... without that tutorial you don't know how it works.</i></p> <p><b>25 years old female, Resident General Practice</b></p>	<p>b. Navigation in the tool and structure of practical issues</p> <p><b>Navigation between layers</b></p> <p><i>And maybe also some link to go back, eg, say I forgot that the patient had heart disease. I don't see any button for going back.</i></p> <p><b>40 years old male, General Practitioner</b></p> <p><b>Maneuvering in the tool</b></p> <p><i>However, if I were to use it, it would have taken a noticeable amount of time to navigate.</i></p> <p><b>41 years old male, General Practitioner</b></p>
<p>c. Definition of key terms</p> <p><i>Usual care... What does it mean? Does it involve no treatment, or does it apply to everyone who uses metformin or DPP4 inhibitors?</i></p> <p><b>40 years old male, General Practitioner</b></p> <p><i>Uhm. Severe gastrointestinal events..</i></p> <p><i>Is it pancreatitis or what?</i></p> <p><b>47 years old female, General Practitioner</b></p> <p><i>I was just wondering, certainty... I think that can be a difficult thing for doctors to interpret. Especially if you're not used to reading a lot of studies or dealing with that kind of terminology, what should you conclude from that? (...) It might be helpful if there were a sentence that gave a little more explanation.</i></p> <p><b>31 years old female, Resident General Practice</b></p>	<p><b>Structure of practical issues</b></p> <p><i>It feels like a bit much to click on, when in a way there is only a little information on each. If in a way you could merge these.</i></p> <p><b>47 years old female, General Practitioner</b></p> <p>d. Adaption of large NMAs to a clinical context</p> <p><i>Again, this is so complex, so the way I know my patients, this would be hard to make a decision. Even I would have trouble with this. (...) It is important to have in mind. Should the doctor use this for himself or with a patient, and I believe doing this with patient is useful, then the presentation must be somewhat prepared. Here there are actually few options to consider, but i.e. not reflected in the presentation.</i></p> <p><b>42 years old male, General Practitioner</b></p> <p><i>From a clinical perspective it is a lot of "noise" when presented with this upfront. It is nice to have all interventions available on request, but upfront it is less relevant.</i></p>

(Continued)

**Table 2.** Continued

<p><i>I would prefer to display the interventions that are in accordance with the recommendations.</i></p> <p><b>27 years old male, Intern Medicine</b></p>
---

MATCH-IT, Making Alternative Treatment Choices Intuitive and Trustworthy; NMA, network meta-analysis.

text support (Fig S2). Although the tutorial proved to be both well-understood and facilitated the use of drag-and-drop in later user tests, a limitation was that several chose not to view it, with some explicitly stating they had negative expectations to a tutorial per se.

To improve navigation between layers, we included a dedicated button to change baseline risk, which later evolved to a drop-down menu. We also included exit buttons on pop-up windows and added a “show all/hide all outcomes” button to more efficiently maneuver the SoF table (Fig S3-S5). These functions were well-received in subsequent user testing.

We redesigned the practical issues section to be embedded as part of the SoF table, each cell displaying the gist of information with the ability to click and view a card/table displaying all information across all categories. Most physicians preferred the redesign over the initial version (Fig S6).

To display definitions of key terms we included hover-over functionality. This was appreciated by those who discovered it, but not all did. In addition, we developed a “frequently asked questions” section that included definitions of outcomes, patient population, and interventions, but also more information about NMAs, GRADE and certainty of the evidence, and a link to the original publication. Feedback indicated that both the information structure and the level of details were appropriate (Fig S7).

We also developed a new pictogram card that in addition to absolute effects and pictogram included a plain language summary, relative effects, and numbers of patients and studies providing evidence (Fig S8). This development was mainly driven by user testing in a parallel study focused on MATCH-IT in guideline development [44], but was well-received.

To help facilitate interpretation of large networks we implemented categorization of interventions with color codes, methodologically based on the principles of the GRADE minimally contextualized framework and graphically based on the gradient color model by Phillips et al (Fig S9) [15,45]. In addition, we developed a filter function to display a selection of interventions based on predefined criteria (eg, recommended interventions, interventions

displaying high/moderate certainty evidence, pharmacological interventions, nonpharmacological interventions) (Fig S10). The combination of these 2 functions made it possible to both graphically highlight the most efficacious interventions and/or focus in on interventions recommended in the associated CPG. The updates were well-received and observed to facilitate interpretation and more efficient browsing of the evidence. Our participants indicated a clear preference for a prefiltered selection of interventions upfront with the possibility to expand.

