

Review

# Reporting Quality of AI Intervention in Randomized Controlled Trials in Primary Care: Systematic Review and Meta-Epidemiological Study

Jinjia Zhong, BMed; Ting Zhu, BMed; Yafang Huang, PhD

School of General Practice and Continuing Education, Capital Medical University, Beijing, China

**Corresponding Author:**

Yafang Huang, PhD

School of General Practice and Continuing Education

Capital Medical University

4th Fl, Jieping Building, Capital Medical University

No.10 You An Men Wai Xi Tou Tiao, Fengtai district

Beijing, 100069

China

Phone: 86 18810673886

Fax: 86 1083911501

Email: [yafang@ccmu.edu.cn](mailto:yafang@ccmu.edu.cn)

## Abstract

**Background:** The surge in artificial intelligence (AI) interventions in primary care trials lacks a study on reporting quality.

**Objective:** This study aimed to systematically evaluate the reporting quality of both published randomized controlled trials (RCTs) and protocols for RCTs that investigated AI interventions in primary care.

**Methods:** PubMed, Embase, Cochrane Library, MEDLINE, Web of Science, and CINAHL databases were searched for RCTs and protocols on AI interventions in primary care until November 2024. Eligible studies were published RCTs or full protocols for RCTs exploring AI interventions in primary care. The reporting quality was assessed using CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) and SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) checklists, focusing on AI intervention–related items.

**Results:** A total of 11,711 records were identified. In total, 19 published RCTs and 21 RCT protocols for 35 trials were included. The overall proportion of adequately reported items was 65% (172/266; 95% CI 59%-70%) and 68% (214/315; 95% CI 62%-73%) for RCTs and protocols, respectively. The percentage of RCTs and protocols that reported a specific item ranged from 11% (2/19) to 100% (19/19) and from 10% (2/21) to 100% (21/21), respectively. The reporting of both RCTs and protocols exhibited similar characteristics and trends. They both lack transparency and completeness, which can be summarized in three aspects: without providing adequate information regarding the input data, without mentioning the methods for identifying and analyzing performance errors, and without stating whether and how the AI intervention and its code can be accessed.

**Conclusions:** The reporting quality could be improved in both RCTs and protocols. This study helps promote the transparent and complete reporting of trials with AI interventions in primary care.

(*J Med Internet Res* 2025;27:e56774) doi: [10.2196/56774](https://doi.org/10.2196/56774)

## KEYWORDS

artificial intelligence; randomized controlled trial; reporting quality; primary care; meta-epidemiological study

## Introduction

Primary care provides a large health care delivery platform for primary care physicians to offer person-centered services for a vast patient population [1]. The practice of primary care generates substantial digital data that can be leveraged for primary care research [2,3]. The growing resources of big data,

mainly consisting of electronic health records, electronic medical data, patient self-reported data, and wearable device data, have paved the way for the advancement of artificial intelligence (AI) in deep machine learning (ML) technology and natural language processing to explore AI- or ML-driven clinical tools dedicated to the improvement of clinical decision-making in the disease screening, diagnosis, prognosis, and management

[4-6]. For example, several studies in primary care have developed the prediction tool for patients presenting with acute cough, the prediction score for head and neck cancer referrals, and the prediction model for asthma exacerbation [7-9]. Nevertheless, in practice, a significant proportion of these tools have not undergone validation through a robust clinical trial, although the majority of these studies use performance indicators to showcase their superiority [10,11].

Randomized controlled trials (RCTs) are widely recognized as the gold standard for evaluating interventions [12]. Many clinical studies that investigate AI- or ML-driven clinical tools as a clinical intervention conduct a confirmatory RCT, aiming to demonstrate the clinical significance [13-15]. Transparency and explainability are essential for the widespread integration of AI systems into clinical practice, as an inaccurate prediction could result in severe consequences [16]. Given the inherent complexity of AI interventions, it is critical for RCT reporting to be comprehensive and transparent. Adhering to reporting guidelines for AI in clinical trials and transparently reporting AI interventions in RCTs are important steps toward improving research quality, fostering scientific discourse, and establishing more reliable foundations for clinical practice and decision-making [17-19].

The CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) extension [17] and SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) extension [18] are newly developed reporting guidelines specifically focusing on AI intervention in trials. CONSORT-AI focuses more on trial results and reporting completeness, while SPIRIT-AI emphasizes pretrial elements, such as trial design and ethical considerations. The relative items in the guidelines should be transparently and completely reported, and qualified reporting is essential for independently evaluating and replicating the trial. However, there is no systematic review and critical appraisal of RCTs for examining AI interventions in primary care, leaving the clarity of their reporting quality uncertain.

This systematic review and meta-epidemiological study aims to evaluate the reporting quality of AI interventions in published RCTs and protocols for RCTs in primary care. The assessment was based on the CONSORT-AI extension and SPIRIT-AI extension guidelines.

## Methods

### Study Design

This study was reported in accordance with the guidelines for reporting meta-epidemiological methodology research [20] and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) reporting guidelines (Multimedia Appendix 1). The protocol for this study has been registered on PROSPERO (ID: CRD42023427694).

### Search Strategy

The searches of PubMed, Embase, and Cochrane Library were conducted up to November 30, 2024. To ensure a comprehensive literature search, we further referenced a previously published study strategy [3], ensuring the completeness and accuracy of

the process of collecting RCTs. This further search was applied across MEDLINE, Web of Science, and CINAHL. There were no restrictions on the year of publication or language. The detailed search strategies are listed in Multimedia Appendix 2.

### Study Selection

#### Overview

To be included in this study, articles had to meet the following criteria: (1) RCTs or protocols for RCTs that used AI interventions or AI-assisted tools to guide a randomized intervention and (2) studies that belonged to primary care research. Based on the 1996 report of the US Institute of Medicine, primary care was defined as “the provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients, and practicing in the context of family and community” [21]. Primary care research was defined as “Research done in a primary care context” [22]. Only published RCTs and published, full RCT protocols were included. We excluded (1) studies with abstracts only; (2) commentaries, letters, editorials, or reviews; and (3) animals or preclinical studies.

#### Stage 1: Title and Abstract Screening

The retrieved records were imported into EndNote 20 (Clarivate). Two independent reviewers (JZ and TZ) conducted the screening of titles and abstracts to identify studies potentially meeting the inclusion criteria. Any discrepancies were resolved through discussion.

#### Stage 2: Full-Text Screening

Using EndNote 20, the reviewers (JZ and TZ) independently assessed the full-text articles identified at stage 1 for eligibility. Any discrepancies were discussed and resolved by consulting the third researcher (YH), and consensus was reached through discussion. Studies meeting the eligibility criteria were included for data extraction.

#### Data Extraction Process and Data Items

Two researchers (JZ and TZ) independently extracted data from each article by using a standardized data extraction form. Discrepancies in data extraction were discussed between JZ and TZ and the remaining conflicts were resolved by YH.

Data were collected on (1) trial name, (2) trial registration identifier assigned to clinical trials registered in the database, (3) first author’s name, (4) published article name, (5) journal name, (6) publication year, (7) research topic, (8) study design (sample size, blinding, and type of comparator arm), (9) primary outcome, (10) classification of primary outcomes by result status (positive or negative), (11) type of AI model (large language model [LLM], ML, deep learning, clinical decision support system [CDSS], or risk prediction model), and (12) deployment context (clinician-assisted decision support, fully automated diagnostic systems, or patient self-management tools).

The reporting quality of included studies was assessed using the CONSORT-AI and SPIRIT-AI checklists. Each checklist consists of two components: (1) the original CONSORT 2010/SPIRIT 2013 items and (2) the AI-specific elaborations

and extension items. This study focused on evaluating the AI-related reporting quality. Only items that are separately elaborated or extended in the CONSORT-AI (14 items) and SPIRIT-AI (15 items) checklists were appraised. These items ensure that critical aspects of AI-supported interventions such as input-data setting are transparently and completely reported. Each item was assessed in all included published articles, as “Yes” or “No.”

### Outcomes

The primary outcomes were as follows: (1) the proportion of adequately reported items, calculated by dividing the number of adequately reported items by the total number of items in each article—high proportions would indicate high reproducibility and quality of the RCTs and protocols for RCTs—and (2) the percentage of articles adequately reporting each item, calculated by dividing the number of articles that adequately reported the item by the total number of articles evaluated for that particular item. Independent analysis was conducted for CONSORT-AI (14 items) and SPIRIT-AI (15 items) checklists based on the primary outcomes described above. Since CONSORT-AI focuses on reporting completed trials and SPIRIT-AI on trial protocol standards, analyzing them separately ensures more precise adherence to each guideline. The secondary outcomes focused on CONSORT-AI included (1) the association between primary outcome status and the reporting quality of each CONSORT-AI item, and (2) factors influencing the percentage of adequately reported items in CONSORT-AI.

