

Review

# AI-Based Clinical Decision Support Systems in Primary Care: Systematic Review of Clinical Implementation

Florian Naye<sup>1,2</sup>, PhD; Steven Ouellet<sup>1†</sup>, PhD; Farzaneh Yousefi<sup>1</sup>, PhD; Chloé Cachinho<sup>1</sup>, MSc; Frédéric Bergeron<sup>3</sup>, MSI; Onil Bhattacharyya<sup>4,5</sup>, PhD; Vess Stamenova<sup>4,6</sup>, PhD; Maxime Sasseville<sup>1,2</sup>, PhD

<sup>1</sup>Faculty of Nursing Sciences, Université Laval, Québec, QC, Canada

<sup>2</sup>VITAM Research Center on Sustainable Health, Québec, QC, Canada

<sup>3</sup>Bibliothèque de l'Université Laval, Université Laval, Québec, QC, Canada

<sup>4</sup>Centre for Digital Health Evaluation, Women's College Hospital, Toronto, QC, Canada

<sup>5</sup>Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

<sup>6</sup>Department of Information Technology Management, Toronto Metropolitan University, Toronto, ON, Canada

<sup>†</sup>deceased

## Corresponding Author:

Maxime Sasseville, PhD

Faculty of Nursing Sciences

Université Laval

1050, avenue de la Médecine

Québec, QC G1V 0A6

Canada

Phone: 1 418-290-3089

Email: [maxime.sasseville@fsi.ulaval.ca](mailto:maxime.sasseville@fsi.ulaval.ca)

## Abstract

**Background:** AI-based clinical decision support systems (AI-CDSSs) can help primary care clinicians provide higher quality care amid rising case complexity and resource constraints. However, evidence regarding their effectiveness and safety in primary care practice is lacking.

**Objective:** We synthesized evidence on the effectiveness and safety of AI-CDSSs compared with usual care or non-AI-based systems in primary care settings.

**Methods:** We conducted a systematic review following the Cochrane Handbook for Systematic Reviews of Interventions and reported the results according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines. We included experimental and quasi-experimental studies evaluating AI-CDSSs used by health care professionals in primary care settings, compared with usual care or non-AI-based decision support systems, and that reported clinician-, patient-, or system-level outcomes, or safety. An information specialist designed the search strategies and systematically reviewed MEDLINE, Embase, CINAHL, Web of Science, and CENTRAL from inception to June 27, 2025. Reviewers independently performed the selection and extraction processes. Data extraction was informed by the DECIDE-AI and CONSORT-AI (Consolidated Standards of Reporting Trials-AI) reporting guidelines and the APPRAISE-AI tool. We independently assessed the risk of bias using Risk of Bias-2 or Risk of Bias in Non-Randomized Studies of Interventions. Owing to substantial heterogeneity, we conducted a synthesis without meta-analysis and explored heterogeneity using an adapted Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability-informed implementation-context classification. The review protocol was registered in PROSPERO (CRD420251119225).

**Results:** We identified 4074 records through database searches and 12 additional records through citation searching and trial registration checking. We included 14 records on 11 distinct studies, including 4 experimental and 7 quasi-experimental studies. The risk of bias was high in all experimental studies and serious to critical in quasi-experimental studies. AI-CDSSs showed more consistent positive effects on workflow-related outcomes in structured and protocolized decision contexts, particularly for screening, referral, and alert optimization tasks. Patient-level effects were limited and inconsistent, particularly in more complex contexts involving heterogeneous conditions or preference-sensitive decisions. System-level outcomes were heterogeneous, with isolated improvements in service delivery and utilization outcomes. Safety evaluation was rare and poorly specified. Only 2 studies prespecified safety outcomes, and the reported safety information mainly involved technical malfunctions or self-reported side effects. No study reported clinician interaction measures such as agreement with, override

of, or reliance on AI outputs. Interpretation of the findings is limited by substantial heterogeneity, short follow-up periods, predominantly high or critical risk of bias, and evidence derived exclusively from high-income countries.

**Conclusions:** AI-CDSSs in primary care may provide more consistent benefits for workflow-related and process-level outcomes in structured and protocolized decision contexts, while evidence on safety remains insufficient to support confident routine implementation. This systematic review reveals an important gap between the rapid development of AI-CDSSs and the lack of robust, comparable evidence to support their reliable use in primary care.

*J Med Internet Res* 2026;28:e96882; doi: [10.2196/96882](https://doi.org/10.2196/96882)

**Keywords:** artificial intelligence ; clinical decision support systems ; primary health care ; systematic review ; patient safety ; effectiveness ; implementation; Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA

## Introduction

Primary care is central to high-performing health systems, contributing to improved population health outcomes, greater equity, and reduced avoidable hospitalizations [1-4]. Yet, primary care is increasingly strained [5], with rising case complexity, workforce shortages, and fragmented information systems placing increasing pressure on clinicians [6-9]. At the same time, the volume of clinical evidence continues to expand at a pace that exceeds clinicians' capacity to remain aligned with current best practices [10]. These challenges have accelerated interest in digital tools to support clinical decision-making and workflow in primary care settings [11, 12].

Clinical decision support systems (CDSSs) can assist clinicians by generating patient-specific recommendations at the point of care [13]. Meta-analyses of controlled trials indicate that traditional CDSSs achieve small to moderate improvements in targeted processes of care (eg, prescribing, test ordering, documentation, and vaccination) [14]. However, these effects do not consistently translate into better clinical outcomes, with evidence indicating small and inconsistent improvements at the patient level [14]. This gap highlights a key limitation of traditional CDSSs, as they have limited capacity to account for variability in patient characteristics and clinical contexts [15,16]. AI methods can be integrated into CDSSs to improve data-driven precision and enable more individualized recommendations [15,16].

Most evidence on AI-based CDSSs (AI-CDSSs) remains concentrated in the early stages of the evaluation pathway [17-19]. Most AI-CDSS studies focus on model development and validation, often reporting strong technical performance [20-24]. Performance evaluations are typically conducted using curated datasets and predefined tasks that do not reflect the complexity and variability of routine clinical care [6,25, 26]. Performance assessments conducted in closed contexts therefore poorly predict how AI-CDSSs will perform when deployed in clinical practice.

To be effective, AI-CDSSs must align with clinical workflows and support, rather than disrupt, the work of clinicians [7,27]. When there is a misalignment, decision support can introduce additional clinical or administrative steps into already time-constrained consultations or generate recommendations that are difficult to implement [28]. Clinical misalignments commonly lead to nonadoption, inappropriate

use, and long-term abandonment of AI-CDSSs [7,29-31]. Even systems with strong technical performance have failed to produce clinical impact because they were not effectively integrated into clinical practice [32].

AI-CDSSs influence clinical decision-making in ways that shape diagnosis and management, raising clinical safety considerations [33]. Errors in predictions or recommendations, if not identified or appropriately contextualized, can lead to inappropriate or delayed care [34,35]. How clinicians interact with AI-CDSSs, such as relying on recommendations without sufficient critical appraisal or disregarding relevant outputs, can shape how these risks materialize in practice [36]. Data from AI incident databases document thousands of AI-related incidents and hazards reported over the past decade, emphasizing the scale of these concerns [37]. However, the processes through which risks are identified, monitored, and mitigated in clinical settings remain insufficiently characterized [37-39], limiting the ability to determine whether AI-CDSSs can be safely integrated into care.

The rapid development, increasing availability, and commercialization of AI-CDSSs are accelerating their clinical use [40,41]. This pace of deployment could exceed the availability of effectiveness and safety evidence in clinical settings [19,42-47]. When evaluation does not keep pace with deployment, systems are introduced into care without adequate evaluation in routine practice [45,48,49], limiting impact assessment on clinical decision-making and health system oversight [43,50]. Previous reviews have mapped AI use in primary care [24,51] and highlighted barriers to AI-CDSS implementation, including workflow integration and feasibility [52,53], but have not systematically synthesized whether AI-CDSSs improve outcomes or how safe they are compared with usual care in routine primary care settings. This evidence gap aligns with recent calls for clinical evaluation of AI-CDSSs to support implementation decisions in health care [41,44,45,54,55]. The aim of this review was to synthesize evidence on the effectiveness and safety of AI-CDSSs compared with usual care or non-AI-based systems in primary care settings.

## Methods

### Study Design

We conducted a systematic review following the Cochrane Handbook for Systematic Reviews of Interventions [56] and reported the results according to PRISMA (Preferred

Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [57]. The completed PRISMA 2020 checklist is provided in [Checklist 1](#). We registered the protocol in PROSPERO [58]. Following protocol registration, departures from the registered methods were limited and are summarized in [Multimedia Appendix 1](#). These consisted of planned analyses that were not undertaken because the characteristics of the included evidence did not support their meaningful application (Risk of Bias due to Missing Evidence and Grading of Recommendations Assessment, Development, and Evaluation [GRADE]), and 1 additional exploratory

analysis introduced post hoc to facilitate interpretation of heterogeneity by classifying implementation contexts using an operationalized approach informed by the Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework.

Eligibility Criteria

We present the eligibility criteria according to the population, intervention, comparator, and outcomes (PICO) framework [59] in [Table 1](#).

Table 1. Eligibility criteria based on the PICO<sup>a</sup> framework.

| Category                | Eligibility criteria                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Population              | <ul style="list-style-type: none"><li>Health care professionals (physicians, nurses, and allied health) or clinical teams.</li><li>Exclusion: only students.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Intervention            | <ul style="list-style-type: none"><li>AI-CDSS<sup>b</sup>.</li><li>CDSSs are commonly defined as “computer systems designed to impact clinician decision making about individual patients at the point in time that these decisions are made” [13]. When CDSSs incorporate AI methods such as machine learning, they are referred to as AI-CDSS.</li><li>For this synthesis, we defined AI-CDSSs as systems meeting all of the following criteria: (1) generates patient-specific outputs, (2) produces decision-directed outputs (including recommendation, prioritization, or optimization), (3) is used at or near the point of care within clinical workflow, (4) is clinician-facing, and (5) incorporates AI-based methods.</li><li>Exclusion: presence of co-interventions (ie, additional interventions delivered alongside the AI-CDSS).</li></ul> |
| Comparators             | <ul style="list-style-type: none"><li>Standard care with or without traditional (non-AI) CDSS.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Outcomes                | <ul style="list-style-type: none"><li>Any quantitative outcomes at the patient level (eg, mortality and readmission), clinician level (eg, adherence and satisfaction), or system level (eg, cost and workflow).</li><li>Adverse outcomes and unintended consequences.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Study designs           | <ul style="list-style-type: none"><li>Randomized controlled trials, nonrandomized controlled trials, mixed methods studies, prospective cohort studies, and retrospective cohort studies with AI-CDSS clinical integration.</li><li>Exclusion: retrospective cohort studies without AI-CDSS clinical integration, reviews, editorials, conference abstracts, preprints, and protocols.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Setting                 | <ul style="list-style-type: none"><li>Primary care setting.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Language of publication | <ul style="list-style-type: none"><li>English and French.</li></ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

<sup>a</sup>PICO: population, intervention, comparator, and outcomes.  
<sup>b</sup>AI-CDSS: AI-based clinical decision support system.

Information Sources and Search Strategy

We developed a search strategy in collaboration with an experienced librarian (FB) and reported it in accordance with the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) guidelines [60] ([Checklist 2](#)). We tested the sensitivity of the strategy [61] using studies identified in a previous scoping review [51] that were expected to be retrieved by our strategy. The search was not updated after the initial sensitivity testing. We provide the full electronic search strategies for all databases in [Multimedia Appendix 2](#).

We searched the following databases on June 27, 2025, without validated filters or date and language restrictions: MEDLINE (Ovid), Embase (Elsevier), CINAHL (EBSCO-host), Web of Science Core Collection (Clarivate), and the CENTRAL. We combined database-specific controlled vocabulary (eg, MeSH and Emtree) and free-text terms related to AI and CDSS with terms capturing study designs relevant to effectiveness and safety (ie, randomized, quasi-experimental, and observational studies). We did not search for gray literature sources (eg, conference proceedings and

institutional reports) because of the focus on peer-reviewed evidence. However, we searched the trial registries of included studies to identify ongoing or unpublished results. We conducted backward and forward citation searching [62] of all included studies and relevant reviews identified during the selection process, using Scopus on August 6, 2025.

All retrieved records were imported into the management software Covidence (Veritas Health Innovation), where duplicates were removed using its automated function.

Selection and Data Collection Processes

We conducted calibration training on eligibility criteria interpretation among reviewers (FN, MS, and SO) using the first 100 references identified. If interrater agreement (percent agreement) was below 80% [63], we clarified the eligibility criteria through discussion and a think-aloud approach [64] and conducted a new training. Pairs of reviewers (FN, MS, and SO) applied the eligibility criteria to independently screen the titles and abstracts of all potentially eligible citations and then to full-text articles retrieved. We resolved discrepancies by team consensus. We documented reasons for full-text

exclusion and reported them using the PRISMA 2020 flow diagram.

## Data Items

We developed a standardized data extraction form informed by existing reporting guidelines, including DECIDE-AI for early-stage clinical evaluation of AI-CDSSs [65], CONSORT-AI (Consolidated Standards of Reporting Trials-AI) for clinical trials of AI interventions [66], and the APPRAISE-AI tool for quantitative evaluation of AI-CDSS studies [67]. Given the small number of eligible studies, the form was pilot tested on the most detailed study [68]. Two independent reviewers (FN and SO) performed the data extraction, and disagreements were resolved by a third reviewer (MS). Extracted variables covered study characteristics (eg, study period and number of participating institutions), participant characteristics (eg, sample size and baseline features), and intervention details (eg, purpose and outputs). We also collected data on outcomes and effects (eg, measurement instruments and between-group comparisons) and safety (eg, risk mitigation strategies and reported harms or unintended effects).

## Risk of Bias Assessment

We assessed individually randomized and cluster-randomized trials using the corresponding versions of Risk of Bias-2 [69] and quasi-experimental studies using ROBINS-I (version 2) [70]. Two reviewers (FN and SO) independently evaluated the risk of bias for the main outcome of each study, and disagreements were resolved by team consensus.

## Synthesis Method

Owing to substantial heterogeneity across studies, we did not conduct a meta-analysis [71,72]. First, there was limited evidence for any given comparison, with most outcomes reported in only 1 or 2 studies, resulting in minimal overlap of comparable effect estimates [72]. Second, studies exhibited marked clinical and methodological diversity, including variation in target conditions, AI-CDSS functions, implementation contexts, study designs, and outcome definitions [72]. Under these conditions, the assumption of a common underlying effect was not met, and quantitative pooling would risk producing misleading summary estimates. This approach is consistent with methodological guidance from the Cochrane Handbook [72].

We therefore performed a narrative synthesis and reported it in accordance with the Synthesis Without Meta-analysis (SWiM) reporting guidelines [71]. We categorized outcomes into 3 levels: patient level, referring to measures affecting patients, such as symptom improvement or function; clinician level, including outcomes related to provider knowledge, behavior, or workload; and system level, encompassing organizational or economic outcomes, such as health care utilization or cost-effectiveness. We synthesized findings

using the direction of effects (improvement, no effect, or worsening), which was determined based on the statistical results reported in the original studies (eg, *P* values or equivalent information). All included studies were considered, and findings were interpreted using study design and risk of bias. We presented findings using structured tables and narrative summaries. We provide the completed SWiM checklist in [Checklist 3](#).

We explored heterogeneity through structured comparison across implementation contexts. To describe the implementation context, we classified each study as simple, complicated, or complex using an operationalized approach informed by selected items from the NASSS-Complexity Assessment Tool (CAT) short version [73,74]. As NASSS-CAT was developed to support prospective assessment of technology projects and includes domains not consistently reported in published studies, we adapted relevant items into 4 observable criteria derived from the extracted data: condition complexity, decision logic, system complexity, and preference sensitivity. Each criterion was coded using predefined definitions ([Multimedia Appendix 3](#)). Two reviewers (FN and MS) independently classified context as *simple* when all criteria indicated low complexity and as *complex* when at least two criteria indicated high complexity. All other scenarios were classified as *complicated*.

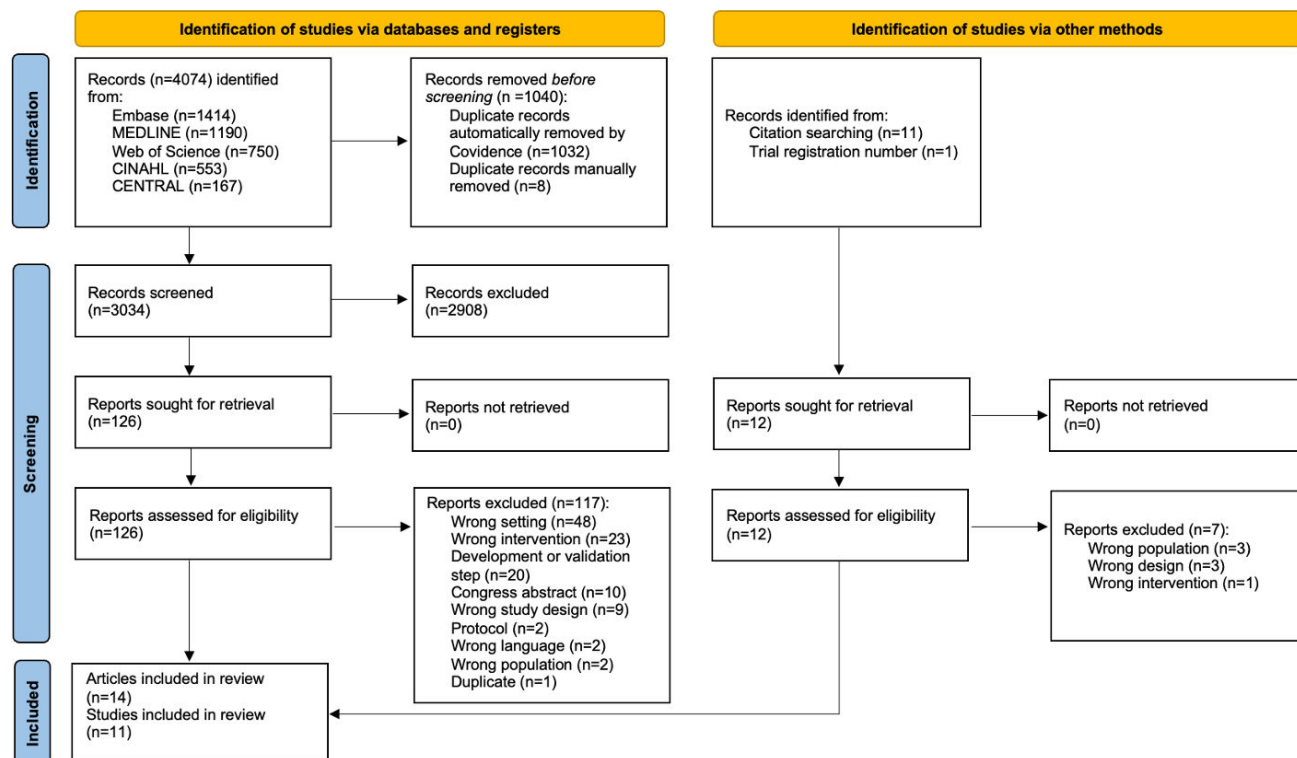
We classified safety information according to the DECIDE-AI reporting guidelines [65] across 3 domains: patient safety (prespecification and reported harms or unintended effects), AI system (risk mitigation and significant errors or malfunctions), and clinician interaction (agreement and override).