## 4. Discussion

### 4.1. Main findings

Our study provides proof of concept that practicing physicians can use MATCH-IT to understand NMA results in a format perceived as useful. Key positive attributes were the ability to dynamically provide an overview of the evidence (further enhanced by inclusion of color coding and filters), the visualization of effects for pairwise comparisons and the display of information on burden of treatments. However, despite updates to improve MATCH-IT, problems with discovering all relevant functionality remained. In addition, limited knowledge about GRADE and time constraints in clinical practice were identified as potential challenges for use. Nevertheless, considering the alternatives and with only few NMA publications providing structured evidence summaries [23], we believe our findings should encourage use of MATCH-IT (or equivalents) in SRs and CPGs to better support well-informed decisions in policy and practice.

### 4.2. Strengths and limitations of this study

A strength of our study is the iterative tool development process, allowing for repeated validation of updates. We also managed to include a sample of physicians balanced by gender, with variable clinical and academic experience, representing both primary and specialist health care. Finally, we took several measures to achieve methodological rigor, including duplicate analysis and harmonized coding practices of the material for each cycle of development.

Our study also has some limitations. A qualitative design with user testing of hypothetical scenarios within 3 selected topics cannot determine to what extent larger samples of physicians—and other categories of health-care professionals—find the tool useful and understandable in real clinical decision making in different clinical contexts. A physician browsing a CPG may have different time and information needs than when exposed to a clinical question at the point of care or in a shared decision-making context. The representativeness of our sample may also be limited by recruitment of willing and/or positive physicians. While such bias cannot be excluded, our results indicate that

participants were willing to pinpoint shortcomings, difficulties, and problems with MATCH-IT.

#### 4.3. Implications for research and practice

Whereas we are currently planning to perform user testing of MATCH-IT in clinical consultations—focusing on shared decision-making between general practitioners and patients with low health literacy—we welcome research to further test the value of the tool in other target groups and contexts. We have already demonstrated the feasibility of evaluating the tool with quantitative methods, performing a combined randomized control trial and survey with medical students [46], allowing for multidimensional comparisons of specific features. To further explore the performance of alternative reporting approaches and design features, comparing MATCH-IT with similar tools/formats represents a highly relevant next step in our development [4,17,45,47].

The results in this study align with findings from other user testing studies of MATCH-IT involving medical students and guideline panels, respectively. These studies also indicate that the tool provided a good overview of the evidence with good comprehensibility among users who had a clear preference for color coding of results [44,46]. Guideline panels also emphasized the importance of having information on burden of treatments. Furthermore, Phillips et al found that a similar (but static) tabular format displaying structured evidence summaries for multiple comparisons, with categorization of interventions using color coding and clustering of outcomes into benefits and harms, was well-understood by users [45]. However, the limited knowledge of GRADE and certainty of the evidence identified in this study indicate that continued efforts to enhance physicians' comprehension of evidence summaries through educational programs are necessary [10].

During this study, MATCH-IT has been published as part of the World Health Organization living guideline for COVID-19 therapeutics, demonstrating the feasibility of publishing and updating this tool within the novel concept of living evidence [48]. But for such process to be sustainable, automatic import of data is essential for authors. In the case of MATCH-IT, recent progress has been made and automation of data import is now in place, making the size and complexity of networks or frequency of updates no longer an issue for application.

Finally, to allow for a successful implementation of MATCH-IT at the point of care, it is crucial to acknowledge physicians' busy schedule. This necessitates providing easy and instant access to the tool. This may be achieved by integrating MATCH-IT into existing Electronic Health Record Systems or EBM resources like BMJ Best Practice or UpToDate. While such integration is still pending, MAGIC is poised to explore such integration in a related project, 'Enhancing the Evidence Ecosystem.'