### Data Synthesis and Analysis

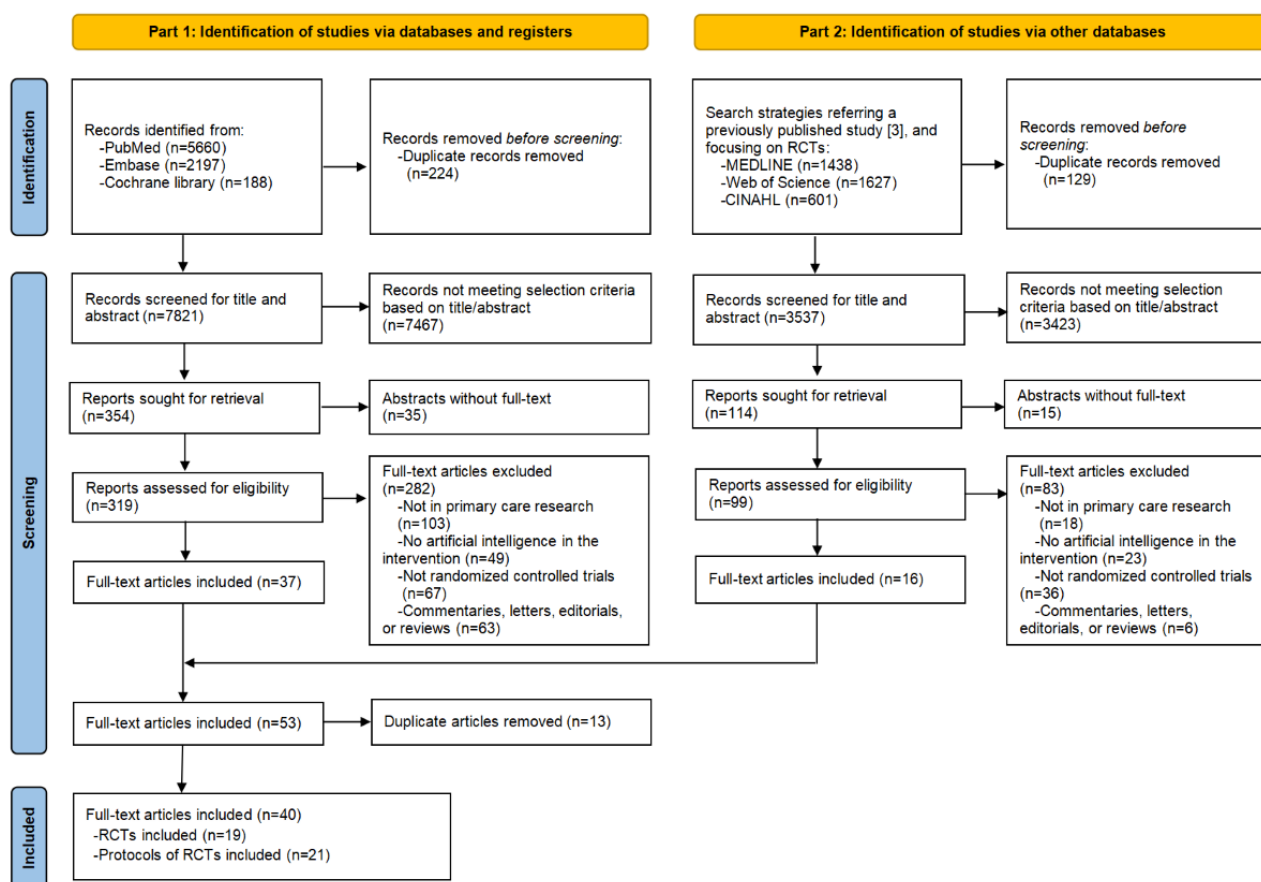
Descriptive statistical methods were used and presented as frequencies, median percentages, and IQRs. The proportion of adequately reported items and the percentage of articles adequately reporting each specific item was reported, with its corresponding 95% CI calculated using the Clopper-Person method. The analyses were performed using Microsoft Excel 2021 and OriginPro 2022 (OriginLab Corporation). The association between primary outcome result status (positive or negative) and the reporting quality of each item was examined using the Fisher exact test. The percentage of adequately reported items was compared across RCT characteristics such as type of disease using the Mann-Whitney *U* test. To assess the main effects and interaction effects between the primary outcome result status and various RCT characteristics on the percentage of adequately reported items, the Scheirer-Ray-Hare test was used. All *P* values were derived from 2-sided tests, with significance defined as  $P < .05$  (R, version 4.4.2; R Foundation for Statistical Computing).

## Results

### Overview

The search strategies retrieved a total of 11,711 records. After screening, 40 articles finally met the inclusion criteria. These 40 full-text articles corresponded to 35 trials, including 19 RCTs and 21 protocols for RCTs (Figure 1 and Multimedia Appendix 3 [23-62]).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram of the selection procedure. RCT: randomized controlled trial.



### Characteristics of Included Articles

Table 1 summarizes the characteristics of the included articles. Among the 19 RCTs, 8 (42%) used LLMs, 10 (53%) were conducted in the context of clinician-assisted decision support, and 11 (58%) reported positive primary outcome results. In

total, 21% (4/19) focused on cardiovascular topics. The median sample size for these RCTs was 335 (IQR 133–487) participants. Among the 21 protocols, 6 (29%) incorporated LLMs, 11 (52%) involved clinician-assisted decision support, and 6 (29%) focused on cardiovascular topics (Table 1).

**Table 1.** Characteristics of included published randomized controlled trials (RCTs) and protocols for RCTs.

Characteristics	Randomized controlled trials (n=19)	Protocols (n=21)	Overall (n=40)
Reported items, median (range)	9 (7-14)	10 (7-13)	9 (7-14)
<b>Publication year, n (%)</b>			
2024	7 (37)	3 (14)	10 (25)
2023	3 (16)	2 (10)	5 (13)
2022	6 (32)	9 (43)	15 (38)
Pre-2021	3 (16)	7 (33)	10 (25)
Sample size, median (IQR)	335 (133-487)	— <sup>a</sup>	—
<b>Blinding, n (%)</b>			
Single blinded	6 (32)	2 (10)	8 (20)
Double-blinded	1 (5)	3 (14)	4 (10)
Open label	12 (63)	16 (76)	28 (70)
<b>Type of comparator arm, n (%)</b>			
Usual care	18 (95)	17 (81)	35 (88)
No treatment	1 (5)	3 (14)	4 (10)
Delay intervention	0 (0)	1 (5)	1 (3)
<b>Type of artificial intelligence model, n (%)</b>			
Large language model	8 (42)	6 (29)	14 (35)
Machine learning	2 (11)	4 (19)	6 (15)
Deep learning	6 (32)	6 (29)	12 (30)
Clinical decision support system	2 (11)	3 (14)	5 (13)
Risk prediction model	1 (5)	2 (10)	3 (8)
<b>Deployment context, n (%)</b>			
Clinician-assisted decision support	10 (53)	11 (52)	21 (53)
Fully automated diagnostic systems	3 (16)	1 (5)	4 (10)
Patient self-management tools	6 (32)	9 (43)	15 (38)
<b>Classification of primary outcomes by result status, n (%)</b>			
Positive	11 (58)	—	—
Negative	8 (42)	—	—
<b>Disease or topic, n (%)</b>			
Cardiovascular disease	4 (21)	6 (29)	10 (25)
Oncology	2 (11)	3 (14)	5 (13)
Pain management	4 (21)	1 (5)	5 (13)
Endocrine disorders	1 (5)	2 (10)	3 (8)
Respiratory diseases	2 (11)	1 (5)	3 (8)
Digestive disorders	2 (11)	0 (0)	2 (5)
Mental illness	1 (5)	2 (10)	3 (8)
Others	3 (16) <sup>b</sup>	6 (29) <sup>c</sup>	9 (23)

<sup>a</sup>Not applicable.<sup>b</sup>Prescription management, diagnostic reasoning, and quitting smoking.<sup>c</sup>Auxiliary diagnosis, prescription management, retinal disease, quit smoking, and drug overdose.

## Proportion of Adequately Reported Items

In the 19 RCTs included in this review, the overall proportion of adequately reported items was 65% (172/266; 95% CI 59%-70%). Only 2 RCTs had more than 90% of items reported adequately. In the 21 protocols for RCTs, the overall proportion of adequately reported items was 68% (214/315; 95% CI 62%-73%). Two protocols of RCTs had more than 85% of items reported adequately (Multimedia Appendix 3).