For AI-CDSS functions reporting, we used the taxonomy of front-end clinical decision support tools [75]. We then complemented this classification by extracting the corresponding operational functions, defined as the concrete actions performed by the system within the study context (eg, alert optimization, monitoring and follow-up, treatment optimization, and documentation or care summary).

## Results

### Study Selection

We retrieved 4074 records from the search. After removing duplicates, 3034 (74.5%) records remained for the selection. Of these records, 126 (4.2%) were retained for full-text screening. We identified 12 supplemental records from handsearching and trial registration checking. At the end of the process, we included 14 records on 11 distinct studies (we present the mapping of included records to unique studies in [Multimedia Appendix 4](#)). We present a flow diagram of the selection process in [Figure 1](#).

**Figure 1.** Flow diagram based on PRISMA 2020.

## Characteristics of the Studies

We present a summary of the characteristics of the included studies in [Table 2](#), and a detailed version is available in [Multimedia Appendix 5](#).

We identified 11 included studies published between 2019 and 2025 and conducted across North America ( $n=7$ , 63.6%) and Europe ( $n=4$ , 36.4%). Four (36.4%) studies used experimental designs, and 7 (63.6%) were

quasi-experimental, applying a variety of comparison strategies, including concurrent control groups, historical or matched controls, and before-after or self-controlled analyses. Study durations ranged from approximately 1 to 12 months. According to the DECIDE-AI reporting items, none of the studies reported patient involvement in the evaluation of the interventions. Five (45.5%) studies reported at least one conflict of interest, and 5 (45.5%) were funded by industry.

**Table 2.** Characteristics of included studies evaluating AI-based clinical decision support systems in primary care, including level of evidence, type of control group, study period, country of recruitment, and reporting of conflicts of interest and funding sources.

| Author                         | Year of publication | Level of evidence                                                  | Study period                         | Country of recruitment | Conflicts of interest | Sources of funding          |
|--------------------------------|---------------------|--------------------------------------------------------------------|--------------------------------------|------------------------|-----------------------|-----------------------------|
| Alcoceba-Herrero et al [76]    | 2024                | Experimental study                                                 | ~2 mo                                | Spain                  | No                    | Academic                    |
| Areias et al <sup>a</sup> [77] | 2024                | Quasi-experimental study with concurrent control group             | Control: ~24 mo; experimental: ~6 mo | United States          | Yes                   | Industrial and governmental |
| Chen et al [78]                | 2020                | Quasi-experimental study with alternating on and off periods       | ~2 mo                                | United States          | No                    | Not reported                |
| Cruz et al [79]                | 2019                | Quasi-experimental before-after study                              | Control: ~16 mo; experimental: ~1 mo | Spain                  | Not reported          | Not reported                |
| Granviken et al [68]           | 2025                | Experimental study                                                 | ~9 mo                                | Norway                 | No                    | Governmental                |
| Herter et al [80]              | 2022                | Quasi-experimental study with matched control group                | ~4 mo                                | Netherlands            | Yes                   | Industrial                  |
| Liu et al [81]                 | 2021                | Quasi-experimental study with historical control group             | ~8 mo                                | United States          | Yes                   | Industrial                  |
| Popescu et al [82]             | 2021                | Quasi-experimental study with self-control                         | ~10 mo                               | Canada                 | Yes                   | Industrial                  |
| Qassim et al [83]              | 2023                | Not reported                                                       | Not reported                         | Not reported           | Not reported          | Governmental                |
| Romero-Brufau et al [84]       | 2020                | Quasi-experimental pre-post survey-based quality improvement study | Control: ~1 mo; experimental: ~1 mo  | United States          | No                    | Academic                    |

| Author                          | Year of publication | Level of evidence  | Study period | Country of recruitment | Conflicts of interest | Sources of funding          |
|---------------------------------|---------------------|--------------------|--------------|------------------------|-----------------------|-----------------------------|
| Seol et al [85]                 | 2021                | Experimental study | ~12 mo       | United States          | No                    | Industrial and governmental |
| Yao et al [86]                  | 2021                | Experimental study | ~8 mo        | United States          | Yes                   | Academic                    |
| Rushlow et al <sup>a</sup> [87] | 2022                | Not reported       | Not reported | Not reported           | Not reported          | Not reported                |
| Thao et al <sup>a</sup> [88]    | 2024                | Not reported       | Not reported | Not reported           | Not reported          | Not reported                |

<sup>a</sup>Post hoc analysis.

Risk of Bias Assessment

We evaluated the overall risk of bias as high for all experimental studies. The main concern for the 2 parallel randomized controlled trials (RCTs) was related to deviations from the intended interventions. For the 2 cluster RCTs, the main concerns were linked to the identification and timing of recruitment and the measurement of the outcomes.

Among quasi-experimental studies, 1 was at serious risk of bias, while all others were at critical risk of bias. Five critical-level assessments were derived from the ROBINS-I screening stage (B1-B3), reflecting concerns related to uncontrolled confounding and/or outcome measurement. In the remaining study rated at critical risk of bias, concerns were related to missing data. We provide overall judgment and detailed domain-level assessments in [Multimedia Appendix 6](#).

Characteristics of the Participants

We present a summary of the characteristics of the participants included in the studies in [Table 3](#), and a detailed version is presented in [Multimedia Appendix 7](#).

AI-CDSSs were primarily evaluated in musculoskeletal conditions (3/11, 27.3%), urinary tract infections (2/11, 18.2%), and diabetes (2/11, 18.2%), with 1 (9.1%) study addressing multiple conditions. Of the 11 studies, 6 (54.5%) reported clinician sample sizes, ranging from 7 to 358. Among these 6 studies, 4 (66.7%) did not report any clinician characteristics. Age and sex were reported in 2 (33.3%) studies, and ethnicity in 1 (16.7%) study. Of the 11 studies, 8 (72.7%) reported patient sample sizes, ranging from 14 to 24,083 in the experimental groups and from 14 to 12,103 in the control groups. Among these 8 studies, age was reported in all (100%), sex and ethnicity in 5 (62.5%), and gender in 4 (50%). When reported (8/11, 72.7%), the number of participating institutions varied from 1 to 45 and included urban, rural, or mixed health care settings.

Table 3. Characteristics of study populations included in studies evaluating AI-based clinical decision support systems in primary care.

| Authors                     | Context                                                                                          | Clinicians: sample size and characteristics                                                    | Patients (experimental group): sample size and characteristics                                                                                                                              | Patients (control group): sample size and characteristics                                                                                                                                     | Number of participating institutions | Setting         |
|-----------------------------|--------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|-----------------|
| Alcoceba-Herrero et al [76] | Acute phase of COVID-19                                                                          | <ul style="list-style-type: none"><li>Clinician sample not described</li></ul>                 | <ul style="list-style-type: none"><li>n=170</li><li>Age:&lt;30 y, n=13, 12.6%; 30-45 y, n=39, 37.9%; 46-65 y, n=46, 44.7%; &gt;65 y, n=5, 4.9%</li><li>Gender: n=51, 49.5% women</li></ul>  | <ul style="list-style-type: none"><li>n=172</li><li>Age:&lt;30 y, n=28, 19.4%; 30-45 y, n=55, 38.2%; 46-65 y, n=57, 39.6%; &gt;65 y, n=4, 2.8%</li><li>Gender: n=90, 62.5% women</li></ul>    | 25                                   | Urban and rural |
| Areias et al [77]           | Chronic musculoskeletal pain                                                                     | <ul style="list-style-type: none"><li>Clinician sample not described</li></ul>                 | <ul style="list-style-type: none"><li>n=24,083</li><li>Age: mean 48.5 (SD 11.6) years</li><li>Gender: women n=14,126, 58.7%</li><li>Ethnicity: non-Hispanic White n=15,991, 67.9%</li></ul> | <ul style="list-style-type: none"><li>n=12,103</li><li>Age: mean 50.0 (SD 11.7) years</li><li>Gender: women n=6984, 57.9% women</li><li>Ethnicity: non-Hispanic White n=6013, 66.3%</li></ul> | Not reported                         | — <sup>a</sup>  |
| Chen et al [78]             | Shingles vaccination                                                                             | <ul style="list-style-type: none"><li>Clinician sample not described</li></ul>                 | <ul style="list-style-type: none"><li>Not reported</li></ul>                                                                                                                                | <ul style="list-style-type: none"><li>Not reported</li></ul>                                                                                                                                  | Not reported                         | Urban academic  |
| Cruz et al [79]             | Multiple conditions (hip or knee osteoarthritis, low back pain, headache, anxiety, osteoporosis, | <ul style="list-style-type: none"><li>n=86 primary care physicians and pediatricians</li></ul> | <ul style="list-style-type: none"><li>Not reported</li></ul>                                                                                                                                | <ul style="list-style-type: none"><li>Not reported</li></ul>                                                                                                                                  | 24                                   | Not reported    |

| Authors                  | Context                                                                                  | Clinicians: sample size and characteristics                                                                                                                                                                                                                   | Patients (experimental group): sample size and characteristics                                                                                                                                                                                                                                                                                                                  | Patients (control group): sample size and characteristics                                                                                                                                                      | Number of participating institutions | Setting                                                             |
|--------------------------|------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|---------------------------------------------------------------------|
| Granviken et al [68]     | dyspepsia, heart failure, heart murmur, urinary tract infection)<br>Musculoskeletal pain | <ul style="list-style-type: none"> <li>Experimental: n=22 physiotherapists</li> <li>Mean age 39.7 (SD 7.9) years</li> <li>Sex: 31.8% women</li> <li>Control: n=22 physiotherapists</li> <li>Mean age 40.8 (SD 8.8) years</li> <li>Sex: 27.3% women</li> </ul> | <ul style="list-style-type: none"> <li>n=358</li> <li>Age: mean 47.0 (SD 14.9) years</li> <li>Sex: women n=238, 66.5%</li> </ul>                                                                                                                                                                                                                                                | <ul style="list-style-type: none"> <li>n=366</li> <li>Age: mean 49.8 (SD 15.1) years</li> <li>Sex: women n=226, 61.7%</li> </ul>                                                                               | 33                                   | Private                                                             |
| Herter et al [80]        | Urinary tract infection                                                                  | <ul style="list-style-type: none"> <li>Clinician sample not described</li> </ul>                                                                                                                                                                              | <ul style="list-style-type: none"> <li>Preimplementation: n=4998</li> <li>Postimplementation: n=3422</li> <li>Age: &lt;30 y, 17%; 30-50 y, 21%; 50-70 y, 36%; &gt;70 y, 26%</li> <li>Sex: 88% women</li> <li>Postimplementation: n=1121 with certainty of AI-CDSS use</li> <li>Age: &lt;30 y, 15%; 30-50 y, 18%; 50-70 y, 39%; &gt;70 y, 28%</li> <li>Sex: 92% women</li> </ul> | <ul style="list-style-type: none"> <li>Preimplementation: n=5044</li> <li>Postimplementation: n=3360</li> <li>Age: &lt;30 y, 17%; 30-50 y, 20%; 50-70 y, 29%; &gt;70 y, 33%</li> <li>Sex: 86% women</li> </ul> | 36 experimental; 29 control          | Not reported                                                        |
| Liu et al [81]           | Type 1 or type 2 diabetes mellitus                                                       | <ul style="list-style-type: none"> <li>Clinician sample not described</li> </ul>                                                                                                                                                                              | <ul style="list-style-type: none"> <li>n=180</li> <li>Age: mean 57.4 (SD 11.4) years</li> <li>Gender: women n=97, 53.9%</li> <li>Ethnicity: Black n=141, 78.3%</li> </ul>                                                                                                                                                                                                       | <ul style="list-style-type: none"> <li>n=974</li> <li>Age: mean 56 (SD 13.8) years</li> <li>Gender: women n=565, 58%</li> <li>Ethnicity: Black n=746, 76.6%</li> </ul>                                         | 1                                    | Academic                                                            |
| Popescu et al [82]       | Major depressive disorder                                                                | <ul style="list-style-type: none"> <li>n=7 primary care physicians and psychiatrists</li> </ul>                                                                                                                                                               | <ul style="list-style-type: none"> <li>n=14 patients</li> </ul>                                                                                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>Self-control</li> </ul>                                                                                                                                                 | Not reported                         | —                                                                   |
| Qassim et al [83]        | Major depressive disorder                                                                | <ul style="list-style-type: none"> <li>Not reported</li> </ul>                                                                                                                                                                                                | <ul style="list-style-type: none"> <li>Mean age 36.4 (SD 14.84) years</li> <li>Sex: women n=9, 64.3%</li> <li>Ethnicity: Caucasian n=10, 76.9 %</li> </ul>                                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>Not reported</li> </ul>                                                                                                                                                 | Not reported                         | University hospitals, primary care clinics, and psychiatric clinics |
| Romero-Brufau et al [84] | Diabetes mellitus                                                                        | <ul style="list-style-type: none"> <li>Preimplementation: n=45 primary care physicians, advanced practice providers, nurses, clinical assistants, and others</li> <li>Postimplementation: n=38 primary care physicians, advanced</li> </ul>                   | <ul style="list-style-type: none"> <li>Not reported</li> </ul>                                                                                                                                                                                                                                                                                                                  | <ul style="list-style-type: none"> <li>Not reported</li> </ul>                                                                                                                                                 | 3                                    | Not reported                                                        |

| Authors            | Context               | Clinicians: sample size and characteristics                                                                                                                                                                                                                                                                                                                                                                            | Patients (experimental group): sample size and characteristics                                                                                                                                                        | Patients (control group): sample size and characteristics                                                                                                                                     | Number of participating institutions | Setting                        |
|--------------------|-----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|--------------------------------|
| Seol et al [85]    | Asthma                | <ul style="list-style-type: none"> <li>practice providers, nurses, clinical assistants, and others</li> <li>n=42 residents and other primary care providers</li> </ul>                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>n=90</li> <li>Age: median 9.3 (Q1<sup>b</sup>=4.8; Q3<sup>c</sup>=13.3) years</li> <li>Sex: female, n=37, 41%</li> <li>Ethnicity: White n=68, 76%</li> <li>n=11,573</li> </ul> | <ul style="list-style-type: none"> <li>n=94</li> <li>Age: median 8.3 (Q1=5.4; Q3=12.7) years</li> <li>Sex: female, n=43, 46%</li> <li>Ethnicity: White n=64, 68%</li> <li>n=11,068</li> </ul> | 1                                    | Not reported                   |
| Rushlow et al [87] | Low ejection fraction | —                                                                                                                                                                                                                                                                                                                                                                                                                      | —                                                                                                                                                                                                                     | —                                                                                                                                                                                             | 45                                   | —                              |
| Thao et al [88]    | Low ejection fraction | <ul style="list-style-type: none"> <li>n=181 physicians, physician assistants, and nurse practitioners</li> </ul>                                                                                                                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>Mean age 60.5 (SD 17.6) years</li> </ul>                                                                                                                                       | <ul style="list-style-type: none"> <li>Mean age 60.5 (SD 17.5) years</li> </ul>                                                                                                               | —                                    | Academic, community, and rural |
| Yao et al [86]     | Low ejection fraction | <ul style="list-style-type: none"> <li>Experimental: n=181 physicians, physician assistants, and nurse practitioners</li> <li>Age: mean 44.8 (SD 10.5) years</li> <li>Sex: 52.5% women</li> <li>Ethnicity: 94.4% White</li> <li>Control: n=177 physicians, physician assistants, and nurse practitioners</li> <li>Age: mean 46.1 (SD 11.01) years</li> <li>Sex: 59.7% women</li> <li>Ethnicity: 90.3% White</li> </ul> | —                                                                                                                                                                                                                     | —                                                                                                                                                                                             | —                                    | —                              |

<sup>a</sup>Not applicable.

<sup>b</sup>Q1: first quartile.

<sup>c</sup>Q3: third quartile.

## Characteristics of the Interventions

We describe the control and experimental interventions in [Table 4](#).

