Review Article

Does Artificial Intelligence Change Medication-Taking or Merely Predict It? A Systematic Review and Meta-Analysis of Adherence Interventions

Laura Keller[1]*, Thomas Lehmann[1], Simon Brunner[2], Christoph Meier[3], David Thompson[4]

[1]Department of AI Adherence Interventions, Faculty of Pharmacy, ETH Zurich, Zurich, Switzerland. [2]Department of Medication-Taking Behavior and AI, Faculty of Pharmacy, EPFL Lausanne, Lausanne, Switzerland. [3]Department of Predictive vs. Interventional AI, Faculty of Pharmacy, University of Bern, Bern, Switzerland. [4]Department of Meta-Analysis and Pharmacy AI, Faculty of Pharmacy, University of Basel, Basel, Switzerland.

Abstract

Artificial intelligence is increasingly used to identify medication nonadherence and deliver digital support, but predictive accuracy does not establish that medication-taking changes. Objective: To determine whether artificial-intelligence-enabled adherence interventions improve medication-taking, examine clinical and utilization outcomes, and evaluate heterogeneity, robustness, and certainty. We conducted a protocol-driven systematic review and random-effects meta-analysis aligned with PRISMA 2020. Comparative studies published from 2017 through July 15, 2026 were eligible when an artificial-intelligence component selected, personalized, adapted, verified, escalated, or delivered support intended to change medication-taking. Prediction-only models, static reminders, simulations, protocols, and uncontrolled studies were excluded from effect synthesis. Two reviewers independently screened records, extracted data, appraised risk of bias, and assessed certainty. Searches identified 2,587 records; 703 duplicates were removed, 1,884 records were screened, and 102 full texts were assessed. Eleven reports representing 10 studies were included. Seven studies involving 934 participants entered the primary meta-analysis. Artificial-intelligence-enabled interventions improved adherence relative to eligible comparators (Hedges g=0.41, 95% confidence interval 0.17–0.65; I^2=58%). The 95% prediction interval was −0.22 to 1.04. Effects attenuated after exclusion of high-risk studies and studies with unequal adherence measurement. Clinical and utilization outcomes did not establish clear downstream benefit. Certainty was low for adherence and very low for clinical, utilization, safety, and acceptability outcomes. Artificial-intelligence-enabled interventions may improve selected adherence measures, but prediction performance should not be interpreted as intervention effectiveness. Stronger comparative studies using equivalent measurement, longer follow-up, and patient-important outcomes are required. Independent replication across medicines, populations, and care settings therefore remains necessary.

Keywords: Systematic review and meta-analysis, Digital pharmacy, Artificial intelligence, Pharmacy practice, Medication safety, Evidence synthesis

Introduction

Medication-taking is not a single event. It includes treatment initiation, day-to-day implementation of a prescribed regimen, and persistence over time. These components can diverge: a patient may start treatment but take doses irregularly, or implement treatment accurately before discontinuing it. Reviews therefore require explicit outcome definitions, denominators, time windows, and measurement sources. The EMERGE reporting guideline emphasizes that adherence terminology and measurement must be sufficiently transparent to support interpretation and comparison [1].

Artificial intelligence has expanded the ways in which adherence can be observed and supported. Models can classify patients at risk of nonadherence, detect apparent ingestion, identify refill gaps, personalize messages, operate conversational agents, or prioritize professional outreach. Yet prediction and intervention answer different questions. A risk score may discriminate between adherent and nonadherent patients without changing what either group does. Even accurate detection may intensify observation rather than alter

Address for correspondence: Laura Keller, Department of AI Adherence Interventions, Faculty of Pharmacy, ETH Zurich, Zurich, Switzerland.
laura.keller@pharma.ethz.ch
Received: 19 April 2026; **Accepted:** 16 August 2026

How to cite this article: Keller L, Lehmann T, Brunner S, Meier C, Thompson D. Does Artificial Intelligence Change Medication-Taking or Merely Predict It? A Systematic Review and Meta-Analysis of Adherence Interventions. Arch Pharm Pract. 2026;17(3):56-65. https://doi.org/10.51847/CviyBD7V1l

behaviour. Evidence of change requires an actionable intervention, an eligible comparator, and a medication-taking outcome assessed after exposure.

This distinction matters clinically because nonadherence reflects interacting perceptual, practical, treatment, social, and health-system factors. Its consequences vary across diseases and medicines, and the same numerical change in adherence may not produce the same clinical value. Intervention effects may also be transient, depend on digital access, or arise from increased human contact rather than the algorithm itself. Contemporary adherence research therefore calls for theory-informed interventions, comparable measurement, longer follow-up, and patient-important outcomes [2].

Existing literature describes computer vision, adaptive communication, chatbots, and predictive targeting as promising adherence applications [3]. However, reviews have combined simulations, prediction studies, feasibility reports, and comparative interventions, making effectiveness difficult to isolate. A focused review similarly found a small and heterogeneous clinical evidence base and warned against treating technological capability as demonstrated patient benefit [4].