## Percentage of Articles Adequately Reporting Each Specific Item

The complete list of the 14 items of the CONSORT-AI extension is shown in Table 2. The percentage of RCTs that reported a

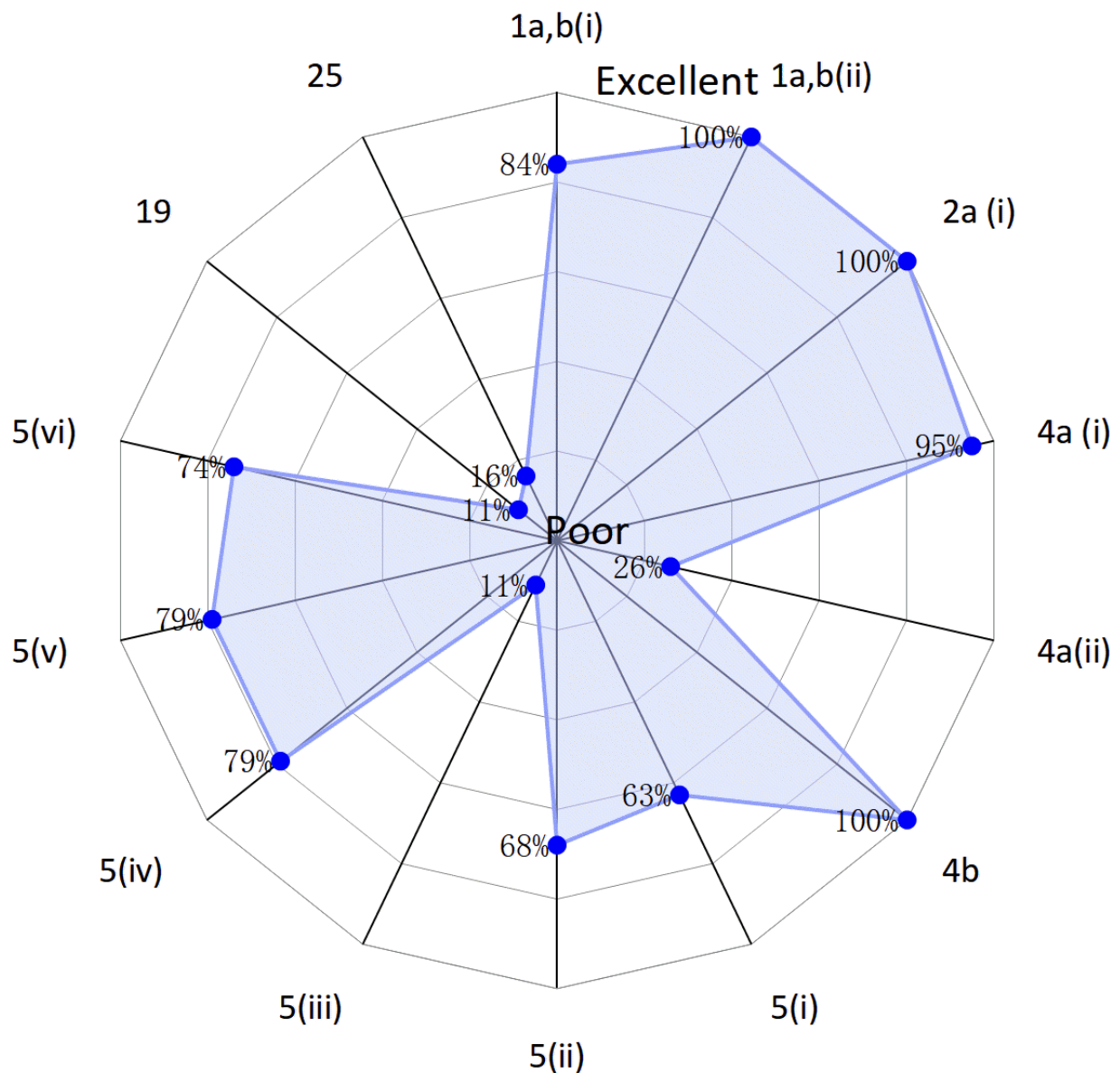
specific item ranged from 11% to 100%. The best-reported sections were the title and abstract [item 1a and item b(ii)], background and objective [item 2a(i)], and participants (item 4b), all being reported in 100% (19/19) of RCTs. The poorly reported sections were intervention [item 5(i), item 5(ii), and item 5(iii)], participants [item 4a(ii)], harms (item 19), and funding (item 25). Among these poorly reported items, 4a(ii), 5(ii), and 5(iii) were related to providing the information for input data (Figure 2 and Table 2). The reporting quality increases as it approaches the outer edges of the radar chart—1a, b(i), 1a,b(ii): title; 2a(i): background and objectives; 4a(i), 4a(ii), 4b: participants; 5(i)-5(iv): intervention; 19: harms; and 25: funding.

**Table 2.** Adherence to CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) extension items.

Item	Item description	Total (n=19), n (%; 95%CI)
CONSORT-AI 1a Elaboration	Indicate that the intervention involves artificial intelligence or machine learning in the title and abstract and specify the type of model.	16 (84; 60-97)
CONSORT-AI 1b(ii) Elaboration	State the intended use of the AI <sup>a</sup> intervention within the trial in the title and abstract.	19 (100; 82-100)
CONSORT-AI 2a(i) Extension	Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (eg, health care professionals, patients, public).	19 (100; 82-100)
CONSORT-AI 4a(i) Elaboration	State the inclusion and exclusion criteria at the level of participants	18 (95; 74-100)
CONSORT-AI 4a(ii) Extension	State the inclusion and exclusion criteria at the level of the input data.	5 (26; 9-51)
CONSORT-AI 4b Extension	Describe how the AI intervention was integrated into the trial setting, including any onsite or offsite requirements.	19 (100; 82-100)
CONSORT-AI 5(i) Extension	State which version of the AI algorithm was used.	12 (63; 38-84)
CONSORT-AI 5(ii) Extension	Describe how the input data were acquired and selected for the AI intervention.	13 (68; 43-87)
CONSORT-AI 5(iii) Extension	Describe how poor quality or unavailable input data were assessed and handled.	2 (11; 1-33)
CONSORT-AI 5(iv) Extension	Specify whether there was human-AI interaction in the handling of the input data, and what level of expertise was required of users.	15 (79; 54-94)
CONSORT-AI 5(v) Extension	Specify the output of the AI intervention.	15 (79; 54-94)
CONSORT-AI 5(vi) Extension	Explain how the AI intervention's outputs contributed to decision-making or other elements of clinical practice.	14 (74; 49-91)
CONSORT-AI 19 Extension	Describe results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, justify why not.	2 (11; 1-33)
CONSORT-AI 25 Extension	State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.	3 (16; 3-40)

<sup>a</sup>AI: artificial intelligence.

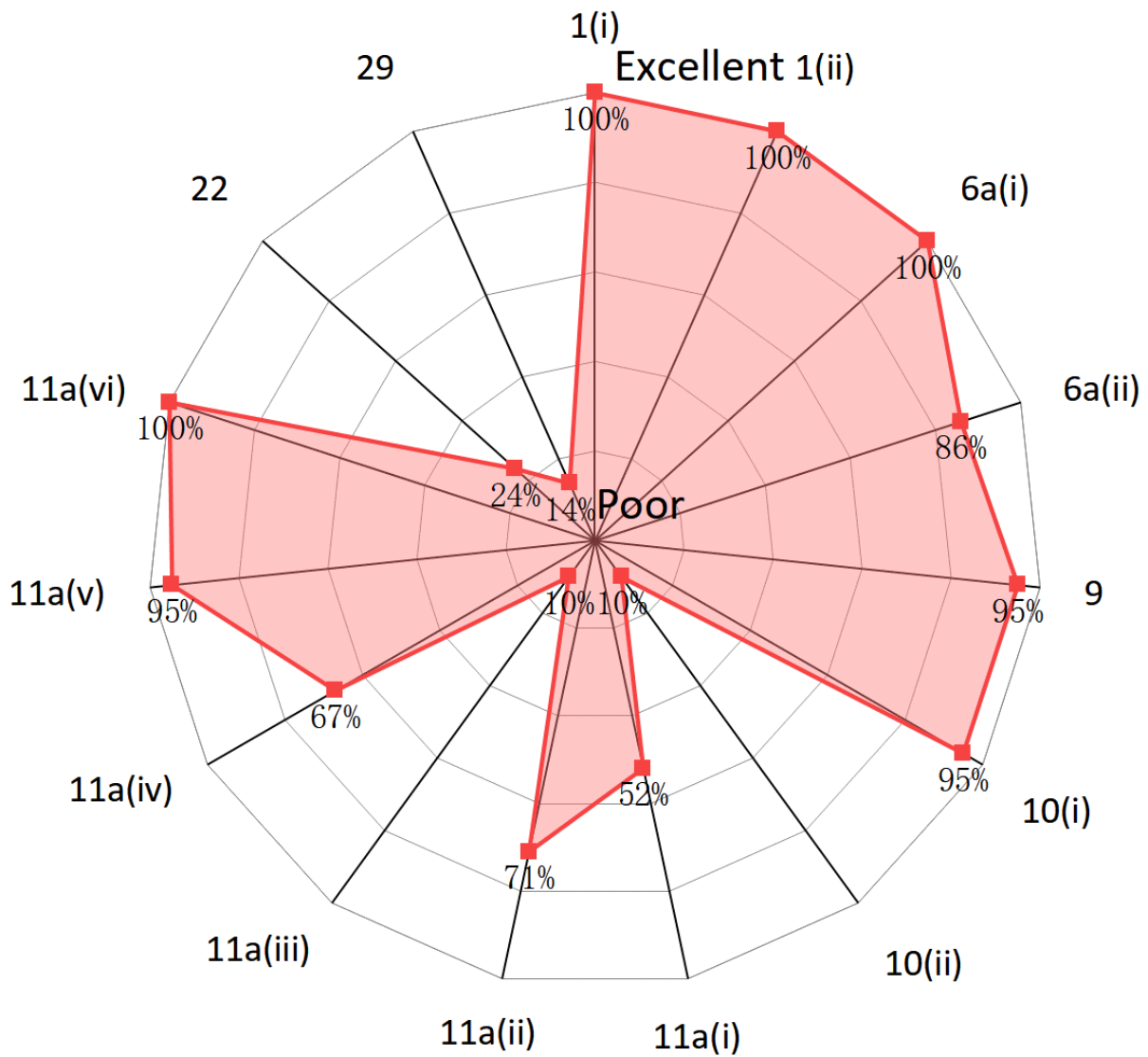
**Figure 2.** Radar chart to show the adequately reported percentage for each CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) item.