**Table 4.** Description of control and experimental interventions in studies evaluating AI-based clinical decision support systems in primary care.

| Authors                     | Control intervention arm                                                                                                                                                                                                                                                                                                                                  | Experimental intervention arm                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Alcoceba-Herrero et al [76] | <ul style="list-style-type: none"> <li>Clinical care: usual primary care follow-up for patients with COVID-19 in home isolation</li> <li>Delivery modality: scheduled telephone follow-up at 24 hours, 48 hours, 96 hours, 7 days, and 10 days; patient self-report of symptoms via a mobile app every 8 hours</li> <li>Decision support: none</li> </ul> | <ul style="list-style-type: none"> <li>Clinical care: same as control</li> <li>Delivery modality: same as control, with the addition of continuous remote monitoring of vital signs using wearable devices</li> <li>Decision support: automatic generation of alerts based on predefined thresholds; alerts triaged by severity (system-resolved, clinician, or emergency)</li> </ul> |
| Areias et al [77]           | <ul style="list-style-type: none"> <li>Clinical care: digital Care Program for chronic MSK<sup>a</sup> conditions</li> </ul>                                                                                                                                                                                                                              | <ul style="list-style-type: none"> <li>Clinical care: same as control</li> <li>Delivery modality: same as control</li> <li>Decision support: none</li> </ul>                                                                                                                                                                                                                          |

| Authors                  | Control intervention arm                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Experimental intervention arm                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Chen et al [78]          | <ul style="list-style-type: none"> <li>• Delivery modality: remote telerehabilitation platform; patient self-management with PT<sup>b</sup> supervision</li> <li>• Decision support: generation of recommendations for care adjustments; recommendation and rationale displayed to PT</li> <li>• Clinical care: usual primary care practice with standard shingles vaccination alerts</li> <li>• Delivery modality: alerts presented within the EHR<sup>c</sup> interface</li> <li>• Decision support: rule-based CDS<sup>d</sup> alert system prompting shingles vaccination; all eligible alerts displayed to clinicians</li> </ul> | <ul style="list-style-type: none"> <li>• Clinical care: same as control</li> <li>• Delivery modality: same EHR interface and alert structure as control</li> <li>• Decision support: prediction of the likelihood of clinician interaction with alerts; suppression of alerts predicted to be low-value based on a predefined threshold; alerts above the threshold displayed to clinicians</li> </ul>                                                                                                                                                                       |
| Cruz et al [79]          | <ul style="list-style-type: none"> <li>• Clinical care: usual primary care practice</li> <li>• Delivery modality: not reported</li> <li>• Decision support: none</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>• Clinical care: same as control</li> <li>• Delivery modality: real-time tool used during consultation</li> <li>• Decision support: processing of clinical notes from the EHR; generation of alerts with recommendations related to clinical pathways; alerts triggered when the patient case meets predefined criteria derived from clinical pathways translated into rules</li> </ul>                                                                                                                                               |
| Granviken et al [68]     | <ul style="list-style-type: none"> <li>• Clinical care: usual physiotherapy care</li> <li>• Delivery modality: in-person physiotherapy consultations</li> <li>• Decision support: none</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                     | <ul style="list-style-type: none"> <li>• Clinical care: same as control</li> <li>• Delivery modality: same as control, with the addition of a web-based CDSS<sup>e</sup> used during consultation</li> <li>• Decision support: identification of similar previous patients; generation of personalized treatment recommendations</li> </ul>                                                                                                                                                                                                                                  |
| Herter et al [80]        | <ul style="list-style-type: none"> <li>• Clinical care: usual primary care practice</li> <li>• Delivery modality: not reported</li> <li>• Decision support: none</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>• Clinical care: same as control</li> <li>• Delivery modality: web-based decision support software used by general practitioners</li> <li>• Decision support: patient characteristics entered manually into the system; estimation of expected outcomes for antibiotic treatment options based on similar patients; presentation of treatment options with predicted outcomes; display of guideline information alongside treatment options</li> </ul>                                                                                |
| Liu et al [81]           | <ul style="list-style-type: none"> <li>• Clinical care: usual primary care practice</li> <li>• Delivery modality: standard process (historical cohort from the same clinic)</li> <li>• Decision support: none</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                              | <ul style="list-style-type: none"> <li>• Clinical care: same as control</li> <li>• Delivery modality: fundus photography performed during the primary care visit</li> <li>• Decision support: use of an automated retinal image analysis system; automated classification of diabetic retinopathy severity; referral recommendations based on results</li> </ul>                                                                                                                                                                                                             |
| Popescu et al [82]       | <ul style="list-style-type: none"> <li>• Clinical care: usual primary care practice</li> <li>• Delivery modality: not reported</li> <li>• Decision support: none</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>• Clinical care: same as control</li> <li>• Delivery modality: use of a web-based clinical decision support system during clinical visits</li> <li>• Decision support: patients complete questionnaires through the app; symptom tracking and visualization; provision of guideline-based treatment information at each visit; generation of individualized probabilities of remission for antidepressant treatments</li> </ul>                                                                                                       |
| Qassim et al [83]        | <ul style="list-style-type: none"> <li>• Clinical care: usual primary care practice</li> <li>• Delivery modality: not reported</li> <li>• Decision support: none</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                           | <ul style="list-style-type: none"> <li>• Clinical care: same as control</li> <li>• Delivery modality: use of a web-based clinical decision support system during clinical visits</li> <li>• Decision support: patients complete questionnaires through the application; symptom tracking and visualization; provision of guideline-based treatment information at each visit; generation of individualized probabilities of remission for antidepressant treatments</li> </ul>                                                                                               |
| Romero-Brufau et al [84] | <ul style="list-style-type: none"> <li>• Clinical care: usual primary care practice</li> <li>• Delivery modality: standard workflow</li> <li>• Decision support: none</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>• Clinical care: same as control</li> <li>• Delivery modality: use of an AI-based clinical decision support system implemented in primary care clinics</li> <li>• Decision support: clinical assistants review lists of patients identified by the system before visits; patient risk profiles and recommendations are printed and provided to clinicians before the consultation; identification of patients at risk of poor glycemic control in the next 3 months; generation of recommendations intended to reduce risk</li> </ul> |

| Authors            | Control intervention arm                                                                                                                                                                                                                    | Experimental intervention arm                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Seol et al [85]    | <ul style="list-style-type: none"> <li>Clinical care: usual primary care practice with or without enrollment in an Asthma Management Program</li> <li>Delivery modality: not reported</li> <li>Decision support: none</li> </ul>            | <ul style="list-style-type: none"> <li>Clinical care: same as control</li> <li>Delivery modality: quarterly report provided to clinicians (every 3 mo)</li> <li>Decision support: prediction of 1-year risk of asthma exacerbation; provision of asthma management options to clinicians</li> </ul>                                                                                                                                                                                                                                                                                                                        |
| Rushlow et al [87] | <ul style="list-style-type: none"> <li>Clinical care: usual primary care practice</li> <li>Delivery modality: standard 12-lead ECG<sup>f</sup> performed and interpreted as part of routine care</li> <li>Decision support: none</li> </ul> | <ul style="list-style-type: none"> <li>Clinical care: same as control</li> <li>Delivery modality: use of an AI algorithm that was automatically applied to ECGs performed during routine care</li> <li>Decision support: clinicians received alerts when patients had a positive AI-ECG<sup>g</sup> result; ECGs were classified as positive or negative for low ejection fraction; clinicians were advised to consider ordering an echocardiogram for positive results; recommendations were provided within the report; weekly email reminders were sent when no action was taken following a positive result</li> </ul> |
| Thao et al [88]    | <ul style="list-style-type: none"> <li>Clinical care: usual primary care practice</li> <li>Delivery modality: standard 12-lead ECG performed and interpreted as part of routine care</li> <li>Decision support: none</li> </ul>             | <ul style="list-style-type: none"> <li>Clinical care: same as control</li> <li>Delivery modality: use of an AI algorithm that was automatically applied to ECGs performed during routine care</li> <li>Decision support: clinicians received alerts when patients had a positive AI-ECG result; ECGs were classified as positive or negative for low ejection fraction; clinicians were advised to consider ordering an echocardiogram for positive results; recommendations were provided within the report; weekly email reminders were sent when no action was taken following a positive result</li> </ul>             |
| Yao et al [86]     | <ul style="list-style-type: none"> <li>Clinical care: usual primary care practice</li> <li>Delivery modality: standard 12-lead ECG performed and interpreted as part of routine care</li> <li>Decision support: none</li> </ul>             | <ul style="list-style-type: none"> <li>Clinical care: same as control</li> <li>Delivery modality: use of an AI algorithm that was automatically applied to ECGs performed during routine care</li> <li>Decision support: clinicians received alerts when patients had a positive AI-ECG result; ECGs were classified as positive or negative for low ejection fraction; clinicians were advised to consider ordering an echocardiogram for positive results; recommendations were provided within the report; weekly email reminders were sent when no action was taken following a positive result</li> </ul>             |

<sup>a</sup>MSK: musculoskeletal.

<sup>b</sup>PT: physiotherapist.

<sup>c</sup>EHR: electronic health record.

<sup>d</sup>CDS: clinical decision support.

<sup>e</sup>CDSS: clinical decision support system.

<sup>f</sup>ECG: electrocardiogram.

<sup>g</sup>AI-ECG: AI-enhanced electrocardiogram.

Of the 11 studies, control interventions were mainly usual primary care practice (n=10, 90.9%). The remaining study had a digital care program as the control group. One study used a rule-based CDSS as the control intervention arm.

Experimental interventions generally preserved the same clinical care as the control while adding an AI-CDSS.

We present the identified AI-CDSS in Table 5.

**Table 5.** Overview of AI-based clinical decision support systems (AI-CDSSs) evaluated in primary care, including system names, intended purpose, and training or support provided for their use.

| Authors                     | Name of the AI-CDSS                           | Purpose                                                                                                             | Training or support to use AI-CDSS          |
|-----------------------------|-----------------------------------------------|---------------------------------------------------------------------------------------------------------------------|---------------------------------------------|
| Alcoceba-Herrero et al [76] | No name                                       | Support early detection of patient deterioration in COVID-19 home isolation                                         | Technical, logistical, and clinical support |
| Areias et al [77]           | No name                                       | Support monitoring and adjustment of digital care programs for individuals living with chronic musculoskeletal pain | Not reported                                |
| Chen et al [78]             | SmartCDS <sup>a</sup>                         | Reduce alert fatigue and increase the signal-to-noise ratio of shingles vaccination reminders                       | Not reported                                |
| Cruz et al [79]             | No name                                       | Support guideline adherence for multiple conditions                                                                 | Not reported                                |
| Granviken et al [68]        | SupportPrim PT <sup>b</sup> CDSS <sup>c</sup> | Support personalized decision-making for musculoskeletal pain management                                            | Training                                    |

| Authors                  | Name of the AI-CDSS            | Purpose                                                                                     | Training or support to use AI-CDSS               |
|--------------------------|--------------------------------|---------------------------------------------------------------------------------------------|--------------------------------------------------|
| Herter et al [80]        | No name                        | Support selection of the most effective antibiotic for urinary tract infections             | Training                                         |
| Liu et al [81]           | EyeArt v2.0 ARIAS <sup>d</sup> | Support early detection of diabetic retinopathy and improve adherence to follow-up eye care | No                                               |
| Popescu et al [82]       | Aifred                         | Support treatment selection and management for major depressive disorders                   | Training                                         |
| Qassim et al [83]        | Aifred                         | Support treatment selection and management for major depressive disorders                   | Training                                         |
| Romero-Brufau et al [84] | No name                        | Support glycemic control in patients with diabetes                                          | Not reported                                     |
| Seol et al [85]          | No name                        | Support asthma management in children                                                       | Not reported                                     |
| Rushlow et al [87]       | No name                        | Support early detection of low ejection fraction in adults without prior heart failure      | Website access with videos and FAQs <sup>e</sup> |
| Thao et al [88]          | No name                        | Support early detection of low ejection fraction in adults without prior heart failure      | Website access with videos and FAQs <sup>e</sup> |
| Yao et al [86]           | No name                        | Support early detection of low ejection fraction in adults without prior heart failure      | Website access with videos and FAQs <sup>e</sup> |

<sup>a</sup>CDS: clinical decision support.

<sup>b</sup>PT: physiotherapist.

<sup>c</sup>CDSS: clinical decision support system.

<sup>d</sup>ARIAS: Automated Retinal Image Analysis System.

<sup>e</sup>FAQ: frequently asked question.

Studies evaluated 11 distinct AI-CDSSs with different purposes. Seven (63.6%) AI-CDSSs aimed to support condition management, 3 (27.3%) aimed at the early detection of a condition or a deterioration, and 1 (9.1%) aimed to reduce alert fatigue. Preimplementation training occurred in

3 (27.3%) studies, and implementation support occurred in 2 (18.2%) studies.

We provide AI characteristics of the identified AI-CDSS in Table 6.

**Table 6.** AI characteristics of AI-based clinical decision support systems evaluated in primary care, including model accessibility, AI system type, year of model development, and interoperability with clinical systems.

| Authors                     | AI model accessibility            | AI system type                                    | AI development date | Interoperability                                                                                         |
|-----------------------------|-----------------------------------|---------------------------------------------------|---------------------|----------------------------------------------------------------------------------------------------------|
| Alcoceba-Herrero et al [76] | Not reported                      | Not reported                                      | Not reported        | Not reported                                                                                             |
| Areias et al [77]           | Commercial                        | Machine learning                                  | Not reported        | Not reported                                                                                             |
| Chen et al [78]             | Institutional but not open source | Machine learning                                  | Not reported        | EHR <sup>a</sup> embedded                                                                                |
| Cruz et al [79]             | Commercial                        | Natural language processing and rule-based system | Not reported        | Uses EHR data and integration within EHR interface not reported                                          |
| Granviken et al [68]        | Not reported                      | Case-based reasoning                              | Not reported        | Stand-alone system                                                                                       |
| Herter et al [80]           | Commercial                        | Machine learning                                  | Not reported        | Uses EHR data and integration within EHR interface not reported                                          |
| Liu et al [81]              | Commercial                        | Deep learning <sup>b</sup>                        | Not reported        | Stand-alone system                                                                                       |
| Popescu et al [82]          | Commercial                        | Deep learning                                     | Not reported        | Stand-alone system                                                                                       |
| Qassim et al [83]           | Commercial                        | Deep learning                                     | Not reported        | Stand-alone system                                                                                       |
| Romero-Brufau et al [84]    | Commercial                        | Not reported                                      | Not reported        | Not reported                                                                                             |
| Seol et al [85]             | Not reported                      | Machine learning                                  | Not reported        | Uses EHR data and outputs delivered via external reports (integration within EHR interface not reported) |
| Rushlow et al [87]          | Institutional but not open source | Deep learning                                     | Not reported        | EHR embedded                                                                                             |
| Thao et al [88]             | Institutional but not open source | Deep learning                                     | Not reported        | EHR embedded                                                                                             |
| Yao et al [86]              | Institutional but not open source | Deep learning                                     | Not reported        | EHR embedded                                                                                             |

<sup>a</sup>EHR: electronic health record.

<sup>b</sup>Assumed based on study description.

Six (54.5%) identified AI-CDSSs were commercial products, 2 (18.2%) were institutionally developed but not open source, and 3 (27.3%) did not report the accessibility of the underlying AI method. Nine (81.8%) studies reported the type of AI system; among these, 4 (44.4%) used machine learning, and 3 (33.3%) used deep learning approaches. No study reported when the AI model was developed. Regarding

interoperability, 4 (36.4%) studies evaluated stand-alone systems, 2 (18.2%) reported electronic health record (EHR)–embedded systems, and 3 (27.3%) used EHR data without specifying integration within the EHR interface. Three studies (27.3%) did not report interoperability information.

We present the CDSS characteristics in Table 7.

**Table 7.** Functional and operational characteristics of AI-based clinical decision support systems evaluated in primary care, including clinical decision support system (CDSS) functions, operational functions, mode of delivery, additional tasks required, and final decision-maker.

| Authors                     | CDSS functions                                                                                                                                 | Operational functions                                                                                                                                               | Mode of delivery                                                            | Additional tasks                                                                                                 | Final decision made by  |
|-----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|-------------------------|
| Alcoceba-Herrero et al [76] | <ul style="list-style-type: none"> <li>Point-of-care alerts or reminders</li> <li>Relevant information display</li> </ul>                      | <ul style="list-style-type: none"> <li>Alert optimization</li> <li>Documentation or care summary</li> <li>Monitoring or follow-up</li> </ul>                        | <ul style="list-style-type: none"> <li>Dashboard</li> <li>Alerts</li> </ul> | <ul style="list-style-type: none"> <li>Review dashboard</li> </ul>                                               | Clinicians <sup>a</sup> |
| Areias et al [77]           | <ul style="list-style-type: none"> <li>Expert systems</li> <li>Relevant information display</li> </ul>                                         | <ul style="list-style-type: none"> <li>Documentation/care summary</li> <li>Monitoring/follow-up</li> <li>Treatment optimization</li> </ul>                          | <ul style="list-style-type: none"> <li>Dashboard</li> </ul>                 | <ul style="list-style-type: none"> <li>Review and navigate dashboard</li> </ul>                                  | Clinicians              |
| Chen et al [78]             | <ul style="list-style-type: none"> <li>Point-of-care alerts/reminders</li> </ul>                                                               | <ul style="list-style-type: none"> <li>Alert optimization</li> </ul>                                                                                                | <ul style="list-style-type: none"> <li>Alerts</li> </ul>                    | <ul style="list-style-type: none"> <li>No</li> </ul>                                                             | N/A <sup>b</sup>        |
| Cruz et al [79]             | <ul style="list-style-type: none"> <li>Point-of-care alerts/reminders</li> </ul>                                                               | <ul style="list-style-type: none"> <li>Treatment optimization</li> </ul>                                                                                            | <ul style="list-style-type: none"> <li>Alerts</li> </ul>                    | <ul style="list-style-type: none"> <li>No</li> </ul>                                                             | Clinicians <sup>a</sup> |
| Granviken et al [68]        | <ul style="list-style-type: none"> <li>Expert systems</li> <li>Relevant information display</li> </ul>                                         | <ul style="list-style-type: none"> <li>Treatment optimization</li> </ul>                                                                                            | <ul style="list-style-type: none"> <li>Dashboard</li> </ul>                 | <ul style="list-style-type: none"> <li>Review and navigate dashboard</li> </ul>                                  | Clinicians <sup>a</sup> |
| Herter et al [80]           | <ul style="list-style-type: none"> <li>Expert systems</li> <li>Relevant information display</li> </ul>                                         | <ul style="list-style-type: none"> <li>Risk prediction</li> <li>Treatment optimization</li> </ul>                                                                   | <ul style="list-style-type: none"> <li>Dashboard</li> </ul>                 | <ul style="list-style-type: none"> <li>Enter patient's data</li> <li>Review bar chart</li> </ul>                 | Clinicians <sup>a</sup> |
| Liu et al [81]              | <ul style="list-style-type: none"> <li>Expert systems</li> </ul>                                                                               | <ul style="list-style-type: none"> <li>Referral</li> <li>Screening</li> </ul>                                                                                       | <ul style="list-style-type: none"> <li>No information</li> </ul>            | <ul style="list-style-type: none"> <li>Capture fundus images<sup>a</sup></li> </ul>                              | Clinicians <sup>a</sup> |
| Popescu et al [82]          | <ul style="list-style-type: none"> <li>Expert systems</li> </ul>                                                                               | <ul style="list-style-type: none"> <li>Documentation/care summary</li> </ul>                                                                                        | <ul style="list-style-type: none"> <li>Dashboard</li> </ul>                 | <ul style="list-style-type: none"> <li>Review and navigate dashboard</li> </ul>                                  | Clinicians              |
| Qassim et al [83]           | <ul style="list-style-type: none"> <li>Relevant information display</li> </ul>                                                                 | <ul style="list-style-type: none"> <li>Monitoring/follow-up</li> <li>Risk prediction</li> <li>Treatment optimization</li> </ul>                                     | <ul style="list-style-type: none"> <li>Dashboard</li> </ul>                 | <ul style="list-style-type: none"> <li>Review and navigate dashboard</li> </ul>                                  | Clinicians              |
| Romero-Brufau et al [84]    | <ul style="list-style-type: none"> <li>Expert systems</li> <li>Relevant information display</li> </ul>                                         | <ul style="list-style-type: none"> <li>Risk prediction</li> <li>Screening</li> <li>Treatment optimization</li> </ul>                                                | <ul style="list-style-type: none"> <li>Printed report</li> </ul>            | <ul style="list-style-type: none"> <li>Clinical assistant print the report</li> <li>Review the report</li> </ul> | Clinicians              |
| Seol et al [85]             | <ul style="list-style-type: none"> <li>Expert systems</li> <li>Relevant information display</li> </ul>                                         | <ul style="list-style-type: none"> <li>Documentation/care summary</li> <li>Monitoring/follow-up</li> <li>Risk prediction</li> <li>Treatment optimization</li> </ul> | <ul style="list-style-type: none"> <li>Quarterly report</li> </ul>          | <ul style="list-style-type: none"> <li>No</li> </ul>                                                             | Clinicians <sup>a</sup> |
| Rushlow et al [87]          | <ul style="list-style-type: none"> <li>Expert systems</li> <li>Point-of-care alerts/reminders</li> <li>Relevant information display</li> </ul> | <ul style="list-style-type: none"> <li>Referral</li> <li>Risk prediction</li> <li>Screening</li> </ul>                                                              | <ul style="list-style-type: none"> <li>Email alerts</li> </ul>              | <ul style="list-style-type: none"> <li>No</li> </ul>                                                             | Clinicians <sup>a</sup> |
| Thao et al [88]             | <ul style="list-style-type: none"> <li>Expert systems</li> <li>Point-of-care alerts/reminders</li> <li>Relevant information display</li> </ul> | <ul style="list-style-type: none"> <li>Referral</li> <li>Risk prediction</li> <li>Screening</li> </ul>                                                              | <ul style="list-style-type: none"> <li>Email alerts</li> </ul>              | <ul style="list-style-type: none"> <li>No</li> </ul>                                                             | Clinicians <sup>a</sup> |
| Yao et al [86]              | <ul style="list-style-type: none"> <li>Expert systems</li> <li>Point-of-care alerts/reminders</li> <li>Relevant information display</li> </ul> | <ul style="list-style-type: none"> <li>Referral</li> <li>Risk prediction</li> <li>Screening</li> </ul>                                                              | <ul style="list-style-type: none"> <li>Email alerts</li> </ul>              | <ul style="list-style-type: none"> <li>No</li> </ul>                                                             | Clinicians <sup>a</sup> |