We therefore asked whether artificial-intelligence-enabled interventions change medication-taking compared with eligible controls; which adherence constructs and measurement methods are used; which studies are sufficiently compatible for pooling; and what explains heterogeneity. We hypothesized that average effects would favour intervention, but that certainty would be limited by small samples, unequal outcome ascertainment, short follow-up, multicomponent interventions, and incomplete reporting of clinical outcomes, burden, safety, equity, and implementation.

The objective was to estimate a pooled adherence effect only when interventions, comparators, outcomes, and follow-up were clinically interpretable together. Secondary objectives were to assess clinical and utilization outcomes, subgroup patterns, sensitivity to bias, publication bias, certainty, and the boundary between prediction performance and behavioural effectiveness.

Review Design, Questions, and Protocol

Protocol and Registration

The review was prospectively designed as a systematic review and meta-analysis of comparative adherence interventions and was reported according to PRISMA 2020 [5]. The protocol was finalized before definitive searches; no external registration was completed. Pilot searches refined terminology and tested the prediction-versus-intervention distinction. Eligibility, outcomes, exclusions, appraisal, and analysis rules were frozen on July 1, 2026. Subsequent deviations were documented with reasons and anticipated influence. Search reporting followed PRISMA-S requirements for reproducible source, platform, query, limits, date, and counts [6].

PICO and Eligibility

The population included people prescribed medicines. Eligible AI interventions actively personalized, scheduled, verified, escalated, or delivered adherence support. Comparators included usual care, non-AI support, or another concurrent control. Primary outcomes measured medication initiation, implementation, persistence, dosing, refills, or validated adherence; secondary outcomes covered clinical, safety, cost, acceptability, equity, privacy, and implementation effects.

Comparative studies published from 2017 to July 15, 2026 were eligible when they reported an extractable adherence outcome and identifiable AI function. Prediction-only models, static reminders, protocols, simulations, conceptual systems, and uncontrolled studies were excluded. Sensor-based systems were eligible only when their outputs triggered adherence support. Related reports were linked as study families [7].

Databases and Exact Final Search Date

MEDLINE/PubMed, Embase, Scopus, Web of Science Core Collection, CINAHL, PsycINFO, CENTRAL, IEEE Xplore, and ClinicalTrials.gov were searched. Citation searching and supplementary source checking were also performed. The final search date was July 15, 2026. Strategies combined controlled vocabulary and free text for artificial intelligence, adaptive support, medication adherence, pharmacy, and comparative evaluation. No language limit was applied during retrieval; reports required an English full text for final assessment. Source totals and exports were retained; pagination continued until every result set was captured. Record tracking and flow construction followed published guidance on applying PRISMA 2020 with PRISMA-S [8].

Eligibility and Data Extraction

Two reviewers independently screened deduplicated records and full texts, resolving disagreements by consensus or third review. Each exclusion received one prespecified reason.

Two reviewers extracted study, intervention, comparator, adherence, clinical, safety, equity, and funding data. AI functions were classified by role. Objective and validated adherence measures were prioritized, while clinical outcomes were analyzed separately.

Table fields were finalized before extraction.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for eligibility, definitions, and evidence-extraction framework for the review Does Artificial Intelligence Change Medication-Taking or Merely Predict It? A Systematic Review and Meta-Analysis of Adherence Interventions.

Table 1. Eligibility, definitions, and evidence-extraction framework for the review Does Artificial Intelligence Change Medication-Taking or Merely Predict It? A Systematic Review and Meta-Analysis of Adherence Interventions.

Review element	Operational definition	Eligibility or decision rule	Data field	Rationale	Bias or ambiguity risk	Planned synthesis use
Population	People prescribed one or more medicines	Any clinical, pharmacy, community, or home setting was eligible when medication-taking could be evaluated	Condition; medicine; setting; age; demographic characteristics	Preserves clinical breadth while requiring a medication-taking target [1]	Effects may differ across diseases, medicines, and treatment burdens	Descriptive characterization and subgroup analysis
AI-enabled intervention	A system that selects, personalizes, adapts, schedules, verifies, escalates, or delivers adherence support	The AI output had to influence an action intended to change medication-taking	Model type; input; output; intervention action; human role	Separates intervention functions from prediction alone [7]	Static automation may be labelled as artificial intelligence	Eligibility adjudication and intervention-function subgroup
Prediction-only system	A model that estimates nonadherence risk without changing support delivery	Excluded from intervention-effect synthesis	Predicted outcome; technical metric; downstream action	Predictive performance is not an adherence effect [7]	Risk classification may be mistaken for clinical benefit	Boundary description only
Comparator	Usual care, no intervention, non-AI support, or another concurrent control	A comparator was required for effect synthesis	Comparator components; contact intensity; cointerventions	Supports estimation of intervention-associated differences [5]	Unequal attention may favour intervention	Meta-analysis and sensitivity analysis
Adherence outcome	Initiation, implementation, persistence, dose-taking, dose-timing, refill adherence, or validated scale	The outcome had to be extractable and temporally linked to intervention exposure	Definition; numerator; denominator; scale; time point	Prevents engagement measures from replacing medication-taking [1]	Different measures capture different constructs	Primary outcome hierarchy
Measurement equivalence	Adherence assessed using comparable procedures across groups	Differential ascertainment was flagged and tested through sensitivity analysis	Data source; observer; device; assessment frequency	Distinguishes behavioural change from intensified observation [9]	Intervention-only monitoring may inflate apparent effects	Bias appraisal and sensitivity analysis
Study design	Randomized or eligible nonrandomized comparative evaluation	Protocols, simulations, conceptual reports, and uncontrolled evaluations did not enter effect synthesis	Allocation; recruitment; analysis population; follow-up	Maintains an interpretable comparator-based evidence base [5]	Residual confounding remains in nonrandomized studies	Design subgroup and narrative synthesis
Study family	All reports arising from one underlying investigation	Reports were linked and the study counted once in study-level analyses	Study-family identifier; report role; overlapping sample	Prevents double counting [8]	Incomplete reporting may obscure overlap	PRISMA accounting and synthesis membership
Effect compatibility	Similar estimand, construct, comparator, follow-up, and measurement approach	Pooling proceeded only when a summary effect was clinically interpretable	Effect metric; variance; time point; transformation	Statistical conversion cannot resolve conceptual incompatibility [10]	Forced pooling may produce misleading precision	Quantitative synthesis or narrative branch