The complete list of the 15 items of the SPIRIT-AI extension is shown in Table 3. The percentage of protocols for RCTs that reported a specific item ranged from 10% to 100%. The best-reported sections were title and abstract [item 1(i) and item 1(ii)], background and rationale [item 6a(i)], and intervention [item 11a(vi)] all being reported in 100% of protocols. The poorly reported sections were intervention [item 11a(i), item 11a(iii), and item 11a(iv)], eligibility [item 10(ii)], harms (item

22), and access to data (item 29). Among these poorly reported items, 10(ii), 11a(iii), and 11a(iv) were related to providing the information for input data (Figure 3 and Table 3). The reporting quality increases as it approaches the outer edges of the radar chart—1(i), 1(ii): title; 6a(i), 6a(ii): background and rationale; 9: study setting; 10(i), 10(ii): eligibility criteria; 11a(i)-11a(vi): interventions; 22: harms; and 29: access to data.

**Figure 3.** Radar chart to show the adequately reported percentage for each SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) item.





**Table 3.** Adherence to SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials—Artificial Intelligence) extension items.

Item	Item description	Total (n=21), n (%; 95% CI)
SPIRIT-AI 1(i) Elaboration	Indicate that the intervention involves artificial intelligence or machine learning and specify the type of model.	21 (100; 84-100)
SPIRIT-AI 1(ii) Elaboration	Specify the intended use of the AI <sup>a</sup> intervention.	21 (100; 84-100)
SPIRIT-AI 6a(i) Extension	Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (eg, health care professionals, patients, public).	21 (100; 84-100)
SPIRIT-AI 6a(ii) Extension	Describe any pre-existing evidence for the AI intervention.	18 (86; 64-97)
SPIRIT-AI 9 Extension	Describe the onsite and offsite requirements needed to integrate the AI intervention into the trial setting.	20 (95; 76-100)
SPIRIT-AI 10(i) Elaboration	State the inclusion and exclusion criteria at the level of participants.	20 (95; 76-100)
SPIRIT-AI 10(ii) Extension	State the inclusion and exclusion criteria at the level of the input data.	2 (10; 1-30)
SPIRIT-AI 11a(i) Extension	State which version of the AI algorithm will be used.	11 (52; 30-74)
SPIRIT-AI 11a(ii) Extension	Specify the procedure for acquiring and selecting the input data for the AI intervention.	15 (71; 48-89)
SPIRIT-AI 11a(iii) Extension	Specify the procedure for assessing and handling poor quality or unavailable input data.	2 (10; 1-30)
SPIRIT-AI 11a(iv) Extension	Specify whether there is human-AI interaction in the handling of the input data, and what level of expertise is required for users.	14 (67; 43-85)
SPIRIT-AI 11a(v) Extension	Specify the output of the AI intervention.	20 (95; 76-100)
SPIRIT-AI 11a(vi) Extension	Explain the procedure for how the AI intervention's output will contribute to decision-making or other elements of clinical practice.	21 (100; 84-100)
SPIRIT-AI 22 Extension	Specify any plans to identify and analyze performance errors. If there are no plans for this, explain why not.	5 (24; 8-47)
SPIRIT-AI 29 Extension	State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.	3 (14; 3-36)

<sup>a</sup>AI: artificial intelligence.

### Association Between Primary Outcome Result Status and the Reporting Quality of Each Item

Multimedia Appendix 4 shows the association between the primary outcome result status in RCTs and the reporting quality of each item. Fisher exact test results indicate no significant association between positive outcomes and higher reporting quality (all  $P > .05$ ).

### Factors Associated With the Percentage of Adequately Reported Items

To further explore potential differences resulting from RCT characteristics, we compared the percentage of adequately reported items by year of publication, blinding, type of AI model, deployment context, primary outcome result status, and disease type (Multimedia Appendix 5). Trials of cardiovascular disease showed a higher percentage of adequately reported items ( $P = .02$ ). The Scheiner-Ray-Hare test did not detect statistical significance for the interaction effects between primary outcome result status and various RCT characteristics on the percentage of adequately reported items (all  $P > .05$ ; Multimedia Appendix 5).

## Discussion

### Principal Results

In this study, we assessed the RCT reporting quality in published 19 RCTs and 21 RCT protocols involved in 35 trials, with 5 trials having both published RCTs and corresponding protocols. This study found the reporting quality in both RCTs and protocols for RCTs of AI intervention could be improved in 3 aspects. The first aspect concerns providing enough information for the input data, which was detailed in CONSORT-AI guidelines [items 4a(ii), 5(ii), and 5(iii)] and SPIRIT-AI guidelines [items 10(ii), 11a(iii), and 11a(iv)]. The second aspect involves the lack of information on how to identify and analyze performance errors, which was demanded by the harm section of CONSORT-AI guidelines (item 19) and SPIRIT-AI guidelines (item 22), respectively. The third aspect involves neglecting to specify whether and how the AI intervention or its code can be accessed, which was needed in CONSORT-AI guidelines (item 25) and SPIRIT-AI guidelines (item 29), respectively. Based on the evaluation of the reporting of 35 trials, our findings revealed a consistent lack of information across these three aspects across both protocols and RCTs, highlighting a widespread gap in adhering to the reporting standards recommended by these guidelines.

## Strength and Limitations

This is the first study focused on the field of AI application in primary care, providing a first systematic review into the existing landscape of RCTs using AI interventions within primary care. This review not only included RCTs of AI interventions in primary care but also covered protocols, providing a more comprehensive survey in this field.

This study has 2 limitations. First, the majority of studies were conducted in the United States, which may lead to a lack of representation for other countries. Second, this study assessed the reporting quality of AI intervention-related items for RCTs. General reporting items from the original CONSORT 2010 and SPIRIT 2013 checklists were not assessed.

## Comparison With Previous Work

We are aware of only one published systematic scoping review on the application of AI in community-based primary health care [3]. The authors highlighted a gap in the development and implementation of AI in primary care. However, reporting quality was not assessed previous study [3]. This systematic review is the first focusing on primary care and indicating the need for more transparent design, conduct, and analysis of AI interventions in primary care.

## Implications for Research and Practice

To apply AI interventions in primary care, accurate and sufficient information is essential to guide the standardized performance of clinical treatment. The findings of this study have several implications for clinical research and practice, concerning AI safety, reliability, and reproducibility, as outlined below.

First, reporting on AI performance errors needs to be improved. As software, AI systems are likely to undergo multiple iterations and updates [10]. For example, although LLMs have successfully answered medical licensing exam questions [63], they still produce errors when stating facts or synthesizing data from medical literature [64]. Since LLMs still have many flaws, performance error reporting is essential for continuous updating. The lack of performance error reporting may lead to misdiagnosis or incorrect treatment recommendations, ultimately jeopardizing patient safety [65]. It is crucial to specify the types of performance errors, the process of identifying them, and how updated versions correct these errors during the trial.

Second, the reporting on the section of input-data setting deserves special attention. AI interventions applied to primary care should provide comprehensive information for input-data settings, including parameter selection and preprocessing before analysis by the AI system. Moreover, during the input-data setting, the inclusion and exclusion criteria should be clearly defined, along with an adequate investigation of patient characteristics relevant to the disease type. It is necessary to report these items completely to ensure the replicability of the intervention beyond the trials in real-world circumstances [19]. It also supports investigators in identifying whether input-data-handling procedures were standardized across trial sites [17]. Overall, the AI system should establish an optimal input set that includes a wide range of parameters and attribute

combinations [66]. For example, ML prediction models can only accurately predict the specific outcomes after knee arthroplasty while the ability to predict more complex outcomes remains inaccuracy [67]. Thus, ML requires comprehensive patient-related indicators and the identification of specific patterns that are suitable for ML analysis. This underscores the importance of high-quality input data setting as a fundamental prerequisite for optimizing ML capabilities in clinical applications. Finally, trials should report the prespecified conditions, particularly if minimum requirements for input data are not met. Failing to transparently report how input data were acquired and selected can compromise the representativeness and generalizability of the AI model, potentially leading to critical errors in AI-driven decision-making, which may adversely affect patient safety and clinical outcomes.[17]

In particular, reporting on assessment and handling of input data needs to be improved. The reporting of AI intervention should mention the amount of poor-quality input data, as well as how this was identified and handled [17]. Poor-quality or unavailable input data can negatively impact the success of AI algorithms and the AI's performance and effectiveness. It is essential to have a plan in place for handling such scenarios, such as implementing data cleaning techniques, using data imputation methods, or seeking alternative data sources.[10] A study on cardiovascular imaging pointed out that the quality and preprocessing of input data such as echocardiography, cardiac magnetic resonance imaging, and cardiac computed tomography should be evaluated before a model can be trained using deep learning techniques [68]. Otherwise, the black-box nature of AI models may lead to a lack of trust among clinicians and sponsors [69,70].