<sup>a</sup>Assumed based on study description.<sup>b</sup>N/A: not applicable.

Expert system functions and relevant information display were each reported in 8 (72.7%) studies, and point-of-care alerts or reminders in 4 (36.4%). Eight (72.7%) of the 11 AI-CDSSs included multiple functions. These functions translated into operational capabilities, such as treatment optimization (7/11, 63.6%), risk prediction (5/11, 45.5%), monitoring or follow-up (4/11, 36.4%), documentation or care summary (4/11, 36.4%), screening (3/11, 27.3%), referral (2/11, 18.2%), and alert optimization (2/11, 18.2%). Modes of delivery included dashboards (5/11, 45.5%), alert-based systems (3/11, 27.3%), printed or periodic reports (2/11, 18.2%), and email notifications (1/11, 9.1%). One (9.1%) study did not report the mode of delivery. All systems aimed to support rather than replace clinician decision-making, as

explicitly stated in 3 (27.3%) studies and implicit in the remainder. In 7 (63.6%) studies, AI-CDSS use required additional clinician tasks, including access to and review of dashboards (5/11, 45.5%), manual data entry (1/11, 9.1%), acquisition of clinical images (1/11, 9.1%), printing of reports (1/11, 9.1%), and review of system-generated reports (1/11, 9.1%).

Using the classification approach adapted from the NASSS framework, included studies were assigned to simple (4/11, 36.4%), complicated (4/11, 36.4%), and complex (3/11, 27.2%) implementation contexts (see [Table 8](#) and [Multimedia Appendix 3](#) for rating details).

**Table 8.** Classification of implementation contexts for AI-based clinical decision support systems in primary care using the simple-complicated-complex typology informed by the NASSS<sup>a</sup> framework.

| Authors                     | Classification | Justification                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Alcoceba-Herrero et al [76] | Simple         | <ul style="list-style-type: none"> <li>Condition: clinically defined acute infectious condition (COVID-19) with relatively standardized monitoring targets</li> <li>Decision logic: protocolized (predefined thresholds for alerts and escalation)</li> <li>System: operationally simple (remote monitoring+app+automated alerts, but rule-based and predictable)</li> <li>Preference sensitivity: low (decisions driven by clinical deterioration thresholds rather than patient preferences)</li> </ul>                                        |
| Areias et al [77]           | Complex        | <ul style="list-style-type: none"> <li>Condition: clinically heterogeneous chronic MSK<sup>b</sup> conditions with variable presentations and trajectories</li> <li>Decision logic: individualized (continuous adjustment of care based on patient response and clinician input)</li> <li>System: multicomponent adaptive system (AI support+digital care program+PT<sup>c</sup> supervision+longitudinal monitoring)</li> <li>Preference sensitivity: high (treatment adaptation depends on patient goals, engagement, and response)</li> </ul> |
| Chen et al [78]             | Simple         | <ul style="list-style-type: none"> <li>Condition: narrow preventive care task (shingles vaccination eligibility)</li> <li>Decision logic: protocolized (rule-based alert suppression optimizing signal-to-noise)</li> <li>System: operationally simple (EHR<sup>d</sup>-integrated alert system)</li> <li>Preference sensitivity: low (clear guideline-based action)</li> </ul>                                                                                                                                                                  |
| Cruz et al [79]             | Complicated    | <ul style="list-style-type: none"> <li>Condition: multiple conditions with varying levels of clinical heterogeneity</li> <li>Decision logic: protocolized (guideline-based pathway adherence recommendations)</li> <li>System: integrated structured system (NLP<sup>e</sup> processes EHR data and generates real-time recommendations, but not fully embedded within the EHR workflow interface)</li> <li>Preference sensitivity: low to high depending on the conditions</li> </ul>                                                           |
| Granviken et al [68]        | Complex        | <ul style="list-style-type: none"> <li>Condition: clinically heterogeneous MSK pain with variable prognosis and response</li> <li>Decision logic: individualized (treatment decisions require integration of multiple patient factors)</li> <li>System: multicomponent adaptive system (case-based reasoning+clinician judgment+patient variability)</li> <li>Preference sensitivity: high (explicit emphasis on shared decision-making and person-centered care)</li> </ul>                                                                     |
| Herter et al [80]           | Complicated    | <ul style="list-style-type: none"> <li>Condition: moderately heterogeneous condition (UTI<sup>f</sup> presentations and risk profiles)</li> <li>Decision logic: structured (prediction-based guidance using decision trees within bounded decision space)</li> <li>System: integrated structured system (uses EHR data, but not described as fully embedded within the EHR workflow interface)</li> <li>Preference sensitivity: low to moderate (some trade-offs, but guideline constrained)</li> </ul>                                          |
| Liu et al [81]              | Simple         | <ul style="list-style-type: none"> <li>Condition: well-defined screening target (diabetic retinopathy)</li> <li>Decision logic: protocolized (binary classification → referral vs no referral)</li> <li>System: operationally simple (point-of-care screening+automated classification)</li> <li>Preference sensitivity: low (screening/referral decision is guideline driven)</li> </ul>                                                                                                                                                        |
| Popescu et al [82]          | Complex        | <ul style="list-style-type: none"> <li>Condition: clinically heterogeneous mental health condition</li> <li>Decision logic: structured+individualized (guideline-based algorithm combined with personalized AI predictions)</li> <li>System: multicomponent adaptive system (AI predictions+measurement-based care+clinician interpretation)</li> <li>Preference sensitivity: high (treatment selection depends on patient preferences, tolerance, and context)</li> </ul>                                                                       |
| Qassim et al [83]           | Complex        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

| Authors                  | Classification | Justification                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|--------------------------|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Romero-Brufau et al [84] | Complicated    | <ul style="list-style-type: none"> <li>Condition: clinically heterogeneous mental health condition</li> <li>Decision logic: structured+individualized (guideline-based algorithm combined with personalized AI predictions)</li> <li>System: multicomponent adaptive system (AI predictions+measurement-based care+clinician interpretation)</li> <li>Preference sensitivity: high (treatment selection depends on patient preferences, tolerance, and context)</li> </ul>    |
| Seol et al [85]          | Complicated    | <ul style="list-style-type: none"> <li>Condition: moderately heterogeneous chronic condition (diabetes with varying control levels)</li> <li>Decision logic: structured (risk stratification+recommended interventions)</li> <li>System: integrated structured system (AI identifies high-risk patients and suggests actions)</li> <li>Preference sensitivity: moderate (some contextual tailoring needed but within structured care pathways)</li> </ul>                     |
| Rushlow et al [87]       | Simple         | <ul style="list-style-type: none"> <li>Condition: clinically heterogeneous but stratified (childhood asthma with severity-based grouping)</li> <li>Decision logic: structured (risk prediction+periodic report guiding management decisions)</li> <li>System: integrated structured system (quarterly reports+ML<sup>g</sup> predictions within workflow)</li> <li>Preference sensitivity: low to moderate (decisions guided by risk but still clinician mediated)</li> </ul> |
| Thao et al [88]          | Simple         | <ul style="list-style-type: none"> <li>Condition: well-defined diagnostic target (low ejection fraction)</li> <li>Decision logic: protocolized (AI screening → echocardiography referral)</li> <li>System: operationally simple (screening tool embedded in routine ECG<sup>h</sup> workflow)</li> <li>Preference sensitivity: low (diagnostic confirmation pathway is standardized)</li> </ul>                                                                               |
| Yao et al [86]           | Simple         | <ul style="list-style-type: none"> <li>Condition: well-defined diagnostic target (low ejection fraction)</li> <li>Decision logic: protocolized (AI screening → echocardiography referral)</li> <li>System: operationally simple (screening tool embedded in routine ECG workflow)</li> <li>Preference sensitivity: low (diagnostic confirmation pathway is standardized)</li> </ul>                                                                                           |

<sup>a</sup>NASSS: Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability.

<sup>b</sup>MSK: musculoskeletal.

<sup>c</sup>PT: physiotherapist.

<sup>d</sup>EHR: electronic health record.

<sup>e</sup>NLP: natural language processing.

<sup>f</sup>UTI: urinary tract infection.

<sup>g</sup>ML: machine learning.

<sup>h</sup>ECG: electrocardiogram.

## Outcomes and Effects

### Overview

We present a summary of between-group effects in [Table 9](#) and provide a detailed table, including within-group results, in [Multimedia Appendix 8](#).

Among the 11 included studies, outcomes were most frequently reported at the clinician level (n=8, 72.7%), followed by the system level (n=6, 54.5%) and the patient level (n=5, 45.4%). Nearly half (n=5, 45.4%) of the studies assessed outcomes across all 3 levels.

**Table 9.** Between-group effects of AI-based clinical decision support systems on clinician-, patient-, and system-level outcomes in primary care, by study design and implementation context.

| Study                       | Implementation context | Clinician-level effects                                                                                                       | Patient-level effects                                                                                                                                                                                                                                                                                                                                                                                | System-level effects                                                                                                        |
|-----------------------------|------------------------|-------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Experimental studies        |                        |                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                             |
| Alcoceba-Herrero et al [76] | Simple                 | <ul style="list-style-type: none"> <li>During implementation period (~2 mo): reduction of alerts<sup>a</sup></li> </ul>       | — <sup>b</sup>                                                                                                                                                                                                                                                                                                                                                                                       | —                                                                                                                           |
| Granviken et al [68]        | Complex                | <ul style="list-style-type: none"> <li>During implementation period (~9 mo): more treatment modalities<sup>a</sup></li> </ul> | <ul style="list-style-type: none"> <li>At week 12: global perceived effect<sup>c</sup>; pain intensity<sup>c</sup>; pain self-efficacy<sup>c</sup>; emotional distress<sup>c</sup>; general MSK<sup>d</sup> health<sup>c</sup>; number of pain sites<sup>c</sup>; health-related quality of life<sup>c</sup>; sleep<sup>c</sup>; vitality<sup>c</sup>; use of pain medication in the last</li> </ul> | <ul style="list-style-type: none"> <li>During implementation period (~9 mo): number of consultations<sup>c</sup></li> </ul> |

| Study                                                   | Implementation context                                            | Clinician-level effects                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Patient-level effects                                                                                                                                                                                                                                                                                | System-level effects                                                                                                                                                                                                                                                                                                                                                                       |
|---------------------------------------------------------|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Seol et al [85]                                         | Complicated                                                       | <ul style="list-style-type: none"> <li>During implementation period (~12 mo): reduction of clinician burden<sup>a</sup></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                            | <ul style="list-style-type: none"> <li>At 12 mo: occurrence of asthma<sup>c</sup></li> <li>Assessed at 4 time points during the implementation period: asthma control status<sup>c</sup></li> <li>During implementation period (~8 mo): newly diagnosed low ejection fraction<sup>a</sup></li> </ul> | <ul style="list-style-type: none"> <li>During implementation period (~12 mo): health care costs<sup>c</sup>; timeliness of follow-up after exacerbation<sup>c</sup></li> <li>During implementation period (~8 mo): echocardiography utilization for high-risk patients<sup>a</sup>; cost-effectiveness<sup>a</sup>; echocardiography utilization for overall cohort<sup>c</sup></li> </ul> |
| Yao et al [86], Rushlow et al [87], and Thao et al [88] | Yao et al [86] and Rushlow et al [87]: —; Thao et al [88]: simple | —                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                            |
| Quasi-experimental studies                              |                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                            |
| Areias et al [77]                                       | Complex                                                           | —                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <ul style="list-style-type: none"> <li>At week 12: satisfaction with the program<sup>a</sup>; pain response rate<sup>c</sup>; mental health<sup>c</sup>; pain intensity<sup>e</sup></li> </ul>                                                                                                       | <ul style="list-style-type: none"> <li>During implementation period (~6 mo): completion rate of the program<sup>a</sup>; reach outs<sup>a</sup>; number of messages initiated by patients<sup>a</sup>; PT<sup>f</sup>-to-patient ratios<sup>a</sup>; safety<sup>c</sup>; sessions performed per week<sup>e</sup></li> </ul>                                                                |
| Chen et al [78]                                         | Simple                                                            | <ul style="list-style-type: none"> <li>During implementation period (~2 mo): order count<sup>c</sup>; interacted alert counts<sup>c</sup></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                          | —                                                                                                                                                                                                                                                                                                    | —                                                                                                                                                                                                                                                                                                                                                                                          |
| Cruz et al [79]                                         | Complicated                                                       | <ul style="list-style-type: none"> <li>During implementation period (~1 mo): adherence rate to 3 recommendations<sup>a</sup>; adherence rate to 5 recommendations<sup>c</sup></li> </ul>                                                                                                                                                                                                                                                                                                                                                      | —                                                                                                                                                                                                                                                                                                    | —                                                                                                                                                                                                                                                                                                                                                                                          |
| Herter et al [80]                                       | Complicated                                                       | <ul style="list-style-type: none"> <li>During implementation period (~4 mo): prescription behavior<sup>a,c,e</sup></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>During implementation period (~4 mo): successful treatment<sup>a</sup></li> </ul>                                                                                                                                                                             | —                                                                                                                                                                                                                                                                                                                                                                                          |
| Liu et al [81]                                          | Simple                                                            | —                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | —                                                                                                                                                                                                                                                                                                    | <ul style="list-style-type: none"> <li>During implementation period (~8 mo): rate of attendance at ophthalmology appointments<sup>a</sup></li> </ul>                                                                                                                                                                                                                                       |
| Popescu et al [82] and Qassim et al [83]                | Popescu et al [82]: —; Qassim et al [83]: complex                 | <ul style="list-style-type: none"> <li>During implementation period (~10 mo): appointment length<sup>c</sup></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                       | —                                                                                                                                                                                                                                                                                                    | —                                                                                                                                                                                                                                                                                                                                                                                          |
| Romero-Brufau et al [84]                                | Complicated                                                       | <ul style="list-style-type: none"> <li>At the end of implementation period (~1 mo): Clinician attitudes: care coordination<sup>a</sup>; patient preparation for managing diabetes<sup>a</sup>; familiarity with AI<sup>a</sup>; reducing complications<sup>c</sup>; intervention effectiveness<sup>c</sup>; right intervention<sup>c</sup>; routinely use AI<sup>c</sup>; AI will complicate my job<sup>c</sup>; AI will make my job obsolete<sup>c</sup>; excited about AI<sup>e</sup>; AI does not understand my job<sup>e</sup></li> </ul> | —                                                                                                                                                                                                                                                                                                    | —                                                                                                                                                                                                                                                                                                                                                                                          |

<sup>a</sup>Favors the experimental group.

<sup>b</sup>Not applicable.

<sup>c</sup>No difference.

<sup>d</sup>MSK: musculoskeletal.

<sup>e</sup>Favors the control group.

<sup>f</sup>PT: physiotherapist.

## Clinician-Level Outcomes

Four (50%) studies reported workflow outcomes assessed during the implementation period, resulting in mixed effects.

Two experimental studies reported reductions in alert burden and clinician workload across implementation periods ranging from short (~2 mo) to longer periods (~12 mo). These studies

were conducted using protocolized or structured decision logic. Two quasi-experimental studies reported no effect on interacted alert counts in a simple context with an automated EHR-embedded AI-CDSS for alert optimization, and no effect on appointment length in a complex context characterized by individualized decision logic, a multicomponent system, and high preference sensitivity.