## 5. Conclusion

MATCH-IT shows promise in making complex NMA evidence understandable for practicing physicians, although there are limitations, tool-specific and non-tool-specific, that may influence its dissemination and use in clinical practice. Further research is needed to explore MATCH-IT's functioning at the point of care.

### CRediT authorship contribution statement

**Per Olav Løvsletten:** Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Birk Stokke Hunskaar:** Writing – review & editing, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anja Fog Heen:** Writing – review & editing, Supervision, Software, Methodology, Conceptualization. **Geertruida Bekkering:** Writing – review & editing, Resources, Investigation, Formal analysis, Data curation. **Louise Vanden Poel:** Writing – review & editing, Resources, Investigation, Formal analysis, Data curation. **Dena Zeraatkar:** Writing – review & editing, Resources, Investigation, Formal analysis, Data curation. **Mieke Vermandere:** Writing – review & editing, Supervision, Resources, Investigation, Conceptualization. **Bert Aertgeerts:** Writing – review & editing, Supervision, Resources, Investigation, Conceptualization. **Nicolas Delvaux:** Writing – review & editing, Supervision, Resources, Investigation, Conceptualization. **Frankie Achille:** Writing – review & editing, Software. **Jason W. Busse:** Writing – review & editing, Supervision, Resources, Conceptualization. **Thomas Agoritsas:** Writing – review & editing, Supervision, Software, Methodology, Conceptualization. **Per Olav Vandvik:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

### Data availability

Data will be made available on request.

### Declaration of competing interest

T.A. is the Chair of the board and Deputy CEO of the nonprofit organization MAGIC Evidence Ecosystem Foundation (<https://www.magicvidence.org>). F.A. is a paid consultant for the MAGIC Evidence Ecosystem Foundation and is a part of the research team in MAGIC conducting the research project "Making Alternative Treatment Choices



Intuitive and Trustworthy” (MATCH-IT). P.O.V is the CEO of the nonprofit organization MAGIC Evidence Ecosystem Foundation (<https://www.magicvidence.org>). P.O.L, B.S.H, and A.F.H are all part of the research team in MAGIC conducting the research project Making Alternative Treatment Choices Intuitive and Trustworthy (MATCH-IT). J.W.B. is partly supported by the Canadian Institutes of Health Research Canada Research Chair in the prevention and management of chronic pain. There are no competing interests for any other author.

## Acknowledgments

The authors thank Jostein Ibsen for his effort with plotting network meta-analysis datasets into the tool, preparing MATCH-IT (“Making Alternative Treatment Choices Intuitive and Trustworthy”) for user testing.

## Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2024.111399>.