Finally, reporting regarding access to and reuse of AI intervention and its code needs to be specified. Human-AI interaction promotes the capabilities of AI systems, as it leverages human expertise to ensure the practical application of AI models. A review on ML highlights three reasons for the necessity of human involvement: real-world problems are inherently complex, ML methods often lack explainability, and AI outputs may not always align with clinical expectations or disease judgment [71]. It is imperative that health care professionals can validate and refine AI methodologies to ensure they are clinically applicable [72,73]. Therefore, specifying whether and how the AI intervention and its underlying code can be accessed by designated health care professionals is crucial. Clear reporting on the accessibility and transparency of AI systems further helps clinicians understand their functionality, enhancing trust in their outputs and fostering the development of user-friendly interfaces, and ensuring the safe and effective use of AI in clinical practice.

## Conclusion

Our study focused on AI intervention in primary care, providing a first systematic review of qualified reporting of RCTs using AI interventions. This study indicated significant gaps between the reporting guidelines and published articles and underscored the crucial items and aspects that were frequently overlooked in the reporting framework. It is said that the whole of medicine depends on the transparent reporting of clinical trials [74]. Our

findings may contribute to the enhancement of quality standards in AI research in primary care trials and help future clinical AI investigators to design, conduct, and analysis of higher-quality AI interventions for primary care.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (82104133).

### Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist.

[\[PDF File \(Adobe PDF File\), 93 KB-Multimedia Appendix 1\]](#)

### Multimedia Appendix 2

Search strategies.

[\[DOCX File , 27 KB-Multimedia Appendix 2\]](#)

### Multimedia Appendix 3

List and characteristics of included randomized controlled trials and protocols.

[\[DOCX File , 55 KB-Multimedia Appendix 3\]](#)

### Multimedia Appendix 4

Association between primary outcome result status and the reporting quality of each item.

[\[DOCX File , 17 KB-Multimedia Appendix 4\]](#)

### Multimedia Appendix 5

Association between trial characteristics and the percentage of adequately reported items.

[\[DOCX File , 17 KB-Multimedia Appendix 5\]](#)

### References

1. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. *Cochrane Database Syst Rev.* 2016;3(3):CD006560. [\[FREE Full text\]](#) [doi: [10.1002/14651858.CD006560.pub4](https://doi.org/10.1002/14651858.CD006560.pub4)] [Medline: [33448337](https://pubmed.ncbi.nlm.nih.gov/33448337/)]
2. Rodriguez JA, Charles J, Bates DW, Lyles C, Southworth B, Samal L. Digital healthcare equity in primary care: implementing an integrated digital health navigator. *J Am Med Inform Assoc.* 2023;30(5):965-970. [\[FREE Full text\]](#) [doi: [10.1093/jamia/ocad015](https://doi.org/10.1093/jamia/ocad015)] [Medline: [36795062](https://pubmed.ncbi.nlm.nih.gov/36795062/)]
3. Abbasgholizadeh Rahimi S, Légaré F, Sharma G, Archambault P, Zomahoun HTV, Chandavong S, et al. Application of artificial intelligence in community-based primary health care: systematic scoping review and critical appraisal. *J Med Internet Res.* 2021;23(9):e29839. [\[FREE Full text\]](#) [doi: [10.2196/29839](https://doi.org/10.2196/29839)] [Medline: [34477556](https://pubmed.ncbi.nlm.nih.gov/34477556/)]
4. Tse G, Lee Q, Chou OHI, Chung CT, Lee S, Chan JSK, et al. Healthcare big data in Hong Kong: development and implementation of artificial intelligence-enhanced predictive models for risk stratification. *Curr Probl Cardiol.* 2024;49(1 Pt B):102168. [doi: [10.1016/j.cpcardiol.2023.102168](https://doi.org/10.1016/j.cpcardiol.2023.102168)] [Medline: [37871712](https://pubmed.ncbi.nlm.nih.gov/37871712/)]
5. Uddin Y, Nair A, Shariq S, Hannan SH. Transforming primary healthcare through natural language processing and big data analytics. *BMJ.* 2023;381:948. [doi: [10.1136/bmj.p948](https://doi.org/10.1136/bmj.p948)] [Medline: [37137492](https://pubmed.ncbi.nlm.nih.gov/37137492/)]
6. Abbasgholizadeh Rahimi S, Cwintal M, Huang Y, Ghadiri P, Grad R, Poenaru D, et al. Application of artificial intelligence in shared decision making: scoping review. *JMIR Med Inform.* 2022;10(8):e36199. [\[FREE Full text\]](#) [doi: [10.2196/36199](https://doi.org/10.2196/36199)] [Medline: [35943793](https://pubmed.ncbi.nlm.nih.gov/35943793/)]
7. Lau K, Wilkinson J, Moorthy R. A web-based prediction score for head and neck cancer referrals. *Clin Otolaryngol.* 2018;43(4):1043-1049. [doi: [10.1111/coa.13098](https://doi.org/10.1111/coa.13098)] [Medline: [29543399](https://pubmed.ncbi.nlm.nih.gov/29543399/)]