Four (50%) studies reported decision behavior outcomes during the implementation period. One experimental and 3 quasi-experimental studies examined prescribing behavior, guideline adherence, or order count with heterogeneous effects across and within studies. These studies were conducted across different implementation contexts. In the simple context, alert suppression was not associated with changes in the order count outcome. Two studies conducted in complicated contexts reported mixed effects on prescribing behavior and positive or neutral effects on adherence to clinical guidelines. The experimental study conducted in a complex context reported increased use of treatment modalities. These complicated and complex contexts were characterized by clinical heterogeneity, structured or individualized decision logic, and moderate to high preference sensitivity.

One quasi-experimental study ( $n=1$ , 12.5%) context evaluated attitudinal outcomes at the end of the short implementation period (~1 mo) in a complicated context, reporting improvements in care coordination and patient preparation, alongside concerns regarding the usefulness of the system and its impact on clinical work.

## Patient-Level Outcomes

Two (40%) studies reported patient-reported outcomes assessed at 12 weeks in musculoskeletal conditions. Pain intensity and mental health were the only outcomes assessed in more than 1 study. Across these studies, findings for pain intensity were inconsistent, with 1 experimental study reporting no effect and 1 quasi-experimental study reporting a negative effect. Both studies reported no effect on mental health outcomes. These studies were conducted in complex contexts characterized by clinical heterogeneity, individualized decision logic, multicomponent systems, and high preference sensitivity.

Two (40%) studies reported clinical outcomes assessed during or after the study period. One experimental study reported no effect on asthma occurrence and control, while 1 quasi-experimental study observed a positive effect on treatment success. Both studies were conducted in complicated contexts characterized by clinically heterogeneous but stratified conditions, protocolized decision logic, and moderate preference sensitivity, and differed in system complexity.

Two (40%) studies reported process-of-care outcomes with mixed effects. One experimental study reported a positive effect on the detection rate of low ejection fraction in a simple implementation context characterized by a well-defined diagnostic target, protocolized decision logic, an operationally simple system, and low preference sensitivity.

An experimental study reported no effect on medication use in a complex context involving clinical heterogeneity, individualized decision logic, a multicomponent system, and high preference sensitivity.

One quasi-experimental study (1/5, 20%) reported a patient-reported experience outcome (satisfaction with the program) at 12 weeks with a positive effect.

## System-Level Outcomes

Three (50%) studies evaluated service delivery outcomes during the implementation period and reported mixed results. One quasi-experimental study reported improvements in program completion and patient engagement (eg, reach-outs and messages), while another quasi-experimental study found increased attendance at follow-up appointments. These effects were observed across different implementation contexts, including both complex and simple contexts. Two studies reported neutral or negative effects on service delivery indicators, such as the number of consultations (experimental study) and sessions performed per week (quasi-experimental study), both conducted in complex contexts.

Two (33.3%) experimental studies evaluated economic outcomes with mixed effects. One study reported that the intervention was cost-effective, while the other found no effect on health care costs. The 2 studies were conducted in different implementation contexts, including 1 simple context involving a well-defined diagnostic target with low preference sensitivity and 1 complicated context with a heterogeneous condition and moderate preference sensitivity.

Two (33.3%) studies reported utilization outcomes with positive effects. An experimental study reported an increased use of diagnostic procedures among high-risk patients but no clinical effect difference in the intervention sample. This result was observed in a simple context characterized by a well-defined diagnostic target and a protocolized decision logic. One quasi-experimental study reported improved physiotherapist-to-patient ratios in a complex context.

One (16.7%) quasi-experimental study reported no impact on safety during the implementation period in a complex context.

## Safety Evaluation

We present the reporting of safety-related information in [Table 10](#). Two (18.2%) studies prespecified structured monitoring of adverse events, primarily relying on patient-reported side effects or standardized questionnaires rather than active surveillance of AI-related harms. Across these 2 studies, as well as 1 additional study reporting post hoc harms and unintended effects, no severe adverse events attributable to the AI-CDSS were identified. Risk mitigation strategies were described in 4 (36.4%) studies and included human-in-the-loop oversight, internal quality checks, and automated safeguards such as validation of severe alerts or continuous retraining with monitoring dashboards. Reports of errors or malfunctions were uncommon ( $n=3$ , 27.3%) and mainly incidental, involving image quality issues, software glitches, or technical exclusions of participants. No study

reported clinician agreement with or override of AI-generated recommendations.

**Table 10.** Reporting of safety-related information for AI-based clinical decision support systems in primary care at patient level, system level, and clinician level.

| Authors                     | Patient safety: prespecification                                                                                                      | Patient safety: harms and unintended effects                                                | AI system: risk mitigation                                                                                                                                         | AI system: significant errors and malfunctions                                | Clinician interaction: agreement and override |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|-----------------------------------------------|
| Alcoceba-Herrero et al [76] | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Patient validation required before escalation of severe alerts to emergency</li> </ul>                                    | Not reported                                                                  | Not reported                                  |
| Areias et al [77]           | Report of any adverse events                                                                                                          | Intervention-related adverse events were rare and were evenly distributed across the groups | <ul style="list-style-type: none"> <li>• Human-in-the-loop</li> <li>• Routine internal quality checks comparing algorithms versus clinicians' decisions</li> </ul> | Not reported                                                                  | Not reported                                  |
| Chen et al [78]             | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Monitoring dashboard for alert outputs</li> <li>• Model retrained daily</li> </ul>                                        | Not reported                                                                  | Not reported                                  |
| Cruz et al [79]             | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Not reported                                                                  | Not reported                                  |
| Granviken et al [68]        | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Two participants in the intervention group excluded due to technical issues   | Not reported                                  |
| Herter et al [80]           | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Not reported                                                                  | Not reported                                  |
| Liu et al [81]              | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Errors related to image quality reported (eg, ungradable/poor-quality images) | Not reported                                  |
| Popescu et al [82]          | Weekly collection of PRISE-20 <sup>a</sup> and FIBSER <sup>b</sup> to assess side effects and UKU <sup>c</sup> scale after each visit | No serious adverse events related to the tool                                               | <ul style="list-style-type: none"> <li>• Software errors identified during the trial were corrected</li> </ul>                                                     | Incorrect questionnaire score comparison in CANMAT <sup>d</sup> algorithm     | Not reported                                  |
| Qassim et al [83]           | Weekly collection of PRISE-20 and FIBSER to assess side effects and UKU scale after each visit                                        | No serious adverse events related to the tool                                               | <ul style="list-style-type: none"> <li>• Software errors identified during the trial were corrected</li> </ul>                                                     | Broken link between pages in the CANMAT algorithm                             | Not reported                                  |
| Romero-Brufau et al [84]    | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Not reported                                                                  | Not reported                                  |
| Seol et al [85]             | Not reported                                                                                                                          | 0 serious adverse event                                                                     | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Not reported                                                                  | Not reported                                  |
| Yao et al [86]              | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Not reported                                                                  | Not reported                                  |
| Rushlow et al [87]          | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Not reported                                                                  | Not reported                                  |
| Thao et al [88]             | Not reported                                                                                                                          | Not reported                                                                                | <ul style="list-style-type: none"> <li>• Not reported</li> </ul>                                                                                                   | Not reported                                                                  | Not reported                                  |

<sup>a</sup>PRISE-20: Patient-Rated Inventory of Side Effects-20 items.

<sup>b</sup>FIBSER: Frequency, Intensity, and Burden of Side Effects Rating scale.

<sup>c</sup>UKU: Udvalg for Kliniske Undersøgelser.

<sup>d</sup>CANMAT: Canadian Network for Mood and Anxiety Treatments.

## Discussion

### Principal Findings

This systematic review synthesized evidence on the effectiveness and safety of AI-CDSSs in primary care across 11 studies. Two RCTs, conducted in a simple implementation context with screening or alert optimization AI-CDSSs, reported positive effects on workflow, economic, and utilization outcomes. In contrast, studies conducted in more complex contexts reported limited or inconsistent

effects, with patient-level outcomes rarely improved. Across studies, most systems were deployed as stand-alone tools requiring additional clinician effort, and follow-up periods were generally short, reflecting early phases of implementation. Safety was not evaluated as an integrated, multilevel construct. Only 2 studies prespecified patient safety outcomes. System-level safeguards were described but not evaluated, and no study examined how clinicians engaged with AI-generated recommendations. The current evidence base does not clarify how system performance, clinician behavior, and patient outcomes interact to shape safety in

practice. The overall strength of evidence is further limited by substantial heterogeneity and consistently high or critical risk of bias, restricting confidence in the observed effects and their applicability across clinical contexts.

## Effectiveness Considerations

Positive effects of AI-CDSS implementation were reported more consistently when clinical decisions followed structured and protocolized action pathways. In our review, the most consistent effects emerged in studies targeting screening, referral, or alert optimization, where clinical rules were explicit and thresholds were predefined. In these contexts, recommendations could be readily integrated into routine workflows and translated into immediate clinical responses. Prior CDSS research similarly suggests that decision support is more likely to improve care when recommendations are delivered at the point of care, embedded within clinical workflows, and linked to clear actions [89-91]. However, structured decision contexts alone may be insufficient to ensure effectiveness, as systems that increase workflow burden, generate excessive interruptions, or poorly align with local practices may reduce clinician responsiveness over time [92-96]. Alert fatigue represents a well-documented failure mode, with studies reporting override rates exceeding 50% to 73% for medication-related alerts, with clinicians becoming progressively less likely to accept alerts as their frequency increases [94,97,98]. This suggests that the effects of AI-CDSSs in structured contexts may depend not only on the clarity of the underlying clinical rules but also on the degree to which system recommendations remain actionable, relevant, and aligned with the realities of clinical practice.

In settings requiring complex or individualized decision-making, AI-CDSS effects on clinical outcomes appear less consistent, potentially not because the systems fail, but because they are used differently in ways that leave a limited measurable trace. Clinicians integrating evolving clinical information, patient goals, and preference-sensitive trade-offs across longitudinal care trajectories face decisions that are less amenable to reduction into a single rule, threshold, or predefined action pathway [99-101]. Several mechanisms may contribute to this pattern. First, as decision tasks become more cognitively demanding, clinicians appear less likely to consult decision support systems and may shift toward situation-assessment tools rather than algorithmic recommendations [102]. Second, physicians tend to remain anchored to pre-existing diagnoses when using diagnostic CDSSs. Berner et al [103] found that system outputs more often reinforced than redirected clinical reasoning, a pattern consistent with confirmation and anchoring biases previously described in clinical reasoning research, particularly under cognitively demanding conditions [104,105]. Third, even when systems are consulted, clinicians may resist outputs perceived as undermining their decision autonomy or conflicting with the clinical context, particularly when system reasoning is opaque or inconsistent [106,107]. These mechanisms may shift AI-CDSSs from a directive toward a more interpretive or confirmatory role. In our review, a study targeting chronic musculoskeletal pain found that clinicians used AI-CDSSs primarily to support existing practice rather than to modify

management decisions [68]. This pattern is consistent with broader evidence showing that CDSS effects on clinical outcomes in chronic disease are often modest and heterogeneous even when improvements in process measures are observed [108,109]. If AI-CDSS outputs primarily validate or refine reasoning rather than consistently triggering management changes, measurable downstream effects on clinical outcomes may be expected to remain small, variable, and difficult to detect. Decision complexity should be explicitly addressed during AI-CDSS design, implementation, and evaluation [29]. Future research may also need evaluation approaches that better capture interpretive and confirmatory influences on clinical reasoning rather than relying exclusively on downstream outcome measures.

Implementation success also depends on alignment with organizational context, clinical workflows, and clinician routines [7,53,110]. Many AI-CDSS failures appear to be sociotechnical rather than algorithmic [19,29,111-113]. In our review, most AI-CDSSs were implemented as stand-alone systems requiring additional clinician effort, including manual data entry, dashboard consultation, or review of system-generated outputs outside existing workflows. In primary care, increased workload consistently emerges as a major barrier to CDSS adoption, with human and organizational factors negatively influencing uptake regardless of technological performance [7]. Workflow integration, usability, and health care system-level factors have also been identified as persistent implementation challenges across clinical domains and CDSS types [53,110], reinforcing the importance of aligning decision support systems with routine clinical work [29,32]. These barriers may become particularly consequential in more complex care settings, where workflows are often less standardized. In these contexts, clinicians must continuously integrate evolving clinical and contextual information across encounters, providers, and competing patient priorities, as in chronic disease management, multimorbidity, or preference-sensitive care [52]. EHR integration is frequently proposed as a strategy to reduce these implementation barriers [111,113], and limited evidence from our review is consistent with this interpretation, as the 2 interventions described as EHR-integrated both reported positive outcomes. However, the small number of studies prevents causal interpretation, and integration alone may be insufficient to ensure effectiveness. A CDSS deployment study reported implementation failure despite high algorithmic performance because only a minority of required input variables were available at the point of care, limiting the system's practical usability during clinical decision-making [114]. EHR integration may also constrain decision support functionality or require substantial local infrastructure investment, particularly in fragmented primary care systems [112,113].

Interpretation of these findings should also consider systematic differences in outcome selection and evaluation timing across implementation contexts. In our review, AI-CDSSs implemented in structured contexts primarily evaluated process-level outcomes, whereas patient-level clinical outcomes were less frequently assessed, a pattern

consistent with prior reviews [115,116]. Process outcomes are more readily measurable within short evaluation windows and are often more immediately responsive to decision support interventions [89,116]. However, improvements in process-level outcomes do not necessarily translate into patient-relevant benefit [89,116,117]. Reliance on intermediate end points has been associated with overestimated effects compared with trials evaluating patient outcomes directly [118]. In individualized contexts, where outcomes are shaped by interacting clinical and contextual factors across longitudinal care trajectories, isolating change specifically to AI-CDSSs becomes methodologically challenging [119]. Short evaluation timing was common in our review and may have captured early implementation phases before stable adoption patterns were established. Clinician engagement with decision support tends to evolve over time as familiarity, workflow demands, and implementation barriers change [31,120,121]. These interpretive challenges may partly reflect a mismatch between the predominantly point-in-time logic of current AI-CDSSs and the longitudinal structure of primary care. Most systems in our review were designed to support discrete decisions. This approach is well suited to structured tasks but less aligned with care that unfolds across encounters, depends on evolving patient goals, and requires ongoing adaptation to changing clinical circumstances [122,123]. Achieving more consistent effects in complex primary care settings may require forms of decision support evolving along with care trajectories, clinician judgment, and shared decision-making. Evaluation strategies may also need longer follow-up periods and clearer prespecified linkage between process-level changes and patient-level outcomes.

## Safety Considerations

Current reporting of AI-CDSS implementation research captures safety through fragmented and inconsistently specified indicators. Only 2 studies in this review prespecified safety outcomes, and most relied on the absence of serious adverse events as a proxy for safety, with limited detail regarding how harms were identified, monitored, or attributed. This approach risks missing harms that are subtle, cumulative, or delayed, such as missed follow-up, delayed diagnosis, or inappropriate treatment, because these events often require prospective monitoring rather than retrospective absence of signals [124]. In primary care, these risks are amplified by the longitudinal and distributed nature of care [125]. Clinical presentations are heterogeneous, care unfolds across repeated encounters, and adverse outcomes often cannot be attributed to a single decision or clinician interaction [126,127]. Diagnostic delays in ambulatory care, for example, frequently emerge from failures in follow-up coordination and clinical sensemaking across complex work systems rather than from isolated technical errors [128]. Similar concerns have been raised across the broader AI-CDSS literature, where safety reporting often remains secondary to performance evaluation, and benefits and harms are frequently assessed separately, limiting interpretation of overall clinical impact [34,129,130]. Recent expert consensus statements have also highlighted the absence of structured adverse event reporting methodologies and postdeployment

monitoring strategies for AI systems in clinical care [38], while calls for systematic “algorithmovigilance” remain only partially operationalized in practice [37].

Our findings further suggest that AI-CDSS safety in primary care cannot be adequately understood through isolated technical or patient-level indicators. Safety events involving digital systems characteristically emerge from interactions among technical performance, clinician behavior, workflow conditions, and organizational context [93,126,131,132]. These dimensions of safety were largely unreported in the studies included in our review. This gap may be particularly important in primary care, where high cognitive load and cumulative alert exposure increase the likelihood of automation bias and alert fatigue [94,133,134]. Harm in primary care can also often accumulate longitudinally across encounters rather than concentrating in discrete events detectable through standard adverse event monitoring [125,128]. Without structured and adapted safety evaluation, implementation studies risk producing evidence for AI-CDSS adoption without establishing whether these systems are, in practice, safe to use. These features of primary care point to evaluation requirements that are not generic but structurally specific to the setting. These include prospective specification of safety outcomes inclusive of process-level harms such as missed follow-up and delayed referral [38,65]; longitudinal monitoring during routine use rather than at deployment alone [127,135]; systematic reporting of clinician engagement with AI recommendations, including override patterns [94,133]; and monitoring intensity calibrated to the clinical consequence of supported decisions, which in primary care spans chronic disease management to acute triage [65,136]. Without evaluation approaches adapted to these features of primary care, the growing body of AI-CDSS implementation evidence risks establishing effectiveness without establishing safety, reversing the expected evidence sequence in clinical research [55].