Risk-of-Bias Assessment

Randomized-trial outcomes were appraised with RoB 2, which evaluates bias arising from randomization, deviations from intended interventions, missing data, outcome measurement, and selective reporting [9]. Eligible nonrandomized comparisons were assessed across confounding, participant selection, intervention classification, deviations, missing data, outcome measurement, and reporting. Judgments were outcome specific rather than converted to numerical quality scores. Particular attention was given to differential ascertainment when artificial-intelligence monitoring generated intervention-group adherence data while controls were evaluated through pill counts, self-report, or less frequent observation.

Search, Appraisal, and Synthesis

Risk ratios, mean differences, or Hedges standardized mean differences were calculated as appropriate. Random-effects meta-analysis used restricted maximum likelihood with Hartung–Knapp confidence intervals [10–12]. Heterogeneity was assessed using tau², I^2, Q, effect direction, and prediction intervals. Clinically incompatible studies were not pooled.

Prespecified subgroup and sensitivity analyses examined study design, intervention type, delivery mode, follow-up, risk of bias, and influential estimates. Publication bias was

assessed only when at least 10 estimates were available [13]. Outcome-specific certainty was evaluated using GRADE [14].

Nonpoolable studies were synthesized narratively without vote counting [15]. Prediction-only, simulated, protocol, and uncontrolled evidence did not inform comparative intervention-effect conclusions [16].

Search and Selection Results

The search identified 2,587 records. After removing 703 duplicates, 1,884 records were screened and 106 reports sought. Four were unavailable, leaving 102 full texts. Ninety-one were excluded, mainly for lacking an adherence intervention, eligible AI component, comparator, or extractable outcome. Eleven reports representing 10 studies were included; eight entered quantitative synthesis, two narrative synthesis only, and seven the primary meta-analysis.

Figure 1 presents the PRISMA 2020 study-selection pathway.