8. Lisspers K, Ställberg B, Larsson K, Janson C, Müller M, Łuczko M, et al. Developing a short-term prediction model for asthma exacerbations from Swedish primary care patients' data using machine learning - based on the ARCTIC study. *Respir Med*. 2021;185:106483. [FREE Full text] [doi: [10.1016/j.rmed.2021.106483](https://doi.org/10.1016/j.rmed.2021.106483)] [Medline: [34077873](https://pubmed.ncbi.nlm.nih.gov/34077873/)]
9. Bruyndonckx R, Hens N, Verheij TJ, Aerts M, Ieven M, Butler CC, et al. Development of a prediction tool for patients presenting with acute cough in primary care: a prognostic study spanning six European countries. *Br J Gen Pract*. 2018;68(670):e342-e350. [FREE Full text] [doi: [10.3399/bjgp18X695789](https://doi.org/10.3399/bjgp18X695789)] [Medline: [29632005](https://pubmed.ncbi.nlm.nih.gov/29632005/)]
10. Yang Z, Silcox C, Sendak M, Rose S, Rehkopf D, Phillips R, et al. Advancing primary care with artificial intelligence and machine learning. *Healthc (Amst)*. 2022;10(1):100594. [doi: [10.1016/j.hjdsi.2021.100594](https://doi.org/10.1016/j.hjdsi.2021.100594)] [Medline: [34954571](https://pubmed.ncbi.nlm.nih.gov/34954571/)]
11. Mohsen F, Al-Absi HRH, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. *NPJ Digit Med*. 2023;6(1):197. [FREE Full text] [doi: [10.1038/s41746-023-00933-5](https://doi.org/10.1038/s41746-023-00933-5)] [Medline: [37880301](https://pubmed.ncbi.nlm.nih.gov/37880301/)]
12. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ*. 1996;313(7057):570-571. [FREE Full text] [doi: [10.1136/bmj.313.7057.570](https://doi.org/10.1136/bmj.313.7057.570)] [Medline: [8806240](https://pubmed.ncbi.nlm.nih.gov/8806240/)]
13. Wu L, Shang R, Sharma P, Zhou W, Liu J, Yao L, et al. Effect of a deep learning-based system on the miss rate of gastric neoplasms during upper gastrointestinal endoscopy: a single-centre, tandem, randomised controlled trial. *Lancet Gastroenterol Hepatol*. 2021;6(9):700-708. [doi: [10.1016/S2468-1253\(21\)00216-8](https://doi.org/10.1016/S2468-1253(21)00216-8)] [Medline: [34297944](https://pubmed.ncbi.nlm.nih.gov/34297944/)]
14. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol*. 2019;20(12):1645-1654. [doi: [10.1016/S1470-2045\(19\)30637-0](https://doi.org/10.1016/S1470-2045(19)30637-0)] [Medline: [31591062](https://pubmed.ncbi.nlm.nih.gov/31591062/)]
15. Mangas-Sanjuan C, de-Castro L, Cubiella J, Díez-Redondo P, Suárez A, Pellisé M, et al. Role of artificial intelligence in colonoscopy detection of advanced neoplasias : a randomized trial. *Ann Intern Med*. 2023;176(9):1145-1152. [doi: [10.7326/M22-2619](https://doi.org/10.7326/M22-2619)] [Medline: [37639723](https://pubmed.ncbi.nlm.nih.gov/37639723/)]
16. Lauritsen SM, Kristensen M, Olsen MV, Larsen MS, Lauritsen KM, Jørgensen MJ, et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nat Commun*. 2020;11(1):3852. [FREE Full text] [doi: [10.1038/s41467-020-17431-x](https://doi.org/10.1038/s41467-020-17431-x)] [Medline: [32737308](https://pubmed.ncbi.nlm.nih.gov/32737308/)]
17. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health*. 2020;2(10):e537-e548. [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30218-1](https://doi.org/10.1016/S2589-7500(20)30218-1)] [Medline: [33328048](https://pubmed.ncbi.nlm.nih.gov/33328048/)]
18. Rivera SC, Liu X, Chan A, Denniston AK, Calvert MJ, SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Lancet Digit Health*. 2020;2(10):e549-e560. [doi: [10.1016/S2589-7500\(20\)30219-3](https://doi.org/10.1016/S2589-7500(20)30219-3)] [Medline: [33328049](https://pubmed.ncbi.nlm.nih.gov/33328049/)]
19. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30-36. [FREE Full text] [doi: [10.1038/s41591-018-0307-0](https://doi.org/10.1038/s41591-018-0307-0)] [Medline: [30617336](https://pubmed.ncbi.nlm.nih.gov/30617336/)]
20. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med*. 2017;22(4):139-142. [FREE Full text] [doi: [10.1136/ebmed-2017-110713](https://doi.org/10.1136/ebmed-2017-110713)] [Medline: [28701372](https://pubmed.ncbi.nlm.nih.gov/28701372/)]
21. Institute of Medicine (US) Committee on the Future of Primary Care. Donaldson MS, Yordy KD, Lohr KN, Vanselow NA, editors. *Primary Care: America's Health in a New Era*. Washington, DC. National Academies Press; 1996.
22. Starfield B. A framework for primary care research. *J Fam Pract*. 1996;42(2):181-185. [Medline: [8606309](https://pubmed.ncbi.nlm.nih.gov/8606309/)]
23. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, et al. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. *JAMA Netw Open*. 2020;3(3):e200255. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.0255](https://doi.org/10.1001/jamanetworkopen.2020.0255)] [Medline: [32119093](https://pubmed.ncbi.nlm.nih.gov/32119093/)]
24. Seol HY, Shrestha P, Muth JF, Wi C-I, Sohn S, Ryu E, et al. Artificial intelligence-assisted clinical decision support for childhood asthma management: a randomized clinical trial. *PLoS One*. 2021;16(8):e0255261. [FREE Full text] [doi: [10.1371/journal.pone.0255261](https://doi.org/10.1371/journal.pone.0255261)] [Medline: [34339438](https://pubmed.ncbi.nlm.nih.gov/34339438/)]
25. Yao X, Rushlow DR, Inselman JW, McCoy RG, Thacher TD, Behnken EM, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med*. 2021;27(5):815-819. [doi: [10.1038/s41591-021-01335-4](https://doi.org/10.1038/s41591-021-01335-4)] [Medline: [33958795](https://pubmed.ncbi.nlm.nih.gov/33958795/)]
26. Itoh N, Mishima H, Yoshida Y, Yoshida M, Oka H, Matsudaira K. Evaluation of the effect of patient education and strengthening exercise therapy using a mobile messaging app on work productivity in Japanese patients with chronic low back pain: open-label, randomized, parallel-group trial. *JMIR Mhealth Uhealth*. 2022;10(5):e35867. [FREE Full text] [doi: [10.2196/35867](https://doi.org/10.2196/35867)] [Medline: [35576560](https://pubmed.ncbi.nlm.nih.gov/35576560/)]
27. Piette JD, Newman S, Krein SL, Marinec N, Chen J, Williams DA, et al. Patient-centered pain care using artificial intelligence and mobile health tools: a randomized comparative effectiveness trial. *JAMA Intern Med*. 2022;182(9):975-983. [FREE Full text] [doi: [10.1001/jamainternmed.2022.3178](https://doi.org/10.1001/jamainternmed.2022.3178)] [Medline: [35939288](https://pubmed.ncbi.nlm.nih.gov/35939288/)]
28. Rushlow DR, Croghan IT, Inselman JW, Thacher TD, Friedman PA, Yao X, et al. Clinician adoption of an artificial intelligence algorithm to detect left ventricular systolic dysfunction in primary care. *Mayo Clin Proc*. 2022;97(11):2076-2085. [FREE Full text] [doi: [10.1016/j.mayocp.2022.04.008](https://doi.org/10.1016/j.mayocp.2022.04.008)] [Medline: [36333015](https://pubmed.ncbi.nlm.nih.gov/36333015/)]

29. Manz CR, Zhang Y, Chen K, Long Q, Small DS, Evans CN, et al. Long-term effect of machine learning-triggered behavioral nudges on serious illness conversations and end-of-life outcomes among patients with cancer: a randomized clinical trial. *JAMA Oncol.* 2023;9(3):414-418. [FREE Full text] [doi: [10.1001/jamaoncol.2022.6303](https://doi.org/10.1001/jamaoncol.2022.6303)] [Medline: [36633868](https://pubmed.ncbi.nlm.nih.gov/36633868/)]
30. Wei MT, Shankar U, Parvin R, Abbas SH, Chaudhary S, Friedlander Y, et al. Evaluation of computer-aided detection during colonoscopy in the community (AI-SEE): a multicenter randomized clinical trial. *Am J Gastroenterol.* 2023;118(10):1841-1847. [doi: [10.14309/ajg.0000000000002239](https://doi.org/10.14309/ajg.0000000000002239)] [Medline: [36892545](https://pubmed.ncbi.nlm.nih.gov/36892545/)]
31. Yang J, Cui Z, Liao X, He X, Wang L, Wei D, et al. Effects of a feedback intervention on antibiotic prescription control in primary care institutions based on a health information system: a cluster randomized cross-over controlled trial. *J Glob Antimicrob Resist.* 2023;33:51-60. [FREE Full text] [doi: [10.1016/j.jgar.2023.02.006](https://doi.org/10.1016/j.jgar.2023.02.006)] [Medline: [36828121](https://pubmed.ncbi.nlm.nih.gov/36828121/)]
32. Lin CS, Liu WT, Tsai DJ, Lou YS, Chang CH, Lee CC, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. *Nat Med.* 2024;30(5):1461-1470. [doi: [10.1038/s41591-024-02961-4](https://doi.org/10.1038/s41591-024-02961-4)] [Medline: [38684860](https://pubmed.ncbi.nlm.nih.gov/38684860/)]
33. Benrimoh D, Whitmore K, Richard M, Golden G, Perlman K, Jalali S, et al. Artificial Intelligence in Depression - Medication Enhancement (AID-ME): a cluster randomized trial of a deep learning enabled clinical decision support system for personalized depression treatment selection and management. *Clinical Decision Support Systems.* 2024. [FREE Full text]
34. Olano-Espinosa E, Avila-Tomas JF, Minue-Lorenzo C, Matilla-Pardo B, Serrano Serrano ME, Martinez-Suberviola FJ, et al. Dejal@ Group. Effectiveness of a conversational chatbot (Dejal@bot) for the adult population to quit smoking: pragmatic, multicenter, controlled, randomized clinical trial in primary care. *JMIR Mhealth Uhealth.* 2022;10(6):e34273. [FREE Full text] [doi: [10.2196/34273](https://doi.org/10.2196/34273)] [Medline: [35759328](https://pubmed.ncbi.nlm.nih.gov/35759328/)]
35. Sandal LF, Bach K, Øverås CK, Svendsen MJ, Dalager T, Stejnicher Drongstrup Jensen J, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfback randomized clinical trial. *JAMA Intern Med.* 2021;181(10):1288-1296. [FREE Full text] [doi: [10.1001/jamainternmed.2021.4097](https://doi.org/10.1001/jamainternmed.2021.4097)] [Medline: [34338710](https://pubmed.ncbi.nlm.nih.gov/34338710/)]
36. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw Open.* 2024;7(10):e2440969. [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.40969](https://doi.org/10.1001/jamanetworkopen.2024.40969)] [Medline: [39466245](https://pubmed.ncbi.nlm.nih.gov/39466245/)]
37. Granviken F, Meisingset I, Bach K, Bones AF, Simpson MR, Hill JC, et al. Personalised decision support in the management of patients with musculoskeletal pain in primary physiotherapy care: a cluster randomised controlled trial (the SupportPrim project). *Pain.* 2024. [doi: [10.1097/j.pain.0000000000003456](https://doi.org/10.1097/j.pain.0000000000003456)] [Medline: [39432806](https://pubmed.ncbi.nlm.nih.gov/39432806/)]
38. Thiruvengadam NR, Solaimani P, Shrestha M, Buller S, Carson R, Reyes-Garcia B, et al. The efficacy of real-time computer-aided detection of colonic neoplasia in community practice: a pragmatic randomized controlled trial. *Clin Gastroenterol Hepatol.* 2024;22(11):2221-2230.e15. [FREE Full text] [doi: [10.1016/j.cgh.2024.02.021](https://doi.org/10.1016/j.cgh.2024.02.021)] [Medline: [38437999](https://pubmed.ncbi.nlm.nih.gov/38437999/)]
39. Nayak A, Vakili S, Nayak K, Nikolov M, Chiu M, Sosseinheimer P, et al. Use of voice-based conversational artificial intelligence for basal insulin prescription management among patients with type 2 diabetes: a randomized clinical trial. *JAMA Netw Open.* 2023;6(12):e2340232. [FREE Full text] [doi: [10.1001/jamanetworkopen.2023.40232](https://doi.org/10.1001/jamanetworkopen.2023.40232)] [Medline: [38039007](https://pubmed.ncbi.nlm.nih.gov/38039007/)]
40. Alcoceba-Herrero I, Coco-Martín MB, Jiménez-Pérez JM, Leal-Vega L, Martín-Gutiérrez A, Dueñas-Gutiérrez C, et al. Randomized controlled trial to assess the feasibility of a novel clinical decision support system based on the automatic generation of alerts through remote patient monitoring. *J Clin Med.* 2024;13(19):5974. [FREE Full text] [doi: [10.3390/jcm13195974](https://doi.org/10.3390/jcm13195974)] [Medline: [39408035](https://pubmed.ncbi.nlm.nih.gov/39408035/)]
41. Ortiz O, Daca-Alvarez M, Rivero-Sanchez L, Gimeno-Garcia AZ, Carrillo-Palau M, Alvarez V, et al. TIMELY study group. An artificial intelligence-assisted system versus white light endoscopy alone for adenoma detection in individuals with Lynch syndrome (TIMELY): an international, multicentre, randomised controlled trial. *Lancet Gastroenterol Hepatol.* 2024;9(9):802-810. [doi: [10.1016/S2468-1253\(24\)00187-0](https://doi.org/10.1016/S2468-1253(24)00187-0)] [Medline: [39033774](https://pubmed.ncbi.nlm.nih.gov/39033774/)]
42. Persell SD, Karmali KN, Stein N, Li J, Peprah YA, Lipiszko D, et al. Design of a randomized controlled trial comparing a mobile phone-based hypertension health coaching application to home blood pressure monitoring alone: The Smart Hypertension Control Study. *Contemp Clin Trials.* 2018;73:92-97. [doi: [10.1016/j.cct.2018.08.013](https://doi.org/10.1016/j.cct.2018.08.013)] [Medline: [30172038](https://pubmed.ncbi.nlm.nih.gov/30172038/)]
43. Fontil V, Khoong EC, Hoskote M, Radcliffe K, Ratanawongsa N, Lyles CR, et al. Evaluation of a health information technology-enabled collective intelligence platform to improve diagnosis in primary care and urgent care settings: protocol for a pragmatic randomized controlled trial. *JMIR Res Protoc.* 2019;8(8):e13151. [FREE Full text] [doi: [10.2196/13151](https://doi.org/10.2196/13151)] [Medline: [31389337](https://pubmed.ncbi.nlm.nih.gov/31389337/)]
44. Sandal LF, Stochkendahl MJ, Svendsen MJ, Wood K, Øverås CK, Nordstoga AL, et al. An app-delivered self-management program for people with low back pain: protocol for the selfback randomized controlled trial. *JMIR Res Protoc.* 2019;8(12):e14720. [FREE Full text] [doi: [10.2196/14720](https://doi.org/10.2196/14720)] [Medline: [31793897](https://pubmed.ncbi.nlm.nih.gov/31793897/)]
45. Aguilera A, Figueroa CA, Hernandez-Ramos R, Sarkar U, Cembali A, Gomez-Pathak L, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. *BMJ Open.* 2020;10(8):e034723. [FREE Full text] [doi: [10.1136/bmjopen-2019-034723](https://doi.org/10.1136/bmjopen-2019-034723)] [Medline: [32819981](https://pubmed.ncbi.nlm.nih.gov/32819981/)]