This review has subject-matter and methodological limitations. First, our search strategy was developed in collaboration with an information specialist (FB) and tested for sensitivity. However, we identified several studies through citation searching and trial registration number. This reflects the indexing inconsistencies and incomplete database coverage for CDSS, despite the existence of a MeSH term since 1997. Our methodology, including backward and forward citation screening in identified knowledge syntheses records, reduces the risk of missing studies that would affect the overall result interpretation. As AI-CDSS implementation studies continue to emerge, the evidence synthesized in this review reflects the literature available up to the predefined search dates (June 27, 2025). Future work should periodically repeat this search approach to monitor newly published evaluations and incorporate emerging evidence over time. Second, substantial heterogeneity across interventions, study designs, implementation contexts, and reported outcomes precluded meta-analysis and limited the use of summary effect estimates for strength-of-evidence interpretation. We therefore used synthesis without meta-analysis based on direction of effect. As emphasized in

the SWiM guidance [71], this synthesis approach answers whether there is evidence of effects across studies, rather than estimating an average intervention effect. Our findings should be interpreted as patterns of reported effects across heterogeneous AI-CDSS implementations, not as a pooled estimate of AI-CDSS effectiveness in primary care. Although this approach was appropriate given the available evidence, it carries a risk of interpretive bias. We mitigated this risk through structured outcome grouping, predefined implementation context criteria, and review by a multidisciplinary team of researchers, clinicians, and people with lived experience. Beyond these methodological considerations, interpretation of the findings should also consider the well-recognized international variation in the organization and delivery of primary care across health systems [1,137]. Although all included studies met our predefined eligibility criteria for primary care, they involved different clinician groups and care settings across countries. AI-CDSSs evaluated in 1 primary care system may not perform similarly when implemented in health systems with different organizational structures or models of care, which may limit the transferability of these findings. Third, limitations related to the included studies constrain interpretation. The included studies were conducted exclusively in high-income countries, which limits the generalizability of these findings to low- and middle-income settings. The restriction to English and French publications may have limited the identification of relevant studies from other regions. This choice was made to ensure accurate data extraction and interpretation, as studies in other languages could not be reliably assessed by the review team. The classification of implementation contexts (simple, complicated, and complex) involved an element of interpretive judgment. To enhance transparency and reproducibility, we applied predefined criteria informed by the NASSS framework, including condition complexity, decision logic, system complexity, and preference sensitivity. Classification was performed using explicit decision rules and reviewer consensus. However, this approach remains dependent on the level of detail reported in primary studies, which was often

limited, and may therefore not fully capture the underlying complexity of implementation contexts. Finally, evidence from the broader CDSS literature suggests that effectiveness may vary over time, with more recent systems showing modest improvements in clinician behavior [138]. However, we were unable to extract information on the development timeline, versioning, or updating procedures of AI models, as these data were not reported in the included studies. This limits the ability to determine whether observed effects reflect implementation context or differences related to the maturity of the systems evaluated.

## Conclusions

Available evidence on AI-CDSSs implemented in primary care demonstrates limited and inconsistent effectiveness, with safety poorly assessed. While positive effects were observed in simple clinical implementation contexts, these were primarily limited to clinician- and system-level outcomes. Improvements in outcomes were not consistent across more complex contexts and rarely translated into patient-level benefits. Safety conclusions are limited by inconsistent evaluation at the AI system level, patient level, and clinician level. The effectiveness conclusions are constrained by heterogeneity, short follow-up, and high or critical risk of bias. By synthesizing comparative evidence from clinically implemented AI-CDSSs across clinician-, patient-, and system-level outcomes and explicitly considering safety and implementation context, this review extends prior reviews that primarily mapped AI applications or implementation barriers in primary care. Clinical implementation research should evaluate safety as a multilevel construct and describe how implementation context shapes the use and impact of AI-CDSSs in clinical practice. Health care systems are adopting AI-CDSSs faster than they can ensure reliable and safe use, a gap demanding rigorous workflow integration, standardized evaluation frameworks, and systematic safety assessment to support trustworthy implementation in primary care.

## Acknowledgments

Generative AI tools were used during manuscript preparation but not during the conduct of the systematic review. ChatGPT (OpenAI; GPT-5.5) was used to assist with language editing, improving clarity and conciseness, formatting, and revising responses to reviewers. Elicit (Ought) was used during manuscript preparation as a supplementary tool to identify potentially contradictory or alternative published evidence when drafting the "Discussion" section and to challenge the authors' interpretation of the review findings. This process was intended to encourage consideration of evidence beyond the authors' initial interpretation. Any AI-generated text suggestions and any publications identified through Elicit were independently reviewed and verified against the original sources before being incorporated into the manuscript. Generative AI was not used to formulate the review question, develop the methodology, establish eligibility criteria, select studies, extract or analyze data, assess risk of bias, synthesize evidence, or determine the conclusions. The authors critically reviewed all content and take full responsibility for the integrity and accuracy of the manuscript.

## Funding

This study was funded by the Centre for Digital Health Evaluation, which was mandated by a governmental institution to commission the review. Two co-authors are affiliated with the Centre for Digital Health Evaluation and contributed to the study by helping align the review methods with the commissioning objectives and by critically revising the manuscript. The review team independently conducted study selection, data extraction, risk of bias assessment, data synthesis, interpretation of the findings, and made the final decision to submit the manuscript for publication.

---

**Data Availability**

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

---

**Authors' Contributions**

Conceptualization: FN, MS, OB, VS.

Methodology: FB, FN, MS.

Formal analysis: FN, MS, SO.

Investigation: FN, MS, SO.

Data curation: FN, MS, SO.

Funding acquisition: FN, MS.

Project administration: FN, MS.

Visualization: FN, SO.

Writing – original draft: FN.

Writing – review & editing: FN, FY, MS, OB, SO, VS.

---

**Conflicts of Interest**

None declared.

---

**Multimedia Appendix 1**

Methodological specifications.

[\[DOCX File \(Microsoft Word File\), 23 KB-Multimedia Appendix 1\]](#)

---

**Multimedia Appendix 2**

Databases search strategy.

[\[DOCX File \(Microsoft Word File\), 22 KB-Multimedia Appendix 2\]](#)

---

**Multimedia Appendix 3**

Definitions and rules to classify the implementation contexts.

[\[DOCX File \(Microsoft Word File\), 26 KB-Multimedia Appendix 3\]](#)

---

**Multimedia Appendix 4**

Mapping of included records to unique studies.

[\[DOCX File \(Microsoft Word File\), 23 KB-Multimedia Appendix 4\]](#)

---

**Multimedia Appendix 5**

Detailed characteristics of the included studies.

[\[DOCX File \(Microsoft Word File\), 25 KB-Multimedia Appendix 5\]](#)

---

**Multimedia Appendix 6**

Risk of bias assessment.

[\[DOCX File \(Microsoft Word File\), 2887 KB-Multimedia Appendix 6\]](#)

---

**Multimedia Appendix 7**

Characteristics of the participants.

[\[DOCX File \(Microsoft Word File\), 43 KB-Multimedia Appendix 7\]](#)

---

**Multimedia Appendix 8**

Outcomes and effects.

[\[DOCX File \(Microsoft Word File\), 45 KB-Multimedia Appendix 8\]](#)

---

**Checklist 1**

PRISMA 2020 checklist.

[\[DOCX File \(Microsoft Word File\), 275 KB-Checklist 1\]](#)

---

**Checklist 2**

PRISMA-S checklist

[\[PDF File \(Adobe File\), 488 KB-Checklist 2\]](#)

---

**Checklist 3**

SWiM checklist

[\[DOCX File \(Microsoft Word File\), 21 KB-Checklist 3\]](#)

## References

- Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Milbank Q*. 2005;83(3):457-502. [doi: [10.1111/j.1468-0009.2005.00409.x](https://doi.org/10.1111/j.1468-0009.2005.00409.x)] [Medline: [16202000](https://pubmed.ncbi.nlm.nih.gov/16202000/)]
- Shi L. The impact of primary care: a focused review. *Scientifica (Cairo)*. 2012;2012:432892. [doi: [10.6064/2012/432892](https://doi.org/10.6064/2012/432892)] [Medline: [24278694](https://pubmed.ncbi.nlm.nih.gov/24278694/)]
- Ferdows NB, Kosar C, Panagiotou O, Rahman M. Greater physician supply associated with lower mortality in rural counties: a 23-year county-level longitudinal observational study. *INQUIRY*. 2025;62:469580251380412. [doi: [10.1177/00469580251380412](https://doi.org/10.1177/00469580251380412)] [Medline: [41170580](https://pubmed.ncbi.nlm.nih.gov/41170580/)]
- Kim AM, Park JH, Yoon TH, Kim Y. Hospitalizations for ambulatory care sensitive conditions as an indicator of access to primary care and excess of bed supply. *BMC Health Serv Res*. Apr 27, 2019;19(1):259. [doi: [10.1186/s12913-019-4098-x](https://doi.org/10.1186/s12913-019-4098-x)] [Medline: [31029134](https://pubmed.ncbi.nlm.nih.gov/31029134/)]
- Kuhlmann E, Falkenbach M, Brînzac MG, et al. Tackling the primary healthcare workforce crisis: time to talk about health systems and governance—a comparative assessment of nine countries in the WHO European region. *Hum Resour Health*. Dec 31, 2024;22(1):83. [doi: [10.1186/s12960-024-00965-2](https://doi.org/10.1186/s12960-024-00965-2)] [Medline: [39741313](https://pubmed.ncbi.nlm.nih.gov/39741313/)]
- Kueper JK, Liaw W, Lizotte DJ, Tsuei SHT. Why is primary care different? Considerations for machine learning development with electronic medical record data. *NEJM AI*. Apr 24, 2025;2(5):AIp2401122. [doi: [10.1056/AIp2401122](https://doi.org/10.1056/AIp2401122)]
- Meunier PY, Raynaud C, Guimaraes E, Gueyffier F, Letrilliart L. Barriers and facilitators to the use of clinical decision support systems in primary care: a mixed-methods systematic review. *Ann Fam Med*. 2023;21(1):57-69. [doi: [10.1370/afm.2908](https://doi.org/10.1370/afm.2908)] [Medline: [36690490](https://pubmed.ncbi.nlm.nih.gov/36690490/)]
- Bazemore AW, Petterson SM, McCulloch KK. US primary care workforce growth: a decade of limited progress, and projected needs through 2040. *J GEN INTERN MED*. Feb 2025;40(2):339-346. [doi: [10.1007/s11606-024-09121-x](https://doi.org/10.1007/s11606-024-09121-x)] [Medline: [39443342](https://pubmed.ncbi.nlm.nih.gov/39443342/)]
- National Center for Health Workforce Analysis. State of the U.S. Health Care Workforce, 2024. Health Resources and Services Administration; 2024. URL: <https://bhwh.hrsa.gov/sites/default/files/bureau-health-workforce/state-of-the-health-workforce-report-2024.pdf> [Accessed 2026-10-10]
- Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLOS Med*. Sep 21, 2010;7(9):e1000326. [doi: [10.1371/journal.pmed.1000326](https://doi.org/10.1371/journal.pmed.1000326)] [Medline: [20877712](https://pubmed.ncbi.nlm.nih.gov/20877712/)]
- Yousefi F, Dehnavieh R, Laberge M, et al. Opportunities, challenges, and requirements for artificial intelligence (AI) implementation in Primary Health Care (PHC): a systematic review. *BMC Prim Care*. Jun 9, 2025;26(1):196. [doi: [10.1186/s12875-025-02785-2](https://doi.org/10.1186/s12875-025-02785-2)] [Medline: [40490689](https://pubmed.ncbi.nlm.nih.gov/40490689/)]
- Katonai G, Arvai N, Mesko B. AI and primary care: scoping review. *J Med Internet Res*. Aug 15, 2025;27:e65950. [doi: [10.2196/65950](https://doi.org/10.2196/65950)] [Medline: [40836794](https://pubmed.ncbi.nlm.nih.gov/40836794/)]
- Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3(1):17. [doi: [10.1038/s41746-020-0221-y](https://doi.org/10.1038/s41746-020-0221-y)] [Medline: [32047862](https://pubmed.ncbi.nlm.nih.gov/32047862/)]
- Kwan JL, Lo L, Ferguson J, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. *BMJ*. Sep 17, 2020;370:m3216. [doi: [10.1136/bmj.m3216](https://doi.org/10.1136/bmj.m3216)] [Medline: [32943437](https://pubmed.ncbi.nlm.nih.gov/32943437/)]
- Ouanes K, Farhah N. Effectiveness of artificial intelligence (AI) in clinical decision support systems and care delivery. *J Med Syst*. Aug 12, 2024;48(1):74. [doi: [10.1007/s10916-024-02098-4](https://doi.org/10.1007/s10916-024-02098-4)] [Medline: [39133332](https://pubmed.ncbi.nlm.nih.gov/39133332/)]
- Ramgopal S, Sanchez-Pinto LN, Horvat CM, Carroll MS, Luo Y, Florin TA. Artificial intelligence-based clinical decision support in pediatrics. *Pediatr Res*. Jan 2023;93(2):334-341. [doi: [10.1038/s41390-022-02226-1](https://doi.org/10.1038/s41390-022-02226-1)] [Medline: [35906317](https://pubmed.ncbi.nlm.nih.gov/35906317/)]
- Marwaha JS, Kvedar JC. Crossing the chasm from model performance to clinical impact: the need to improve implementation and evaluation of AI. *NPJ Digit Med*. Mar 3, 2022;5(1):25. [doi: [10.1038/s41746-022-00572-2](https://doi.org/10.1038/s41746-022-00572-2)] [Medline: [35241790](https://pubmed.ncbi.nlm.nih.gov/35241790/)]
- Rahamtalla BM, Medani IE, Abdelhag ME, et al. The AI-powered healthcare ecosystem: bridging the chasm between technical validation and systemic integration—a systematic review. *Future Internet*. 2025;17(12):550. [doi: [10.3390/fi17120550](https://doi.org/10.3390/fi17120550)]
- van de Sande D, Chung EFF, Oosterhoff J, van Bommel J, Gommers D, van Genderen ME. To warrant clinical adoption AI models require a multi-faceted implementation evaluation. *NPJ Digit Med*. Mar 6, 2024;7(1):58. [doi: [10.1038/s41746-024-01064-1](https://doi.org/10.1038/s41746-024-01064-1)] [Medline: [38448743](https://pubmed.ncbi.nlm.nih.gov/38448743/)]

20. Akay EMZ, Hilbert A, Carlisle BG, Madai VI, Mutke MA, Frey D. Artificial intelligence for clinical decision support in acute ischemic stroke: a systematic review. *Stroke*. Jun 2023;54(6):1505-1516. [doi: [10.1161/STROKEAHA.122.041442](https://doi.org/10.1161/STROKEAHA.122.041442)] [Medline: [37216446](https://pubmed.ncbi.nlm.nih.gov/37216446/)]
21. Moazemi S, Vahdati S, Li J, et al. Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: a systematic review. *Front Med (Lausanne)*. 2023;10:1109411. [doi: [10.3389/fmed.2023.1109411](https://doi.org/10.3389/fmed.2023.1109411)] [Medline: [37064042](https://pubmed.ncbi.nlm.nih.gov/37064042/)]
22. Lin X, Liang C, Liu J, Lyu T, Ghumman N, Campbell B. Artificial intelligence-augmented clinical decision support systems for pregnancy care: systematic review. *J Med Internet Res*. Sep 16, 2024;26:e54737. [doi: [10.2196/54737](https://doi.org/10.2196/54737)] [Medline: [39283665](https://pubmed.ncbi.nlm.nih.gov/39283665/)]
23. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev*. Oct 2020;43(5):1235-1253. [doi: [10.1007/s10143-019-01163-8](https://doi.org/10.1007/s10143-019-01163-8)] [Medline: [31422572](https://pubmed.ncbi.nlm.nih.gov/31422572/)]
24. Iannone S, Kaur A, Johnson KB. Artificial intelligence in outpatient primary care: a scoping review on applications, challenges, and future directions. *J Gen Intern Med*. Jan 2026;41(2):364-373. [doi: [10.1007/s11606-025-09938-0](https://doi.org/10.1007/s11606-025-09938-0)] [Medline: [41152567](https://pubmed.ncbi.nlm.nih.gov/41152567/)]
25. Park SH, Cha WC. Application strategies for artificial intelligence- based clinical decision support system: from the simulation to the real-world. *Healthc Inform Res*. Jul 2022;28(3):185-187. [doi: [10.4258/hir.2022.28.3.185](https://doi.org/10.4258/hir.2022.28.3.185)] [Medline: [35982592](https://pubmed.ncbi.nlm.nih.gov/35982592/)]
26. Susanto AP, Lyell D, Widyanoro B, Berkovsky S, Magrabi F. How well do AI-enabled decision support systems perform in clinical settings. In: Bichel-Findlay J, Otero P, Scott P, Huesing E, editors. *MEDINFO 2023—The Future Is Accessible: Proceedings of the 19th World Congress on Medical and Health Informatics*. SAGE Publications; 2024.
27. Wang L, Zhang Z, Wang D, et al. Human-centered design and evaluation of AI-empowered clinical decision support systems: a systematic review. *Front Comput Sci*. Jun 2, 2023;5:2023. [doi: [10.3389/fcomp.2023.1187299](https://doi.org/10.3389/fcomp.2023.1187299)]
28. Lees AF, White A, Leu M, Robinson J, Hall MK, Doerning R. The costs and benefits of clinical decision support for radiology appropriate use criteria: a retrospective observational study. *Appl Clin Inform*. Oct 2025;16(5):1658-1663. [doi: [10.1055/a-2635-3820](https://doi.org/10.1055/a-2635-3820)] [Medline: [40523639](https://pubmed.ncbi.nlm.nih.gov/40523639/)]
29. Parsons CS, Zuiderwijk A, Orchard NA, Oosterhoff JHF, de Reuver M. Task-technology fit of artificial intelligence-based clinical decision support systems: a review of qualitative studies. *BMC Med Inform Decis Mak*. Oct 28, 2025;25(1):397. [doi: [10.1186/s12911-025-03237-8](https://doi.org/10.1186/s12911-025-03237-8)] [Medline: [41152858](https://pubmed.ncbi.nlm.nih.gov/41152858/)]
30. Shi Y, Amill-Rosario A, Rudin RS, et al. Barriers to using clinical decision support in ambulatory care: do clinics in health systems fare better? *J Am Med Inform Assoc*. Jul 30, 2021;28(8):1667-1675. [doi: [10.1093/jamia/ocab064](https://doi.org/10.1093/jamia/ocab064)] [Medline: [33895828](https://pubmed.ncbi.nlm.nih.gov/33895828/)]
31. Fernando M, Abell B, McPhail SM, Tyack Z, Tariq A, Naicker S. Applying the non-adoption, abandonment, scale-up, spread, and sustainability framework across implementation stages to identify key strategies to facilitate clinical decision support system integration within a large metropolitan health service: interview and focus group study. *JMIR Med Inform*. Oct 17, 2024;12:e60402. [doi: [10.2196/60402](https://doi.org/10.2196/60402)] [Medline: [39419497](https://pubmed.ncbi.nlm.nih.gov/39419497/)]
32. Bayor AA, Li J, Yang IA, Varnfield M. Designing clinical decision support systems (CDSS)-a user-centered lens of the design characteristics, challenges, and implications: systematic review. *J Med Internet Res*. Jun 20, 2025;27:e63733. [doi: [10.2196/63733](https://doi.org/10.2196/63733)] [Medline: [40540451](https://pubmed.ncbi.nlm.nih.gov/40540451/)]
33. Choudhury A, Asan O. Role of artificial intelligence in patient safety outcomes: systematic literature review. *JMIR Med Inform*. Jul 24, 2020;8(7):e18599. [doi: [10.2196/18599](https://doi.org/10.2196/18599)] [Medline: [32706688](https://pubmed.ncbi.nlm.nih.gov/32706688/)]
34. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. Mar 2019;28(3):231-237. [doi: [10.1136/bmjqs-2018-008370](https://doi.org/10.1136/bmjqs-2018-008370)]
35. Macrae C. Governing the safety of artificial intelligence in healthcare. *BMJ Qual Saf*. Jun 2019;28(6):495-498. [doi: [10.1136/bmjqs-2019-009484](https://doi.org/10.1136/bmjqs-2019-009484)] [Medline: [30979783](https://pubmed.ncbi.nlm.nih.gov/30979783/)]
36. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc*. 2012;19(1):121-127. [doi: [10.1136/amiajnl-2011-000089](https://doi.org/10.1136/amiajnl-2011-000089)] [Medline: [21685142](https://pubmed.ncbi.nlm.nih.gov/21685142/)]
37. Balendran A, Benchoufi M, Evgeniou T, Ravaut P. Algorithmovigilance, lessons from pharmacovigilance. *NPJ Digit Med*. Oct 2, 2024;7(1):270. [doi: [10.1038/s41746-024-01237-y](https://doi.org/10.1038/s41746-024-01237-y)] [Medline: [39358559](https://pubmed.ncbi.nlm.nih.gov/39358559/)]
38. Labkoff S, Oladimeji B, Kannry J, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. *J Am Med Inform Assoc*. Nov 1, 2024;31(11):2730-2739. [doi: [10.1093/jamia/ocae209](https://doi.org/10.1093/jamia/ocae209)] [Medline: [39325508](https://pubmed.ncbi.nlm.nih.gov/39325508/)]
39. Li J, Guo Z, Guo Y, et al. The absence of full lifecycle risk management for AI-based medical devices in radiology. *npj Digit Med*. 2026;9(1). [doi: [10.1038/s41746-026-02712-4](https://doi.org/10.1038/s41746-026-02712-4)]
40. Coderre-Ball A, Ravi Chandran A, Gnanapragasam V, Ataman R, Baser K, Scott-Meuser P. Do we have enough information? Assessing AI clinical decision support systems for implementation in primary care. In: Copeland A, Keshavjee K, Zavar A, Lin G, editors. *Presented at: From Silos to Synergy: Building Bridges Proceedings of the Future*