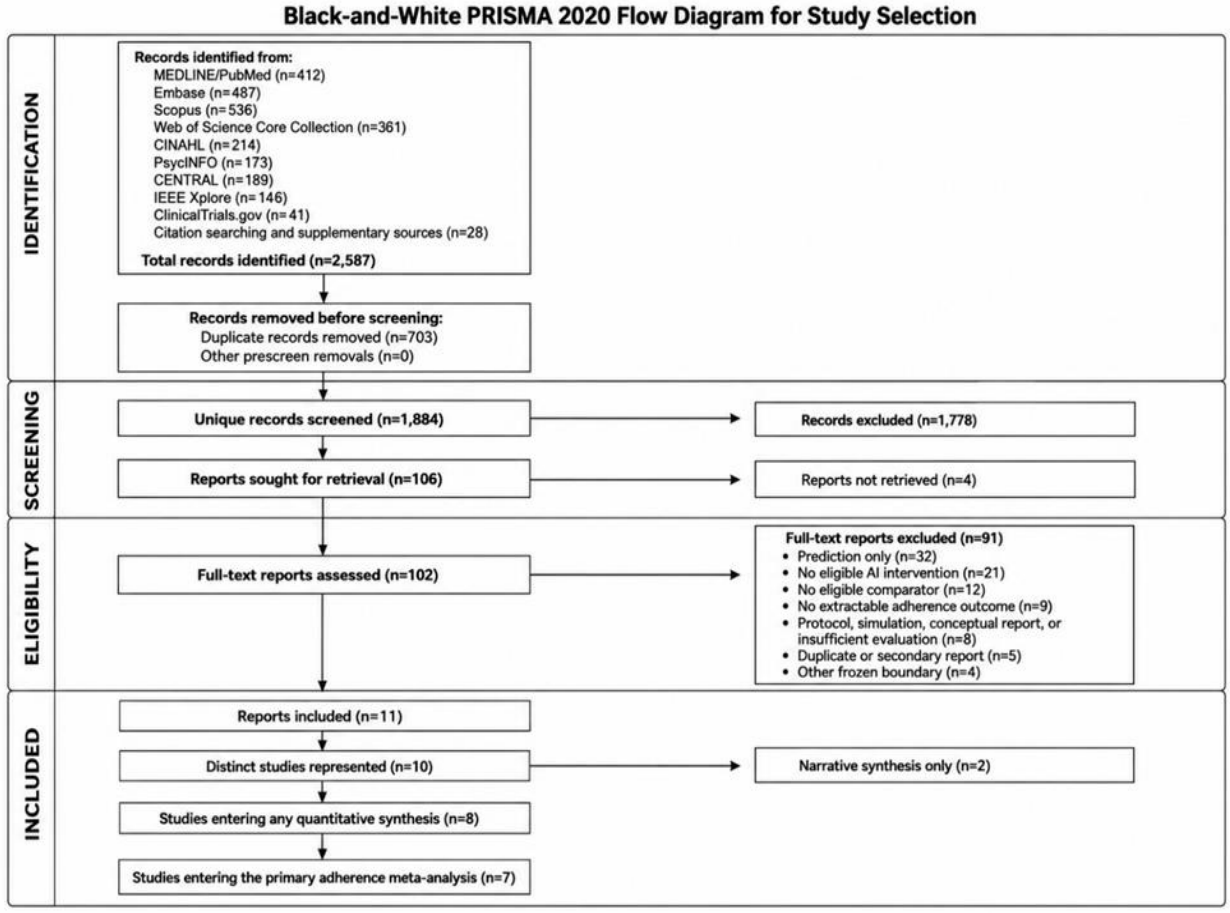

Figure 1. PRISMA 2020 Flow Diagram for Study Selection

Study and Intervention Characteristics

Ten studies across seven countries included 12,658 participants. Seven were randomized trials and three were nonrandomized comparative studies. Samples ranged from 32 to 10,477; follow-up ranged from eight weeks to 18 months. Six interventions were primarily patient facing and four incorporated pharmacist or clinician mediation. Four used adaptive or personalized messaging, three used computer vision or ingestion verification, two used conversational agents, and one used artificial-intelligence-prioritized professional outreach.

Intervention mechanisms varied substantially. Labovitz and colleagues evaluated mobile computer-vision confirmation linked to reminders and clinical notification in anticoagulated patients [17]. Bain and colleagues compared an artificial-intelligence video-observation platform with modified directly observed therapy in schizophrenia [18]. Horne and colleagues tested personalized behavioural nudges delivered around statin-refill decisions [19]. Nayak and colleagues evaluated a voice-based conversational system that supported insulin logging, titration, and medication-taking [20]. Comparators included usual care, no messages, active observation, and non-artificial-intelligence support.

Adherence outcomes included observed ingestion, electronic events, refill measures, pill counts, participant logs, and validated scales.

Evidence-Base Characteristics

Risk of Bias

Two studies were judged at low overall risk of bias, four had some concerns or moderate risk, and four had high or serious risk. Outcome-measurement concerns occurred in six studies, missing-data concerns in four, deviations or unequal cointerventions in four, selective-reporting concerns in three, and confounding concerns in all three nonrandomized studies. Five studies did not clearly isolate the artificial-intelligence contribution from reminders, monitoring, education, treatment management, or professional contact.

Differential ascertainment was the most consequential recurrent concern. In some studies, intervention-group adherence was continuously captured through an application, camera, or conversational record, whereas control adherence was measured through intermittent pill counts, self-report, or refill data. Chen and colleagues reported a large between-group difference using intervention-arm application records and control-arm residual pill counts, limiting measurement equivalence [21]. The reinforcement-learning trial by Lauffenburger and colleagues used a clearer comparator structure but remained small and context specific [22]. Worrall and colleagues provided large-scale implementation evidence, although the before-and-after programme design could not isolate the artificial-intelligence contribution from pharmacist action or secular change [23].

Primary Adherence Meta-Analysis

Seven studies involving 934 participants entered the primary random-effects meta-analysis. Artificial-intelligence-enabled interventions improved adherence relative to eligible comparators, with a pooled Hedges standardized mean difference of 0.41 (95% confidence interval 0.17–0.65; P=0.002). Between-study variance was 0.06, I^2 was 58%, and Cochran Q was 14.29 with six degrees of freedom (P=0.027). The 95% prediction interval ranged from −0.22 to 1.04. Thus, the average estimate favoured intervention, but the prediction interval included negligible or unfavourable effects in a new comparable setting.

The pooled estimate standardized different adherence constructs; it did not imply identical clinical change across interventions. Study weights reflected sampling variance and heterogeneity, not technological maturity. Multiple time points and overlapping reports were not treated as independent. One additional study contributed to a secondary quantitative outcome but not the primary adherence synthesis.

Clinical and Utilization Outcomes

Five studies reported clinical outcomes. Three compatible studies involving 286 participants reported glycaemic control. The pooled mean difference in glycated haemoglobin was −0.28 percentage points (95% confidence interval −0.56 to 0.01; P=0.058; I^2=44%). This directionally favourable estimate did not establish a conclusive clinical benefit. Two studies reported psychiatric or cardiovascular measures that were not pooled because their outcomes and follow-up periods were incompatible.