46. Yao X, McCoy RG, Friedman PA, Shah ND, Barry BA, Behnken EM, et al. ECG AI-guided screening for low ejection fraction (EAGLE): rationale and design of a pragmatic cluster randomized trial. *Am Heart J*. 2020;219:31-36. [doi: [10.1016/j.ahj.2019.10.007](https://doi.org/10.1016/j.ahj.2019.10.007)] [Medline: [31710842](https://pubmed.ncbi.nlm.nih.gov/31710842/)]
47. Lauffenburger JC, Yom-Tov E, Keller PA, McDonnell ME, Bessette LG, Fontanet CP, et al. Reinforcement learning to improve non-adherence for diabetes treatments by optimising response and customising engagement (REINFORCE): study protocol of a pragmatic randomised trial. *BMJ Open*. 2021;11(12):e052091. [FREE Full text] [doi: [10.1136/bmjopen-2021-052091](https://doi.org/10.1136/bmjopen-2021-052091)] [Medline: [34862289](https://pubmed.ncbi.nlm.nih.gov/34862289/)]
48. Cheng W, Zhang Z, Hoelzer S, Tang W, Liang Y, Du Y, et al. Evaluation of a village-based digital health kiosks program: a protocol for a cluster randomized clinical trial. *Digit Health*. 2022;8:20552076221129100. [FREE Full text] [doi: [10.1177/20552076221129100](https://doi.org/10.1177/20552076221129100)] [Medline: [36211797](https://pubmed.ncbi.nlm.nih.gov/36211797/)]
49. Chang Y, Yao Y, Cui Z, Yang G, Li D, Wang L, et al. Changing antibiotic prescribing practices in outpatient primary care settings in China: study protocol for a health information system-based cluster-randomised crossover controlled trial. *PLoS One*. 2022;17(1):e0259065. [FREE Full text] [doi: [10.1371/journal.pone.0259065](https://doi.org/10.1371/journal.pone.0259065)] [Medline: [34995279](https://pubmed.ncbi.nlm.nih.gov/34995279/)]
50. Han JED, Liu X, Bunce C, Douiri A, Vale L, Blandford A, et al. Teleophthalmology-enabled and artificial intelligence-ready referral pathway for community optometry referrals of retinal disease (HERMES): a cluster randomised superiority trial with a linked diagnostic accuracy study-hermes study report 1-study protocol. *BMJ Open*. 2022;12(2):e055845. [FREE Full text] [doi: [10.1136/bmjopen-2021-055845](https://doi.org/10.1136/bmjopen-2021-055845)] [Medline: [35105593](https://pubmed.ncbi.nlm.nih.gov/35105593/)]
51. Kleiman MJ, Plewes AD, Owora A, Grout RW, Dexter PR, Fowler NR, et al. Digital detection of dementia (D): a study protocol for a pragmatic cluster-randomized trial examining the application of patient-reported outcomes and passive clinical decision support systems. *Trials*. 2022;23(1):868. [FREE Full text] [doi: [10.1186/s13063-022-06809-5](https://doi.org/10.1186/s13063-022-06809-5)] [Medline: [36221141](https://pubmed.ncbi.nlm.nih.gov/36221141/)]
52. Laranjo L, Shaw T, Trivedi R, Thomas S, Charlston E, Klimis H, et al. Coordinating health care with artificial intelligence-supported technology for patients with atrial fibrillation: protocol for a randomized controlled trial. *JMIR Res Protoc*. 2022;11(4):e34470. [FREE Full text] [doi: [10.2196/34470](https://doi.org/10.2196/34470)] [Medline: [35416784](https://pubmed.ncbi.nlm.nih.gov/35416784/)]
53. Marshall BDL, Alexander-Scott N, Yedinak JL, Hallowell BD, Goedel WC, Allen B, et al. Preventing overdose using information and data from the environment (PROVIDENT): protocol for a randomized, population-based, community intervention trial. *Addiction*. 2022;117(4):1152-1162. [FREE Full text] [doi: [10.1111/add.15731](https://doi.org/10.1111/add.15731)] [Medline: [34729851](https://pubmed.ncbi.nlm.nih.gov/34729851/)]
54. Ru X, Zhu L, Ma Y, Wang T, Pan Z. Effect of an artificial intelligence-assisted tool on non-valvular atrial fibrillation anticoagulation management in primary care: protocol for a cluster randomized controlled trial. *Trials*. 2022;23(1):316. [FREE Full text] [doi: [10.1186/s13063-022-06250-8](https://doi.org/10.1186/s13063-022-06250-8)] [Medline: [35428315](https://pubmed.ncbi.nlm.nih.gov/35428315/)]
55. Soto-Ruiz N, Escalada-Hernández P, Martín-Rodríguez LS, Ferraz-Torres M, García-Vivar C. Web-based personalized intervention to improve quality of life and self-efficacy of long-term breast cancer survivors: study protocol for a randomized controlled trial. *Int J Environ Res Public Health*. 2022;19(19):12240. [FREE Full text] [doi: [10.3390/ijerph191912240](https://doi.org/10.3390/ijerph191912240)] [Medline: [36231542](https://pubmed.ncbi.nlm.nih.gov/36231542/)]
56. Wong CK, Hai JJ, Lau YM, Zhou M, Lui HW, Lau KK, et al. Protocol for home-based solution for remote atrial fibrillation screening to prevent recurrence stroke (HUA-TUO AF Trial): a randomised controlled trial. *BMJ Open*. 2022;12(7):e053466. [FREE Full text] [doi: [10.1136/bmjopen-2021-053466](https://doi.org/10.1136/bmjopen-2021-053466)] [Medline: [35840293](https://pubmed.ncbi.nlm.nih.gov/35840293/)]
57. Gordon BR, Qiu L, Doerksen SE, Kanski B, Lorenzo A, Truica CI, et al. Addressing metastatic individuals everyday: rationale and design of the nurse AMIE for Amazon Echo Show trial among metastatic breast cancer patients. *Contemp Clin Trials Commun*. 2023;32:101058. [FREE Full text] [doi: [10.1016/j.conctc.2023.101058](https://doi.org/10.1016/j.conctc.2023.101058)] [Medline: [36698743](https://pubmed.ncbi.nlm.nih.gov/36698743/)]
58. Heinzen EP, Wilson PM, Storlie CB, Demuth GO, Asai SW, Schaeferle GM, et al. Impact of a machine learning algorithm on time to palliative care in a primary care population: protocol for a stepped-wedge pragmatic randomized trial. *BMC Palliat Care*. 2023;22(1):9. [FREE Full text] [doi: [10.1186/s12904-022-01113-0](https://doi.org/10.1186/s12904-022-01113-0)] [Medline: [36737744](https://pubmed.ncbi.nlm.nih.gov/36737744/)]
59. Doe G, El-Emir E, Edwards GD, Topalovic M, Evans RA, Russell R, et al. Comparing performance of primary care clinicians in the interpretation of SPIROmetry with or without artificial intelligence decision support software (SPIRO-AID): a protocol for a randomised controlled trial. *BMJ Open*. 2024;14(6):e086736. [FREE Full text] [doi: [10.1136/bmjopen-2024-086736](https://doi.org/10.1136/bmjopen-2024-086736)] [Medline: [38950987](https://pubmed.ncbi.nlm.nih.gov/38950987/)]
60. Avila-Tomas JF, Olano-Espinosa E, Minué-Lorenzo C, Martinez-Suberbiola FJ, Matilla-Pardo B, Serrano-Serrano ME, et al. Group Dej@lo. Effectiveness of a chat-bot for the adult population to quit smoking: protocol of a pragmatic clinical trial in primary care (Dejal@). *BMC Med Inform Decis Mak*. Dec 03, 2019;19(1):249. [FREE Full text] [doi: [10.1186/s12911-019-0972-z](https://doi.org/10.1186/s12911-019-0972-z)] [Medline: [31796061](https://pubmed.ncbi.nlm.nih.gov/31796061/)]
61. Kwan YH, Yoon S, Tai BC, Tan CS, Phang JK, Tan WB, et al. Empowering patients with comorbid diabetes and hypertension through a multi-component intervention of mobile app, health coaching and shared decision-making: protocol for an effectiveness-implementation of randomised controlled trial. *PLoS One*. 2024;19(2):e0296338. [FREE Full text] [doi: [10.1371/journal.pone.0296338](https://doi.org/10.1371/journal.pone.0296338)] [Medline: [38408067](https://pubmed.ncbi.nlm.nih.gov/38408067/)]
62. Zhang H, Huo X, Ren L, Lu J, Li J, Zheng X, et al. Design and rationale of the Comprehensive Intelligent Hypertension Management System (CHESS) evaluation study: a cluster randomized controlled trial for hypertension management in primary care. *Am Heart J*. 2024;273:90-101. [FREE Full text] [doi: [10.1016/j.ahj.2024.03.018](https://doi.org/10.1016/j.ahj.2024.03.018)] [Medline: [38575049](https://pubmed.ncbi.nlm.nih.gov/38575049/)]