- of Health Leadership, Informatics and Policy Conference 2026. Feb 20, 2026; IOS Press. Toronto, Canada. 2026.[doi: [10.3233/SHTI260019](https://doi.org/10.3233/SHTI260019)]
41. Liu TYA, Hswen Y, Bonow RO, et al. Artificial intelligence across the JAMA network. *JAMA*. Nov 11, 2025;334(18):1626. [doi: [10.1001/jama.2025.15074](https://doi.org/10.1001/jama.2025.15074)]
  42. Weissman GE. Evaluation and regulation of artificial intelligence medical devices for clinical decision support. *Annu Rev Biomed Data Sci*. Aug 2025;8(1):81-99. [doi: [10.1146/annurev-biodatasci-103123-095824](https://doi.org/10.1146/annurev-biodatasci-103123-095824)] [Medline: [39971383](https://pubmed.ncbi.nlm.nih.gov/39971383/)]
  43. Muralidharan V, Adewale BA, Huang CJ, et al. A scoping review of reporting gaps in FDA-approved AI medical devices. *NPJ Digit Med*. Oct 3, 2024;7(1):273. [doi: [10.1038/s41746-024-01270-x](https://doi.org/10.1038/s41746-024-01270-x)] [Medline: [39362934](https://pubmed.ncbi.nlm.nih.gov/39362934/)]
  44. Sokol K, Fackler J, Vogt JE. Artificial intelligence should genuinely support clinical reasoning and decision making to bridge the translational gap. *NPJ Digit Med*. Jun 10, 2025;8(1):345. [doi: [10.1038/s41746-025-01725-9](https://doi.org/10.1038/s41746-025-01725-9)] [Medline: [40494886](https://pubmed.ncbi.nlm.nih.gov/40494886/)]
  45. Leuchter RK, Turner WB, Ouyang D. Evaluating translational AI: a two-way moving target problem. *NEJM AI*. Dec 2025;2(12):Aip2500705. [doi: [10.1056/aip2500705](https://doi.org/10.1056/aip2500705)] [Medline: [41613162](https://pubmed.ncbi.nlm.nih.gov/41613162/)]
  46. Bhuyan SS, Sateesh V, Mukul N, et al. Generative artificial intelligence use in healthcare: opportunities for clinical excellence and administrative efficiency. *J Med Syst*. Jan 16, 2025;49(1):10. [doi: [10.1007/s10916-024-02136-1](https://doi.org/10.1007/s10916-024-02136-1)] [Medline: [39820845](https://pubmed.ncbi.nlm.nih.gov/39820845/)]
  47. Hu M, Ma C, Li W, Xu W, Wu J, Hu J, et al. A survey of scientific large language models: from data foundations to agent frontiers. *arXiv*. Preprint posted online on Oct 18, 2025. [doi: [10.48550/arXiv.2508.21148](https://doi.org/10.48550/arXiv.2508.21148)]
  48. Raz A, Inbar Y, Paz Z, Hofman B, Itzhakov Z, Weinstein O. Use first, trust later? Exploring how health care providers view the gaps between AI's regulation and its implementation. *JMIR Form Res*. Nov 10, 2025;9:e74038. [doi: [10.2196/74038](https://doi.org/10.2196/74038)] [Medline: [41213090](https://pubmed.ncbi.nlm.nih.gov/41213090/)]
  49. Herman DS, Reece JT, Weissman GE. Lessons for local oversight of AI in medicine from the regulation of clinical laboratory testing. *NPJ Digit Med*. Dec 13, 2024;7(1):359. [doi: [10.1038/s41746-024-01369-1](https://doi.org/10.1038/s41746-024-01369-1)] [Medline: [39672919](https://pubmed.ncbi.nlm.nih.gov/39672919/)]
  50. Zhang J, Mattie H, Shuaib H, Hensman T, Teo JT, Celi LA. Addressing the “elephant in the room” of AI clinical decision support through organisation-level regulation. *PLOS Digit Health*. Sep 2022;1(9):e0000111. [doi: [10.1371/journal.pdig.0000111](https://doi.org/10.1371/journal.pdig.0000111)] [Medline: [36812576](https://pubmed.ncbi.nlm.nih.gov/36812576/)]
  51. Gomez-Cabello CA, Borna S, Pressman S, Haider SA, Haider CR, Forte AJ. Artificial-intelligence-based clinical decision support systems in primary care: a scoping review of current clinical implementations. *EJIHPE*. 2024;14(3):685-698. [doi: [10.3390/ejihpe14030045](https://doi.org/10.3390/ejihpe14030045)] [Medline: [268482155](https://pubmed.ncbi.nlm.nih.gov/268482155/)]
  52. Derksen C, Le Noury KA, Akbar AB, Round T, Walter FM, Scott SE. Why aren't they used? Systematic review of barriers to implementation of clinical decision support systems for early cancer detection in primary care. *Br J Gen Pract*. Apr 1, 2026;76(765):e264-e274. [doi: [10.3399/BJGP.2025.0579](https://doi.org/10.3399/BJGP.2025.0579)] [Medline: [41494775](https://pubmed.ncbi.nlm.nih.gov/41494775/)]
  53. Derksen C, Walter FM, Akbar AB, et al. The implementation challenge of computerised clinical decision support systems for the detection of disease in primary care: systematic review and recommendations. *Implement Sci*. Jul 17, 2025;20(1):33. [doi: [10.1186/s13012-025-01445-4](https://doi.org/10.1186/s13012-025-01445-4)] [Medline: [40671071](https://pubmed.ncbi.nlm.nih.gov/40671071/)]
  54. Show us the evidence for the value of medical AI. *Nat Med*. Apr 2026;32(4):1163. [doi: [10.1038/s41591-026-04389-4](https://doi.org/10.1038/s41591-026-04389-4)]
  55. You JG, Hernandez-Boussard T, Pfeffer MA, Landman A, Mishuris RG. Clinical trials informed framework for real world clinical implementation and deployment of artificial intelligence applications. *NPJ Digit Med*. Feb 17, 2025;8(1):107. [doi: [10.1038/s41746-025-01506-4](https://doi.org/10.1038/s41746-025-01506-4)] [Medline: [39962232](https://pubmed.ncbi.nlm.nih.gov/39962232/)]
  56. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions* Version. Cochrane; 2024. URL: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current> [Accessed 2025-06-17]
  57. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. Mar 29, 2021;372:n71. [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
  58. Artificial intelligence–based clinical decision support systems in primary care: a systematic review of clinical, provider, and system-level outcomes. PROSPERO – International Prospective Register of Systematic Reviews. URL: <https://www.crd.york.ac.uk/PROSPERO/view/CRD420251119225> [Accessed 2025-08-04]
  59. Hosseini MS, Jahanshahlou F, Akbarzadeh MA, Zarei M, Vaez-Gharamaleki Y. Formulating research questions for evidence-based studies. *Journal of Medicine, Surgery, and Public Health*. Apr 2024;2:100046. [doi: [10.1016/j.glmedi.2023.100046](https://doi.org/10.1016/j.glmedi.2023.100046)]
  60. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev*. Jan 26, 2021;10(1):39. [doi: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z)] [Medline: [33499930](https://pubmed.ncbi.nlm.nih.gov/33499930/)]
  61. Briscoe S, Featherstone R, Littlewood A, et al. Chapter 4: searching for and selecting studies. In: Higgins TH, Chandler J, Cumpston M, Li T, Page MJT, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane; 2025. URL: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-04> [Accessed 2025-06-17]

62. Briscoe S, Bethel A, Rogers M. Conduct and reporting of citation searching in Cochrane systematic reviews: a cross-sectional study. *Res Synth Methods*. Mar 2020;11(2):169-180. [doi: [10.1002/jrsm.1355](https://doi.org/10.1002/jrsm.1355)] [Medline: [31127978](https://pubmed.ncbi.nlm.nih.gov/31127978/)]
63. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. Sep 21, 2017;358:j4008. [doi: [10.1136/bmj.j4008](https://doi.org/10.1136/bmj.j4008)] [Medline: [28935701](https://pubmed.ncbi.nlm.nih.gov/28935701/)]
64. Meadows K. Cognitive interviewing methodologies. *Clin Nurs Res*. May 2021;30(4):375-379. [doi: [10.1177/10547738211014099](https://doi.org/10.1177/10547738211014099)] [Medline: [33998325](https://pubmed.ncbi.nlm.nih.gov/33998325/)]
65. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*. May 18, 2022;377:e070904. [doi: [10.1136/bmj-2022-070904](https://doi.org/10.1136/bmj-2022-070904)] [Medline: [35584845](https://pubmed.ncbi.nlm.nih.gov/35584845/)]
66. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. Sep 2020;26(9):1364-1374. [doi: [10.1038/s41591-020-1034-x](https://doi.org/10.1038/s41591-020-1034-x)] [Medline: [32908283](https://pubmed.ncbi.nlm.nih.gov/32908283/)]
67. Kwong JCC, Khondker A, Lajkosz K, et al. APPRAISE-AI tool for quantitative evaluation of AI studies for clinical decision support. *JAMA Netw Open*. Sep 5, 2023;6(9):e2335377. [doi: [10.1001/jamanetworkopen.2023.35377](https://doi.org/10.1001/jamanetworkopen.2023.35377)] [Medline: [37747733](https://pubmed.ncbi.nlm.nih.gov/37747733/)]
68. Granviken F, Meisingset I, Bach K, 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*. May 1, 2025;166(5):1167-1178. [doi: [10.1097/j.pain.0000000000003456](https://doi.org/10.1097/j.pain.0000000000003456)] [Medline: [39432806](https://pubmed.ncbi.nlm.nih.gov/39432806/)]
69. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. Aug 28, 2019;366:14898. [doi: [10.1136/bmj.l4898](https://doi.org/10.1136/bmj.l4898)] [Medline: [31462531](https://pubmed.ncbi.nlm.nih.gov/31462531/)]
70. Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A. Chapter 7: considering bias and conflicts of interest among the included studies. In: *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane; 2024. URL: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-07> [Accessed 2025-06-19]
71. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. Jan 16, 2020;368:l6890. [doi: [10.1136/bmj.l6890](https://doi.org/10.1136/bmj.l6890)] [Medline: [31948937](https://pubmed.ncbi.nlm.nih.gov/31948937/)]
72. McKenzie JE, Brennan SE. Chapter 12: synthesizing and presenting findings using other methods. In: *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane; 2024. URL: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-12> [Accessed 2026-04-16]
73. Greenhalgh T, Abimbola S. The NASSS framework - a synthesis of multiple theories of technology implementation. In: Scott P, editor. *Applied Interdisciplinary Theory in Health Informatics*. IOS Press; 2019. [doi: [10.3233/SHTI190123](https://doi.org/10.3233/SHTI190123)]
74. Greenhalgh T, Maylor H, Shaw S, et al. The NASSS-CAT tools for understanding, guiding, monitoring, and researching technology implementation projects in health and social care: protocol for an evaluation study in real-world settings. *JMIR Res Protoc*. May 13, 2020;9(5):e16861. [doi: [10.2196/16861](https://doi.org/10.2196/16861)] [Medline: [32401224](https://pubmed.ncbi.nlm.nih.gov/32401224/)]
75. Wright A, Sittig DF, Ash JS, et al. Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems. *J Am Med Inform Assoc*. May 1, 2011;18(3):232-242. [doi: [10.1136/amiajnl-2011-000113](https://doi.org/10.1136/amiajnl-2011-000113)] [Medline: [21415065](https://pubmed.ncbi.nlm.nih.gov/21415065/)]
76. Alcoceba-Herrero I, Coco-Martín MB, Jiménez-Pérez JM, 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*. Oct 8, 2024;13(19):5974. [doi: [10.3390/jcm13195974](https://doi.org/10.3390/jcm13195974)] [Medline: [39408035](https://pubmed.ncbi.nlm.nih.gov/39408035/)]
77. Areias AC, Janela D, Moulder RG, et al. Applying AI to safely and effectively scale care to address chronic MSK conditions. *J Clin Med*. Jul 26, 2024;13(15):4366. [doi: [10.3390/jcm13154366](https://doi.org/10.3390/jcm13154366)] [Medline: [39124635](https://pubmed.ncbi.nlm.nih.gov/39124635/)]
78. Chen J, Chokshi S, Hegde R, et al. Development, implementation, and evaluation of a personalized machine learning algorithm for clinical decision support: case study with shingles vaccination. *J Med Internet Res*. Apr 29, 2020;22(4):e16848. [doi: [10.2196/16848](https://doi.org/10.2196/16848)] [Medline: [32347813](https://pubmed.ncbi.nlm.nih.gov/32347813/)]
79. Cruz NP, Canales L, Muñoz JG, Pérez B, Arnott I. Improving adherence to clinical pathways through natural language processing on electronic medical records. *Stud Health Technol Inform*. Aug 21, 2019;264:561-565. [doi: [10.3233/SHTI190285](https://doi.org/10.3233/SHTI190285)] [Medline: [31437986](https://pubmed.ncbi.nlm.nih.gov/31437986/)]
80. Herter WE, Khuc J, Cinà G, et al. Impact of a machine learning-based decision support system for urinary tract infections: prospective observational study in 36 primary care practices. *JMIR Med Inform*. May 4, 2022;10(5):e27795. [doi: [10.2196/27795](https://doi.org/10.2196/27795)] [Medline: [35507396](https://pubmed.ncbi.nlm.nih.gov/35507396/)]
81. Liu J, Gibson E, Ramchal S, et al. Diabetic retinopathy screening with automated retinal image analysis in a primary care setting improves adherence to ophthalmic care. *Ophthalmol Retina*. Jan 2021;5(1):71-77. [doi: [10.1016/j.oret.2020.06.016](https://doi.org/10.1016/j.oret.2020.06.016)] [Medline: [32562885](https://pubmed.ncbi.nlm.nih.gov/32562885/)]