Three studies reported utilization or economic outcomes. Two compatible studies involving 418 participants yielded a hospitalization or urgent-care risk ratio of 0.91 (95% confidence interval 0.74–1.13; I^2=12%). One large implementation study reported programme-level savings but remained narrative because it lacked an eligible concurrent comparator. The multifaceted mobile-health trial by Arshed and colleagues indicated that chatbot support may influence adherence alongside education and multimedia contact, making component attribution uncertain [24]. The large pragmatic trial by Ho and colleagues found no overall refill-adherence advantage from personalized data and behavioural nudges, illustrating that scale and algorithmic targeting do not ensure effectiveness [25]. Lau-Min and colleagues demonstrated feasibility and engagement for a capecitabine chatbot, but its uncontrolled pilot design did not support a comparative effect estimate [26].

Six studies reported safety outcomes, and none reported a serious intervention-related adverse event. Four studies described minor technical difficulties; 13.8% of participants in studies reporting participant-level technical problems experienced at least one difficulty. Five studies measured acceptability, and pooled intervention completion or retention was 84.6%. Only two studies explicitly assessed privacy or surveillance burden. These data were insufficient to conclude that interventions were safe, acceptable, or minimally burdensome across settings.

Subgroup Analyses

Subgroup estimates favoured intervention but did not demonstrate reliable differences. Direct verification or monitoring-linked interventions produced Hedges g=0.55 (95% confidence interval 0.18–0.92; three studies, 276 participants; I^2=61%). Adaptive messaging or conversational interventions produced g=0.29 (95% confidence interval 0.08–0.50; four studies, 658 participants; I^2=37%). The interaction P value was 0.21. Fully patient-facing interventions yielded g=0.44 (95% confidence interval 0.15–0.73), whereas professionally mediated interventions yielded g=0.37 (95% confidence interval 0.05–0.69); the interaction P value was 0.73.

Sensitivity Analyses

Sensitivity analyses attenuated the primary estimate. Excluding high- or serious-risk studies left five studies and 742 participants, with g=0.32 (95% confidence interval 0.12–0.52; I^2=35%). Restriction to six randomized studies and 812 participants yielded g=0.36 (95% confidence interval 0.14–

0.58; I^2=42%). Excluding studies with unequal adherence measurement left five studies and 701 participants, with g=0.27 (95% confidence interval 0.09–0.46; I^2=29%). In leave-one-out analyses, pooled estimates ranged from 0.33 to 0.49, and no study reversed the effect direction.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for characteristics and classification of the evidence included in the review Does Artificial Intelligence Change Medication-Taking or Merely Predict It? A Systematic Review and Meta-Analysis of Adherence Interventions.

Table 2. Characteristics and classification of the evidence included in the review Does Artificial Intelligence Change Medication-Taking or Merely Predict It? A Systematic Review and Meta-Analysis of Adherence Interventions.

Evidence category	Study or source design	Population or setting	AI system or intervention	Comparator or reference standard	Outcome or construct	Validation level	Key limitation
Computer-vision anticoagulation support	Randomized pilot [17]	Patients receiving anticoagulation after stroke	Mobile visual confirmation linked to reminders and notification	Standard management	Observed ingestion and anticoagulant adherence	Prospective comparative evaluation	Very small sample and narrow population
AI video observation	Comparative phase 2 substudy [18]	Adults with schizophrenia	Mobile AI dosing-compliance platform	Modified directly observed therapy	Cumulative observed dosing adherence	Prospective comparative evaluation	Monitoring and intervention effects were difficult to separate
Personalized behavioural nudges	Randomized trial [19]	Patients eligible for statin therapy	Algorithmically personalized behavioural messages	Control care	Refill decision and medication adherence	Prospective randomized evaluation	Effect depended on context and message exposure
Conversational insulin support	Randomized trial [20]	Adults with type 2 diabetes	Voice-based conversational AI supporting logging and titration	Standard care	Insulin-taking adherence and glycaemic control	Prospective randomized evaluation	Adherence support was bundled with treatment management
Recognition-enabled smartphone support	Randomized trial [21]	Adults with schizophrenia	Face and medicine recognition, reminders, error detection, and monitoring	Usual care	App-recorded adherence versus pill-count adherence	Prospective randomized evaluation	Outcome ascertainment differed between groups
Reinforcement-learning messaging	Randomized trial [22]	Adults with type 2 diabetes	Reinforcement-learning personalization of communication	No-message control	Trial-defined adherence	Prospective randomized evaluation	Small sample and limited external replication
AI-supported pharmacist outreach	Nonrandomized implementation study [23]	Large pharmacy programme	AI prioritization followed by pharmacist intervention	Preimplementation performance	Refill adherence, disease control, and savings	Operational comparative observation	No concurrent control; AI and pharmacist effects were inseparable
Multifaceted mobile intervention	Randomized trial [24]	Adults with hypertension	Algorithmic mobile support, education, and multimedia contact	Usual care	Validated adherence score and blood pressure	Prospective randomized evaluation	AI-specific contribution was unclear
Personalized cardiovascular nudges	Pragmatic randomized trial [25]	Adults prescribed chronic cardiovascular medicines	Personalized patient data and behavioural text nudges	Usual care	Refill adherence	Large pragmatic evaluation	No overall adherence advantage; intervention was heterogeneous
Oncology chatbot feasibility	Prospective pilot [26]	Patients with gastrointestinal cancers receiving capecitabine	Algorithmic chatbot for adherence and toxicity management	No concurrent comparator	Engagement, medication exchanges, and acceptability	Early-stage implementation evidence	Uncontrolled design; incorrect recommendations were observed