## References

- [1] Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet* 2017;390:415–23.
- [2] Chaimani A, Caldwell DM, Li T, Higgins JP, Salanti G. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023), n.d. Cochrane; 2023.
- [3] Bartoszko JJ, Siemieniuk RA, Kum E, Qasim A, Zeraatkar D, Martinez JPD, et al. Prophylaxis against covid-19: living systematic review and network meta-analysis. *BMJ* 2021;373:n949.
- [4] Siemieniuk RA, Bartoszko JJ, Ge L, Zeraatkar D, Izcovich A, Kum E, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ* 2020;370.
- [5] Shi Q, Nong K, Vandvik PO, Guyatt GH, Schnell O, Rydén L, et al. Benefits and harms of drug treatment for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ* 2023;381:e074068.
- [6] Riaz IB, He H, Ryu AJ, Siddiqi R, Naqvi SAA, Yao Y, et al. A living, interactive systematic review and network meta-analysis of first-line treatment of metastatic renal cell carcinoma. *Eur Urol* 2021;80:712–23.
- [7] Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357–66.
- [8] Cohen AT, Hamilton M, Mitchell SA, Phatak H, Liu X, Bird A, et al. Comparison of the novel oral anticoagulants apixaban, dabigatran, edoxaban, and rivaroxaban in the initial and long-term treatment and prevention of venous thromboembolism: systematic review and network meta-analysis. *PLoS One* 2015;10:e0144856.
- [9] Noone C, Leahy J, Morrissey EC, Newell J, Newell M, Dwyer CP, et al. Comparative efficacy of exercise and anti-hypertensive pharmacological interventions in reducing blood pressure in people with hypertension: a network meta-analysis. *Eur J Prev Cardiol* 2020;27:247–55.
- [10] Tikkinen KAO, Guyatt GH. Understanding of research results, evidence summaries and their applicability—not critical appraisal—are core skills of medical curriculum. *BMJ Evid-Based Med* 2021;26:231–3.
- [11] Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2017;4:CD001431.
- [12] Li S, Vandvik PO, Lytvyn L, Guyatt GH, Palmer SC, Rodriguez-Gutierrez R, et al. SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline. *BMJ* 2021;373:n1091.
- [13] Hao Q, Aertgeerts B, Guyatt G, Bekkering GE, Vandvik PO, Khan SU, et al. PCSK9 inhibitors and ezetimibe for the reduction of cardiovascular events: a clinical practice guideline with risk-stratified recommendations. *BMJ* 2022;377:e069066.
- [14] Qaseem A, McLean RM, O’Gurek D, Batur P, Lin K, Kansagara DL. Nonpharmacologic and pharmacologic management of acute pain from non-low back, musculoskeletal injuries in adults: a clinical guideline from the American College of Physicians and American Academy of Family Physicians. *Ann Intern Med* 2020;173:739–48.
- [15] Brignardello-Petersen R, Florez ID, Izcovich A, Santesso N, Hazlewood G, Alhazanni W, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. *BMJ* 2020;371:m3900.
- [16] Brignardello-Petersen R, Izcovich A, Rochwerf B, Florez ID, Hazlewood G, Alhazanni W, et al. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualised framework. *BMJ* 2020;371:m3907.
- [17] Yepes-Nuñez JJ, Li S-A, Guyatt G, Jack SM, Brozek JL, Beyene J, et al. Development of the summary of findings table for network meta-analysis. *J Clin Epidemiol* 2019;115:1–13.
- [18] Schünemann HJ, Higgins JP, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Chapter 14: Completing ‘Summary of findings’ tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane; 2023.
- [19] Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing Summary of Findings tables—binary outcomes. *J Clin Epidemiol* 2013;66:158–72.
- [20] Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles—continuous outcomes. *J Clin Epidemiol* 2013;66:173–83.
- [21] Heen AF, Vandvik PO, Brandt L, Achille F, Guyatt GH, Akl EA, et al. Decision aids linked to evidence summaries and clinical practice guidelines: results from user-testing in clinical encounters. *BMC Med Inform Decis Mak* 2021;21:202.
- [22] Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *J Clin Epidemiol* 2010;63:620–6.
- [23] Løvstetten PO, Wang X, Pitre T, Ødegaard M, Veroniki AA, Lunny C, et al. [Conference Abstract] Reporting of network meta-analysis results in systematic reviews: a cross-sectional sample of evidence summary presentation formats. In: Abstr. Accept. GIN 2023 Hybrid Conf. Glasg. UK. *Guidelines International Network*; 2023:265–6.
- [24] Helsing LM, Vandvik PO, Jodal HC, Agoritas T, Lytvyn L, Anderson JC, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ* 2019;367:l5515.
- [25] Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han PK, et al. Presenting quantitative information about