63. Kung TH, Cheatham M, Medenilla A, Sillos C, de Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for ai-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198. [FREE Full text] [doi: [10.1371/journal.pdig.0000198](https://doi.org/10.1371/journal.pdig.0000198)] [Medline: [36812645](https://pubmed.ncbi.nlm.nih.gov/36812645/)]
64. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. *Nature*. 2023;614(7947):224-226. [doi: [10.1038/d41586-023-00288-7](https://doi.org/10.1038/d41586-023-00288-7)] [Medline: [36737653](https://pubmed.ncbi.nlm.nih.gov/36737653/)]
65. Moura L, Jones DT, Sheikh IS, Murphy S, Kalfin M, Kummer BR, et al. Implications of large language models for quality and efficiency of neurologic care: emerging issues in neurology. *Neurology*. 2024;102(11):e209497. [doi: [10.1212/WNL.0000000000209497](https://doi.org/10.1212/WNL.0000000000209497)] [Medline: [38759131](https://pubmed.ncbi.nlm.nih.gov/38759131/)]
66. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. *Inf Fusion*. 2023;99:101805. [doi: [10.1016/j.inffus.2023.101805](https://doi.org/10.1016/j.inffus.2023.101805)]
67. Hinterwimmer F, Lazic I, Suren C, Hirschmann MT, Pohlig F, Rueckert D, et al. Machine learning in knee arthroplasty: specific data are key—a systematic review. *Knee Surg Sports Traumatol Arthrosc*. 2022;30(2):376-388. [FREE Full text] [doi: [10.1007/s00167-021-06848-6](https://doi.org/10.1007/s00167-021-06848-6)] [Medline: [35006281](https://pubmed.ncbi.nlm.nih.gov/35006281/)]
68. Wehbe RM, Katsaggelos AK, Hammond KJ, Hong H, Ahmad FS, Ouyang D, et al. Deep learning for cardiovascular imaging: a review. *JAMA Cardiol*. 2023;8(11):1089-1098. [doi: [10.1001/jamacardio.2023.3142](https://doi.org/10.1001/jamacardio.2023.3142)] [Medline: [37728933](https://pubmed.ncbi.nlm.nih.gov/37728933/)]
69. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *J Am Med Inform Assoc*. 2020;27(4):592-600. [FREE Full text] [doi: [10.1093/jamia/ocz229](https://doi.org/10.1093/jamia/ocz229)] [Medline: [32106285](https://pubmed.ncbi.nlm.nih.gov/32106285/)]
70. Adler-Milstein J, Aggarwal N, Ahmed M, Castner J, Evans BJ, Gonzalez AA, et al. Meeting the moment: addressing barriers and facilitating clinical adoption of artificial intelligence in medical diagnosis. *NAM Perspect*. 2022;2022:10.31478/202209c. [FREE Full text] [doi: [10.31478/202209c](https://doi.org/10.31478/202209c)] [Medline: [36713769](https://pubmed.ncbi.nlm.nih.gov/36713769/)]
71. Maadi M, Akbarzadeh Khorshidi H, Aickelin U. A review on human-AI interaction in machine learning and insights for medical applications. *Int J Environ Res Public Health*. 2021;18(4):2121. [FREE Full text] [doi: [10.3390/ijerph18042121](https://doi.org/10.3390/ijerph18042121)] [Medline: [33671609](https://pubmed.ncbi.nlm.nih.gov/33671609/)]
72. Kim HM, Kang H, Lee C, Park JH, Chung MK, Kim M, et al. Evaluation of the clinical efficacy and trust in AI-assisted embryo ranking: survey-based prospective study. *J Med Internet Res*. 2024;26:e52637. [FREE Full text] [doi: [10.2196/52637](https://doi.org/10.2196/52637)] [Medline: [38830209](https://pubmed.ncbi.nlm.nih.gov/38830209/)]
73. Champion JR, O'Connor DB, Lahiff C. Human-artificial intelligence interaction in gastrointestinal endoscopy. *World J Gastrointest Endosc*. 2024;16(3):126-135. [FREE Full text] [doi: [10.4253/wjge.v16.i3.126](https://doi.org/10.4253/wjge.v16.i3.126)] [Medline: [38577646](https://pubmed.ncbi.nlm.nih.gov/38577646/)]
74. Rennie D. CONSORT revised—improving the reporting of randomized trials. *JAMA*. 2001;285(15):2006-2007. [doi: [10.1001/jama.285.15.2006](https://doi.org/10.1001/jama.285.15.2006)] [Medline: [11308440](https://pubmed.ncbi.nlm.nih.gov/11308440/)]

## Abbreviations

**AI:** artificial intelligence

**CDSS:** clinical decision support system

**CONSORT-AI:** Consolidated Standards of Reporting Trials—Artificial Intelligence

**LLM:** large language model

**ML:** machine learning

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analysis

**RCT:** randomized controlled trial

**SPIRIT-AI:** Standard Protocol Items: Recommendations for Interventional Trials—Artificial Intelligence

*Edited by T de Azevedo Cardoso; submitted 25.01.24; peer-reviewed by VT Obulareddy, X Ying, C-M Chu, S Papineni; comments to author 08.11.24; revised version received 21.12.24; accepted 22.01.25; published 25.02.25*

*Please cite as:*

Zhong J, Zhu T, Huang Y

Reporting Quality of AI Intervention in Randomized Controlled Trials in Primary Care: Systematic Review and Meta-Epidemiological Study

*J Med Internet Res* 2025;27:e56774

URL: <https://www.jmir.org/2025/1/e56774>

doi: [10.2196/56774](https://doi.org/10.2196/56774)

PMID: [39998876](https://pubmed.ncbi.nlm.nih.gov/39998876/)

©Jinja Zhong, Ting Zhu, Yafang Huang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 25.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.