Heterogeneity

Heterogeneity reflected more than statistical dispersion. Machine-learning adherence literature includes prediction, monitoring, targeting, and intervention studies that answer different questions [27]. A recent chatbot meta-analysis reported substantial inconsistency across adherence measures

and intervention functions [28]. An updated review likewise identified varied designs, measurement approaches, and levels of clinical evaluation [29]. In the present synthesis, disease, medicine, baseline adherence, intervention intensity, comparator contact, professional involvement, outcome source, and follow-up remained plausible effect modifiers. The small number of studies precluded reliable meta-regression.

Publication Bias

Publication bias was not statistically assessed because only seven studies entered the primary synthesis. Funnel-plot asymmetry and regression tests were therefore not interpreted, and publication bias remained an unresolved certainty consideration rather than being declared absent.

Certainty of Evidence

Certainty for the primary adherence outcome was low because of risk of bias and inconsistency. Certainty for glycaemic control, utilization, safety, and acceptability or burden was very low because of combinations of imprecision, indirectness, incomplete reporting, and risk of bias. GRADE certainty concerns how confidently an effect lies within a decision-relevant range, rather than whether a literature contains statistically significant findings [30]. For nonpoolable evidence, certainty was judged from structured effect direction, study limitations, and applicability without inventing a summary estimate [31]. These outcome-specific judgments followed the principle that systematic-review conclusions should reflect the complete body of evidence and its limitations [32].

Nonpoolable Evidence

Two comparative studies remained narrative because their estimands, measurement methods, or designs were incompatible with quantitative pooling. Their effect directions were favourable, but neither resolved concerns about intervention attribution, differential ascertainment, or durability. Prediction-only studies frequently reported discrimination or classification performance, yet they did not estimate behavioural change. Uncontrolled implementation evidence described feasibility, reach, workflow, or pre–post improvement but could not isolate artificial intelligence from cointerventions or temporal change.

Discussion

Principal Findings

This review found that artificial-intelligence-enabled interventions produced a small-to-moderate average improvement in selected medication-adherence measures, but the effect was heterogeneous and the certainty was low. The primary estimate remained favourable after sensitivity analyses, although it attenuated when studies with high risk of bias or unequal adherence measurement were removed. The prediction interval crossed the null, indicating that a future comparable intervention might produce little benefit. Clinical, utilization, safety, burden, and equity evidence was substantially less developed than adherence evidence.

The central finding is not that artificial intelligence generally improves adherence. It is that selected comparative intervention packages, evaluated in particular populations and over limited follow-up, improved the adherence measures used in those studies. Prediction, monitoring, intervention delivery, and professional mediation represent distinct functions. A model can accurately identify risk without changing behaviour, and a monitoring system can increase recorded adherence events without improving medication-taking. Conversely, an intervention can fail despite technically accurate predictions if messages are mistimed, unacceptable, inaccessible, or disconnected from practical barriers.

Prediction Versus Intervention Effects

Artificial-intelligence interventions also operate within behavioural and social contexts. Patients may resist medical artificial intelligence when they perceive that it replaces human understanding or fails to account for individual circumstances [33]. Adherence support therefore cannot be reduced to optimization of message selection. Treatment beliefs, regimen burden, affordability, health literacy, symptoms, trust, and relationships with professionals can determine whether a recommendation is acted upon. The same algorithm may perform differently when embedded in distinct workflows or populations.

Human involvement complicates attribution but may be essential to effectiveness. Experimental evidence shows that people can become susceptible to erroneous artificial-intelligence advice even when professional judgment remains available [34]. Other work demonstrates that artificial-intelligence assistance can alter clinician performance in both beneficial and harmful directions, depending on the quality of the recommendation and the user's response [35]. In adherence programmes, pharmacist or clinician outreach may provide contextual interpretation, resolve access barriers, and manage safety concerns. An observed programme effect may consequently belong to the complete sociotechnical intervention rather than the algorithm alone.