- decision outcomes: a risk communication primer for patient decision aid developers. *BMC Med Inform Decis Mak* 2013;13:S7.
- [26] Song HS, Pusic M, Nick MW, Sarpel U, Plass JL, Kalet AL. The cognitive impact of interactive design features for learning complex materials in medical education. *Comput Educ* 2014;71:198–205.
- [27] Lustria MLA. Can interactivity make a difference? Effects of interactivity on the comprehension of and attitudes toward online health content. *J Am Soc Inf Sci Technol* 2007;58:766–76.
- [28] Domagk S, Schwartz RN, Plass JL. Interactivity in multimedia learning: an integrated model. *Comput Hum Behav* 2010;26:1024–33.
- [29] Bernard AL, Barbour AK, Meernik C, Madeira JL, Lindheim SR, Goodman LR. The impact of an interactive multimedia educational platform on patient comprehension and anxiety during fertility treatment: a randomized controlled trial. *FS Rep* 2022;3:214–22.
- [30] Kristiansen A, Brandt L, Alonso-Coello P, Agoritsas T, Akl EA, Conboy T, et al. Development of a novel, multilayered presentation format for clinical practice guidelines. *Chest* 2015;147:754–63.
- [31] Brandt L, Vandvik PO, Alonso-Coello P, Akl EA, Thornton J, Rigau D, et al. Multilayered and digitally structured presentation formats of trustworthy recommendations: a combined survey and randomised trial. *BMJ Open* 2017;7:e011569.
- [32] Seljelid B, Varsi C, Solberg Nes L, Stenehjem A-E, Bollerslev J, Børøsdund E. Content and system development of a digital patient-provider communication tool to support shared decision making in chronic health care: InvolveMe. *BMC Med Inform Decis Mak* 2020;20:46.
- [33] Breslin M, Mullan RJ, Montori VM. The design of a decision aid about diabetes medications for use during the consultation with patients with type 2 diabetes. *Patient Educ Couns* 2008;73:465–72.
- [34] Rosenbaum SE, Glenton C, Cracknell J. User experiences of evidence-based online resources for health professionals: user testing of the Cochrane Library. *BMC Med Inform Decis Mak* 2008;8:34.
- [35] Treweek S, Oxman AD, Alderson P, Bossuyt PM, Brandt L, Brožek J, et al. Developing and evaluating communication strategies to support informed decisions and practice based on evidence (DECIDE): protocol and preliminary results. *Implement Sci* 2013;8:6.
- [36] Heen AF, Vandvik PO, Brandt L, Montori VM, Lytvyn L, Guyatt G, et al. A framework for practical issues was developed to inform shared decision-making tools and clinical guidelines. *J Clin Epidemiol* 2021;129:104–13.
- [37] O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014;89:1245.
- [38] Palmer SC, Tendal B, Mustafa RA, Vandvik PO, Li S, Hao Q, et al. Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ* 2021;372:m4573.
- [39] Khan SU, Yedlapati SH, Lone AN, Hao Q, Guyatt G, Delvaux N, et al. PCSK9 inhibitors and ezetimibe with or without statin therapy for cardiovascular risk reduction: a systematic review and network meta-analysis. *BMJ* 2022;377:e069116.
- [40] Busse JW, Sadeghirad B, Oparin Y, Chen E, Goshua A, May C, et al. Management of acute pain from non-low back, musculoskeletal injuries: a systematic review and network meta-analysis of randomized trials. *Ann Intern Med* 2020;173:730–8.
- [41] Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005;15:1277–88.
- [42] Assaroudi A, Heshmati Nabavi F, Armat MR, Ebadi A, Vaismoradi M. Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. *J Res Nurs* 2018;23:42–55.
- [43] Rosenbaum S. Improving the user experience of evidence. A design approach to evidence-informed health care. Oslo, Norway: Oslo College of Architecture and Design; 2010: Phd thesis.
- [44] Stokke HB, Løvsletten PO, Heen AF, Achille F, Agoritsas T, Vandvik PO. [Conference Abstract] User-testing of a decision support tool for multiple treatment options with NMA data to facilitate guideline development: a qualitative study. In: *Abstr. Accept. GIN 2023 Hybrid Conf. Glasg. UK. Guidelines International Network*; 2023:329.
- [45] Phillips MR, Sadeghirad B, Busse JW, Brignardello-Petersen R, Cuello-Garcia CA, Kenji Nampo F, et al. Development and design validation of a novel network meta-analysis presentation tool for multiple outcomes: a qualitative descriptive study. *BMJ Open* 2022;12:e056400.
- [46] Stokke Hunskaar B, Løvsletten PO, Muller A, Vandvik PO. Interpretation and use of a decision support tool for multiple treatment options: a combined randomised controlled trial and survey of medical students. *BMJ Evid-Based Med* 2024;29:29–36.
- [47] Seo M, Furukawa TA, Veroniki AA, Pillinger T, Tomlinson A, Salanti G, et al. The Kilim plot: a tool for visualizing network meta-analysis results for multiple outcomes. *Res Synth Methods* 2021;12:86–95.
- [48] Agarwal A, Hunt BJ, Stegemann M, Rochwerg B, Lamontagne F, Siemieniuk RA, et al. A living WHO guideline on drugs for covid-19. *BMJ* 2020;370:m3379.