82. Popescu C, Golden G, Benrimoh D, et al. Evaluating the clinical feasibility of an artificial intelligence-powered, web-based clinical decision support system for the treatment of depression in adults: longitudinal feasibility study. *JMIR Form Res*. Oct 25, 2021;5(10):e31862. [doi: [10.2196/31862](https://doi.org/10.2196/31862)] [Medline: [34694234](https://pubmed.ncbi.nlm.nih.gov/34694234/)]
83. Qassim S, Golden G, Slowey D, et al. A mixed-methods feasibility study of a novel AI-enabled, web-based, clinical decision support system for the treatment of major depression in adults. *Journal of Affective Disorders Reports*. Dec 2023;14:100677. [doi: [10.1016/j.jadr.2023.100677](https://doi.org/10.1016/j.jadr.2023.100677)]
84. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. *Int J Med Inform*. May 2020;137:104072. [doi: [10.1016/j.ijmedinf.2019.104072](https://doi.org/10.1016/j.ijmedinf.2019.104072)] [Medline: [32200295](https://pubmed.ncbi.nlm.nih.gov/32200295/)]
85. Seol HY, Shrestha P, Muth JF, et al. Artificial intelligence-assisted clinical decision support for childhood asthma management: a randomized clinical trial. *PLoS One*. 2021;16(8):e0255261. [doi: [10.1371/journal.pone.0255261](https://doi.org/10.1371/journal.pone.0255261)] [Medline: [34339438](https://pubmed.ncbi.nlm.nih.gov/34339438/)]
86. Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med*. May 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/)]
87. Rushlow DR, Croghan IT, Inselman JW, et al. Clinician adoption of an artificial intelligence algorithm to detect left ventricular systolic dysfunction in primary care. *Mayo Clin Proc*. Nov 2022;97(11):2076-2085. [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/)]
88. Thao V, Zhu Y, Tseng AS, et al. Cost-effectiveness of artificial intelligence-enabled electrocardiograms for early detection of low ejection fraction: a secondary analysis of the electrocardiogram artificial intelligence-guided screening for low ejection fraction trial. *Mayo Clin Proc Digit Health*. Dec 2024;2(4):620-631. [doi: [10.1016/j.mcpdig.2024.10.001](https://doi.org/10.1016/j.mcpdig.2024.10.001)] [Medline: [40206537](https://pubmed.ncbi.nlm.nih.gov/40206537/)]
89. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. *Ann Intern Med*. Jul 3, 2012;157(1):29-43. [doi: [10.7326/0003-4819-157-1-201207030-00450](https://doi.org/10.7326/0003-4819-157-1-201207030-00450)] [Medline: [22751758](https://pubmed.ncbi.nlm.nih.gov/22751758/)]
90. Miller A, Moon B, Anders S, Walden R, Brown S, Montella D. Integrating computerized clinical decision support systems into clinical work: a meta-synthesis of qualitative research. *Int J Med Inform*. Dec 2015;84(12):1009-1018. [doi: [10.1016/j.ijmedinf.2015.09.005](https://doi.org/10.1016/j.ijmedinf.2015.09.005)] [Medline: [26391601](https://pubmed.ncbi.nlm.nih.gov/26391601/)]
91. Reese TJ, Liu S, Steitz B, et al. Conceptualizing clinical decision support as complex interventions: a meta-analysis of comparative effectiveness trials. *J Am Med Inform Assoc*. Sep 12, 2022;29(10):1744-1756. [doi: [10.1093/jamia/ocac089](https://doi.org/10.1093/jamia/ocac089)] [Medline: [35652167](https://pubmed.ncbi.nlm.nih.gov/35652167/)]
92. Militello LG, Arbuckle NB, Saleem JJ, et al. Sources of variation in primary care clinical workflow: implications for the design of cognitive support. *Health Informatics J*. Mar 2014;20(1):35-49. [doi: [10.1177/1460458213476968](https://doi.org/10.1177/1460458213476968)] [Medline: [24105625](https://pubmed.ncbi.nlm.nih.gov/24105625/)]
93. Jeffries M, Salema NE, Laing L, et al. Using sociotechnical theory to understand medication safety work in primary care and prescribers' use of clinical decision support: a qualitative study. *BMJ Open*. Apr 27, 2023;13(4):e068798. [doi: [10.1136/bmjopen-2022-068798](https://doi.org/10.1136/bmjopen-2022-068798)] [Medline: [37105697](https://pubmed.ncbi.nlm.nih.gov/37105697/)]
94. Ancker JS, Edwards A, Nosal S, et al. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Med Inform Decis Mak*. Apr 10, 2017;17(1):36. [doi: [10.1186/s12911-017-0430-8](https://doi.org/10.1186/s12911-017-0430-8)] [Medline: [28395667](https://pubmed.ncbi.nlm.nih.gov/28395667/)]
95. Olakotan OO, Mohd Yusof M. The appropriateness of clinical decision support systems alerts in supporting clinical workflows: a systematic review. *Health Informatics J*. 2021;27(2):14604582211007536. [doi: [10.1177/14604582211007536](https://doi.org/10.1177/14604582211007536)] [Medline: [33853395](https://pubmed.ncbi.nlm.nih.gov/33853395/)]
96. Salwei ME, Ancker JS, Weinger MB. The decision aid is the easy part: workflow challenges of shared decision making in cancer care. *J Natl Cancer Inst*. Nov 8, 2023;115(11):1271-1277. [doi: [10.1093/jnci/djad133](https://doi.org/10.1093/jnci/djad133)] [Medline: [37421403](https://pubmed.ncbi.nlm.nih.gov/37421403/)]
97. Nanji KC, Slight SP, Seger DL, et al. Overrides of medication-related clinical decision support alerts in outpatients. *J Am Med Inform Assoc*. 2014;21(3):487-491. [doi: [10.1136/amiainl-2013-001813](https://doi.org/10.1136/amiainl-2013-001813)] [Medline: [24166725](https://pubmed.ncbi.nlm.nih.gov/24166725/)]
98. Nanji KC, Seger DL, Slight SP, et al. Medication-related clinical decision support alert overrides in inpatients. *J Am Med Inform Assoc*. May 1, 2018;25(5):476-481. [doi: [10.1093/jamia/ocx115](https://doi.org/10.1093/jamia/ocx115)] [Medline: [29092059](https://pubmed.ncbi.nlm.nih.gov/29092059/)]
99. van Baalen S, Boon M, Verhoef P. From clinical decision support to clinical reasoning support systems. *J Eval Clin Pract*. Jun 2021;27(3):520-528. [doi: [10.1111/jep.13541](https://doi.org/10.1111/jep.13541)] [Medline: [33554432](https://pubmed.ncbi.nlm.nih.gov/33554432/)]
100. Grace A, Mahony C, O'Donoghue J, Heffernan T, Molony D, Carroll T. Evaluating the effectiveness of clinical decision support systems: the case of multimorbidity care. *Journal of Decision Systems*. Apr 2013;22(2):97-108. [doi: [10.1080/12460125.2013.780320](https://doi.org/10.1080/12460125.2013.780320)]
101. Fraccaro P, Arguello Casteleiro M, Ainsworth J, Buchan I. Adoption of clinical decision support in multimorbidity: a systematic review. *JMIR Med Inform*. Jan 7, 2015;3(1):e4. [doi: [10.2196/medinform.3503](https://doi.org/10.2196/medinform.3503)] [Medline: [25785897](https://pubmed.ncbi.nlm.nih.gov/25785897/)]

102. Sintchenko V, Coiera E. Decision complexity affects the extent and type of decision support use. *AMIA Annu Symp Proc.* 2006;2006:724-728. [Medline: [17238436](#)]
103. Berner ES, Maisiak RS, Heuderbert GR, Young KR Jr. Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system. *AMIA Annu Symp Proc.* 2003;2003:76-80. [Medline: [14728137](#)]
104. Mamede S, Zandbergen A, de Carvalho-Filho MA, et al. Role of knowledge and reasoning processes as predictors of resident physicians' susceptibility to anchoring bias in diagnostic reasoning: a randomised controlled experiment. *BMJ Qual Saf.* Aug 16, 2024;33(9):563-572. [doi: [10.1136/bmjqs-2023-016621](#)] [Medline: [38365449](#)]
105. Ly DP, Shekelle PG, Song Z. Evidence for anchoring bias during physician decision-making. *JAMA Intern Med.* Aug 1, 2023;183(8):818-823. [doi: [10.1001/jamainternmed.2023.2366](#)] [Medline: [37358843](#)]
106. Burton JW, Stein MK, Jensen TB. A systematic review of algorithm aversion in augmented decision making. *Behavioral Decision Making.* Apr 2020;33(2):220-239. [doi: [10.1002/bdm.2155](#)]
107. Bankuoru Egala S, Liang D. Algorithm aversion to mobile clinical decision support among clinicians: a choice-based conjoint analysis. *European Journal of Information Systems.* Nov 2024;33(6):1016-1032. [doi: [10.1080/0960085X.2023.2251927](#)]
108. Roshanov PS, Misra S, Gerstein HC, et al. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. *Implementation Sci.* Dec 2011;6(1):92. [doi: [10.1186/1748-5908-6-92](#)]
109. Chen W, Howard K, Gorham G, et al. Design, effectiveness, and economic outcomes of contemporary chronic disease clinical decision support systems: a systematic review and meta-analysis. *J Am Med Inform Assoc.* Sep 12, 2022;29(10):1757-1772. [doi: [10.1093/jamia/ocac110](#)] [Medline: [35818299](#)]
110. Abell B, Naicker S, Rodwell D, et al. Identifying barriers and facilitators to successful implementation of computerized clinical decision support systems in hospitals: a NASSS framework-informed scoping review. *Implementation Sci.* 2023;18(1):32. [doi: [10.1186/s13012-023-01287-y](#)]
111. Finkelstein J, Gabriel A, Schmer S, Truong TT, Dunn A. Identifying facilitators and barriers to implementation of AI-assisted clinical decision support in an electronic health record system. *J Med Syst.* Sep 18, 2024;48(1):89. [doi: [10.1007/s10916-024-02104-9](#)] [Medline: [39292314](#)]
112. Wenderott K, Krups J, Luetkens JA, Weigl M. Radiologists' perspectives on the workflow integration of an artificial intelligence-based computer-aided detection system: a qualitative study. *Appl Ergon.* May 2024;117:104243. [doi: [10.1016/j.apergo.2024.104243](#)] [Medline: [38306741](#)]
113. Goldberg DG, Soylu TG, Hoffman CF, Kishton RE, Cronholm PF. Clinicians' perspectives on the adoption and implementation of EMR-integrated clinical decision support tools in primary care. *Digit HEALTH.* 2025;11:20552076251334043. [doi: [10.1177/20552076251334043](#)] [Medline: [40297364](#)]
114. Spijkerboer F, Van Es B, Haitjema S, Guglielmo M. Practical implementation lessons from the trenches of cardiovascular decision support deployment. *European Heart Journal - Digital Health.* Jan 12, 2026;7(Supplement\_1). [doi: [10.1093/ehjdh/ztaf143.174](#)]
115. Wilhelm C, Steckelberg A, Rebitschek FG. Benefits and harms associated with the use of AI-related algorithmic decision-making systems by healthcare professionals: a systematic review. *Lancet Reg Health Eur.* Jan 2025;48:101145. [doi: [10.1016/j.lanepe.2024.101145](#)] [Medline: [39687669](#)]
116. Jaspers MWM, Smeulders M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc.* May 1, 2011;18(3):327-334. [doi: [10.1136/amiajnl-2011-000094](#)] [Medline: [21422100](#)]
117. Freundlich RE, Wanderer JP, Ehrenfeld JM. Clinical decision support tools need to improve more than just process outcomes. *Anesthesiology.* Sep 2018;129(3):614. [doi: [10.1097/ALN.0000000000002349](#)] [Medline: [30106787](#)]
118. Ciani O, Buyse M, Garside R, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: meta-epidemiological study. *BMJ.* Jan 29, 2013;346:f457. [doi: [10.1136/bmj.f457](#)] [Medline: [23360719](#)]
119. Susanto AP, Lyell D, Widyanoto B, Berkovsky S, Magrabi F. Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: a scoping review. *J Am Med Inform Assoc.* Nov 17, 2023;30(12):2050-2063. [doi: [10.1093/jamia/ocad180](#)] [Medline: [37647865](#)]
120. Newton N, Bamgboje-Ayodele A, Forsyth R, Tariq A, Baysari MT. A systematic review of clinicians' acceptance and use of clinical decision support systems over time. *NPJ Digit Med.* May 26, 2025;8(1):309. [doi: [10.1038/s41746-025-01662-7](#)] [Medline: [40419669](#)]
121. Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. *Implementation Sci.* Dec 2017;12(1):113. [doi: [10.1186/s13012-017-0644-2](#)]
122. Bujold M, Pluye P, Légaré F, Participatory Review Team. Decision-making and related outcomes of patients with complex care needs in primary care settings: a systematic literature review with a case-based qualitative synthesis. *BMC Prim Care.* Nov 9, 2022;23(1):279. [doi: [10.1186/s12875-022-01879-5](#)] [Medline: [36352376](#)]

123. Loftus TJ, Balch JA, Marquard JL, et al. Longitudinal clinical decision support for assessing decisions over time: State-of-the-art and future directions. *Digit HEALTH*. 2024;10:20552076241249925. [doi: [10.1177/20552076241249925](https://doi.org/10.1177/20552076241249925)] [Medline: [38708184](https://pubmed.ncbi.nlm.nih.gov/38708184/)]
124. Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *J Am Med Inform Assoc*. Mar 1, 2017;24(2):246-250. [doi: [10.1093/jamia/ocw154](https://doi.org/10.1093/jamia/ocw154)] [Medline: [28011595](https://pubmed.ncbi.nlm.nih.gov/28011595/)]
125. Garzón González G, Alonso Safont T, Aguado Arroyo O, et al. Post-pandemic patient safety: have the characteristics of incidents with harm changed? Comparative observational study in primary care via review of medical records with a trigger tool. *BMC Prim Care*. Nov 7, 2024;25(1):392. [doi: [10.1186/s12875-024-02639-3](https://doi.org/10.1186/s12875-024-02639-3)] [Medline: [39511497](https://pubmed.ncbi.nlm.nih.gov/39511497/)]
126. Magrabi F, Ammenwerth E, McNair JB, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. *Yearb Med Inform*. Aug 2019;28(1):128-134. [doi: [10.1055/s-0039-1677903](https://doi.org/10.1055/s-0039-1677903)] [Medline: [31022752](https://pubmed.ncbi.nlm.nih.gov/31022752/)]
127. Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med*. Jul 15, 2021;385(3):283-286. [doi: [10.1056/NEJMc2104626](https://doi.org/10.1056/NEJMc2104626)] [Medline: [34260843](https://pubmed.ncbi.nlm.nih.gov/34260843/)]
128. Patterson ES, Su G, Sarkar U. Reducing delays to diagnosis in ambulatory care settings: a macrocognition perspective. *Appl Ergon*. Jan 2020;82:102965. [doi: [10.1016/j.apergo.2019.102965](https://doi.org/10.1016/j.apergo.2019.102965)] [Medline: [31605828](https://pubmed.ncbi.nlm.nih.gov/31605828/)]
129. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. Oct 29, 2019;17(1):195. [doi: [10.1186/s12916-019-1426-2](https://doi.org/10.1186/s12916-019-1426-2)] [Medline: [31665002](https://pubmed.ncbi.nlm.nih.gov/31665002/)]
130. Syrowatka A, Motala A, Lawson E, Shekelle P. Computerized Clinical Decision Support To Prevent Medication Errors and Adverse Drug Events: Rapid Review. Agency for Healthcare Research and Quality; 2023. URL: <https://www.ncbi.nlm.nih.gov/books/NBK600580/> [Accessed 2026-04-28]
131. Coiera E, Westbrook J, Wyatt J. The safety and quality of decision support systems. *Yearb Med Inform*. 2006;2006:20-25. [Medline: [17051290](https://pubmed.ncbi.nlm.nih.gov/17051290/)]
132. Salwei ME, Carayon P. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. *J Cogn Eng Decis Mak*. Dec 2022;16(4):194-206. [doi: [10.1177/15553434221097357](https://doi.org/10.1177/15553434221097357)] [Medline: [36704421](https://pubmed.ncbi.nlm.nih.gov/36704421/)]
133. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *J Am Med Inform Assoc*. Mar 1, 2017;24(2):423-431. [doi: [10.1093/jamia/ocw105](https://doi.org/10.1093/jamia/ocw105)] [Medline: [27516495](https://pubmed.ncbi.nlm.nih.gov/27516495/)]
134. Murad DA, Tsugawa Y, Elashoff DA, Baldwin KM, Bell DS. Distinct components of alert fatigue in physicians' responses to a noninterruptive clinical decision support alert. *J Am Med Inform Assoc*. Dec 13, 2022;30(1):64-72. [doi: [10.1093/jamia/ocac191](https://doi.org/10.1093/jamia/ocac191)] [Medline: [36264258](https://pubmed.ncbi.nlm.nih.gov/36264258/)]
135. Wright A, Ai A, Ash J, et al. Clinical decision support alert malfunctions: analysis and empirically derived taxonomy. *J Am Med Inform Assoc*. May 1, 2018;25(5):496-506. [doi: [10.1093/jamia/ocx106](https://doi.org/10.1093/jamia/ocx106)] [Medline: [29045651](https://pubmed.ncbi.nlm.nih.gov/29045651/)]
136. Sittig DF, Singh H. Recommendations to ensure safety of AI in real-world clinical care. *JAMA*. Feb 11, 2025;333(6):457-458. [doi: [10.1001/jama.2024.24598](https://doi.org/10.1001/jama.2024.24598)] [Medline: [39602298](https://pubmed.ncbi.nlm.nih.gov/39602298/)]
137. Kringos D, Boerma W, Bourgueil Y, et al. The strength of primary care in Europe: an international comparative study. *Br J Gen Pract*. Nov 2013;63(616):e742-50. [doi: [10.3399/bjgp13X674422](https://doi.org/10.3399/bjgp13X674422)] [Medline: [24267857](https://pubmed.ncbi.nlm.nih.gov/24267857/)]
138. Ronan CE, Crable EL, Drainoni ML, Walkey AJ. The impact of clinical decision support systems on provider behavior in the inpatient setting: a systematic review and meta-analysis. *J Hosp Med*. May 2022;17(5):368-383. [doi: [10.1002/jhm.12825](https://doi.org/10.1002/jhm.12825)] [Medline: [35514024](https://pubmed.ncbi.nlm.nih.gov/35514024/)]

## Abbreviations

**AI-CDSS:** AI-based clinical decision support system

**CAT:** Complexity Assessment Tool

**CDSS:** clinical decision support systems

**CONSORT-AI:** Consolidated Standards of Reporting Trials–AI

**EHR:** electronic health record

**GRADE:** Grading of Recommendations Assessment, Development, and Evaluation

**NASSS:** Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability

**PICO:** population, intervention, comparator, and outcomes

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

**PRISMA-S:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension

**RCT:** randomized controlled trial

**SWiM:** Synthesis Without Meta-analysis

*Edited by Tiffany Leung, Stefano Brini, Rachele Hendricks-Sturup; peer-reviewed by Abinav Yadamani, Yihan Hu, Zhao Liu; submitted 01 Apr.2026; final revised version received 22 Jul.2026; accepted 31 Jul.2026; published 19 Oct.2026*

*Please cite as:*

Naye F, Ouellet S, Yousefi F, Cachinho C, Bergeron F, Bhattacharyya O, Stamenova V, Sasseville M  
*AI-Based Clinical Decision Support Systems in Primary Care: Systematic Review of Clinical Implementation*  
*J Med Internet Res* 2026;28:e96882  
URL: <https://www.jmir.org/2026/1/e96882>  
doi: [10.2196/96882](https://doi.org/10.2196/96882)

© Florian Naye, Steven Ouellet, Farzaneh Yousefi, Chloé Cachinho, Frédéric Bergeron, Onil Bhattacharyya, Vess Stamenova, Maxime Sasseville. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 19.Oct.2026. 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.