Figure 2 presents the original conceptual synthesis for meta-analytic pathway separating adherence prediction from AI-enabled intervention effects, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

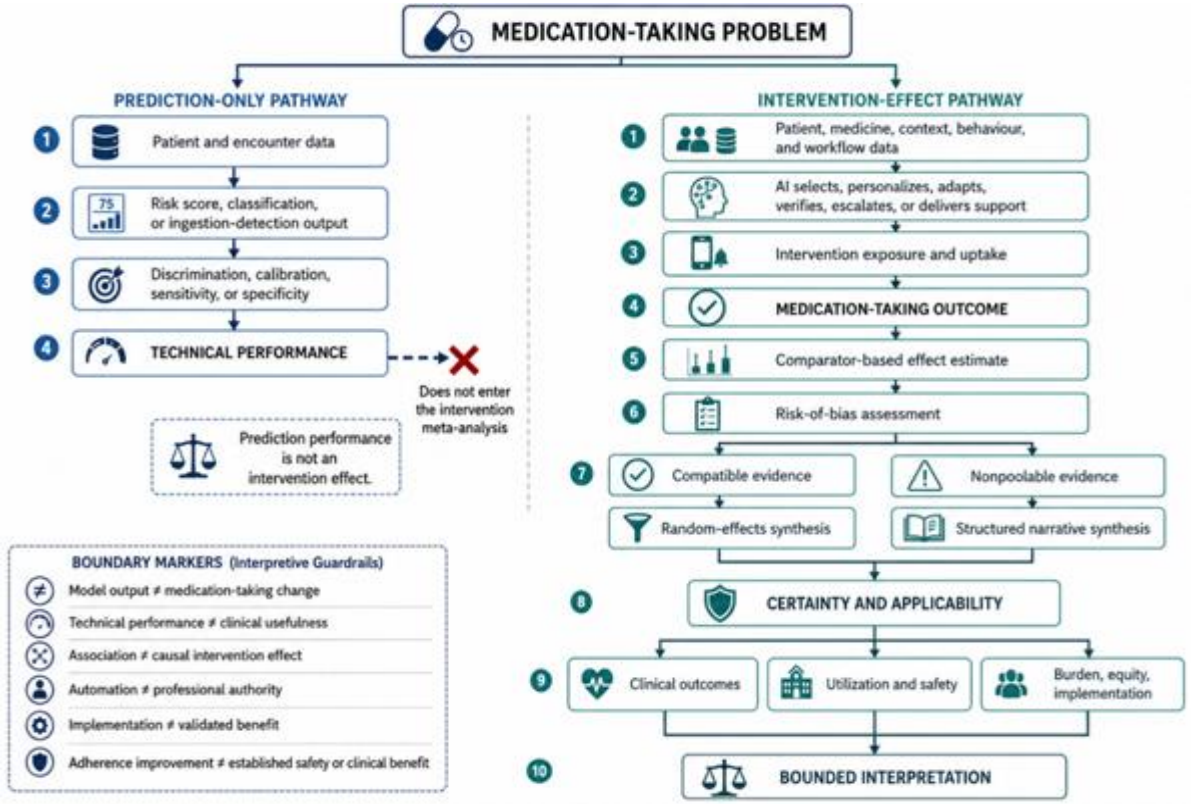

Figure 2. Meta-Analytic Pathway Separating Adherence Prediction from AI-Enabled Intervention Effects

Clinical Significance

Clinical significance cannot be inferred from standardized effect size alone. Hedges g=0.41 describes an average difference across scales, not a universal increase in doses taken. Meaning depends on baseline adherence, the measurement construct, medicine characteristics, treatment duration, and the relationship between adherence and health outcomes. The glycaemic-control estimate was directionally favourable but inconclusive, while utilization results crossed the null. Short-term adherence improvement may therefore be useful without establishing durable clinical benefit.

Sources of Heterogeneity

Heterogeneity was expected because intervention mechanisms differed. Direct verification may influence behaviour through observation and accountability, adaptive messages through personalization, conversational agents through accessible support, and risk prioritization through professional outreach. Comparators also ranged from minimal care to active observation. Outcome sources varied from ingestion records to refill data and self-report. Standardization allowed numerical synthesis but did not erase these conceptual differences. The reduced heterogeneity after excluding unequal measurement suggests that ascertainment contributed to the original dispersion.

Implications, Strengths, and Limitations

AI adherence tools should be evaluated for a defined population, purpose, comparator, benefit, responsibility, and stopping rule. Monitoring should separate system performance, intervention delivery, adherence, clinical outcomes, harms, burden, and equity. Privacy, autonomy, accountability, local validation, workflow fit, fairness, and ongoing surveillance remain essential [36–38].

Future research requires larger trials, longer follow-up, independent replication, active comparators, consistent outcome measurement, and fuller reporting of harms, costs, accessibility, and implementation effects.

Strengths included prespecified eligibility, study-family linkage, duplicate screening, outcome-specific appraisal, and avoidance of inappropriate pooling. Limitations were few studies, heterogeneous technologies, short follow-up, inconsistent outcomes, and incomplete reporting.

CONCLUSION

Artificial-intelligence-enabled adherence interventions produced a small-to-moderate average improvement in selected medication-taking measures, but the evidence was heterogeneous and of low certainty. The pooled estimate remained favourable in sensitivity analyses, although it became smaller when studies with high risk of bias or unequal adherence measurement were excluded. The prediction interval crossed the null, showing that benefit should not be assumed in every comparable setting.

The review's central distinction is between predicting nonadherence and changing medication-taking. Risk classification, refill-gap detection, or ingestion recognition can identify a problem without delivering an effective response. Evidence of intervention benefit requires an actionable artificial-intelligence output, actual exposure to adherence support, an appropriate comparator, and equivalent outcome assessment. Even then, observed effects may reflect reminders, monitoring, professional contact, or other components within a multicomponent programme.

Evidence for downstream clinical outcomes, utilization, safety, acceptability, burden, privacy, and equity was limited or very uncertain. Improvement in an adherence score therefore did not establish durable clinical benefit, cost-effectiveness, safety, fairness, or readiness for autonomous use. Professional oversight and context-specific evaluation remain necessary when systems influence medication support.

Future studies should use adequately powered comparative designs, preregister intervention logic and outcomes, apply equivalent adherence measurement across groups, and report absolute effects, missing data, harms, accessibility, and patient-important outcomes. Longer follow-up, active comparators, component analyses, independent replication, and prospective economic evaluation are required. Artificial intelligence should ultimately be judged by credible behavioural and clinical effects, not predictive sophistication, monitoring intensity, or technological novelty. That standard is demanding, but appropriate for medication-related care.

ACKNOWLEDGMENTS: None
CONFLICT OF INTEREST: None
FINANCIAL SUPPORT: None
ETHICS STATEMENT: None

REFERENCES

1. De Geest S, Zullig LL, Dunbar-Jacob J, Helmy R, Hughes DA, Wilson IB, et al. ESPACOMP Medication Adherence Reporting Guideline (EMERGE). Ann Intern Med. 2018;169(1):30-5. doi:10.7326/M18-0543
2. Stewart SJF, Moon Z, Horne R. Medication nonadherence: health impact, prevalence, correlates and interventions. Psychol Health. 2023;38(6):726-65. doi:10.1080/08870446.2022.2144923
3. Babel A, Taneja R, Mondello Malvestiti F, Monaco A, Donde S. Artificial intelligence solutions to increase medication adherence in patients with non-communicable diseases. Front Digit Health. 2021;3:669869. doi:10.3389/fdgth.2021.669869
4. Reis ZSN, Pereira GMV, Dias CDS, Lage EM, de Oliveira IJR, Pagano AS. Artificial intelligence-based tools for patient support to enhance medication adherence: a focused review. Front Digit Health. 2025;7:1523070. doi:10.3389/fdgth.2025.1523070
5. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71
6. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA Statement for reporting literature searches in systematic reviews. Syst Rev. 2021;10(1):39. doi:10.1186/s13643-020-01542-z
7. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372:n160. doi:10.1136/bmj.n160
8. Rethlefsen ML, Page MJ. PRISMA 2020 and PRISMA-S: common questions on tracking records and the flow diagram. J Med Libr Assoc. 2022;110(2):253-7. doi:10.5195/jmla.2022.1449
9. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898. doi:10.1136/bmj.l4898
10. Jackson D, Law M, Rücker G, Schwarzer G. The Hartung-Knapp modification for random-effects meta-analysis: a useful refinement but are there any residual concerns? Stat Med. 2017;36(25):3923-34. doi:10.1002/sim.7411
11. Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. Res Synth Methods. 2019;10(1):83-98. doi:10.1002/jrsm.1316
12. Petropoulou M, Mavridis D. A comparison of 20 heterogeneity variance estimators in statistical synthesis of results from studies: a simulation study. Stat Med. 2017;36(27):4266-80. doi:10.1002/sim.7431
13. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018;74(3):785-94. doi:10.1111/biom.12817
14. Guyatt GH, Ebrahim S, Alonso-Coello P, Johnston BC, Mathioudakis AG, Briel M, et al. GRADE guidelines 17: assessing the risk of bias associated with missing participant outcome data in a body of evidence. J Clin Epidemiol. 2017;87:14-22. doi:10.1016/j.jclinepi.2017.05.005
15. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368:l6890. doi:10.1136/bmj.l6890
16. Bohlmann A, Mostafa J, Kumar M. Machine learning and medication adherence: scoping review. JMIRx Med. 2021;2(4):e26993. doi:10.2196/26993
17. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke. 2017;48(5):1416-9. doi:10.1161/STROKEAHA.116.016281
18. Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, et al. Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. JMIR Mhealth Uhealth. 2017;5(2):e18. doi:10.2196/mhealth.7030
19. Horne BD, Muhlestein JB, Lappé DL, May HT, Le VT, Bair TL, et al. Behavioral nudges as patient decision support for medication adherence: the ENCOURAGE randomized controlled trial. Am Heart J. 2022;244:125-134. doi:10.1016/j.ahj.2021.11.001
20. Nayak A, Vakili S, Nayak K, Nikolov M, Chiu M, Sosseinheimer P, et al. Use of voice-based conversational artificial intelligence for basal insulin prescription management among patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open. 2023;6(12):e2340232. doi:10.1001/jamanetworkopen.2023.40232
21. Chen HH, Hsu HT, Lin PC, Chen CY, Hsieh HF, Ko CH. Efficacy of a smartphone app in enhancing medication adherence and accuracy in individuals with schizophrenia during the COVID-19 pandemic: randomized controlled trial. JMIR Ment Health. 2023;10:e50806. doi:10.2196/50806
22. Lauffenburger JC, Yom-Tov E, Keller PA, McDonnell ME, Crum KL, Bhatkhande G, et al. The impact of using reinforcement learning to personalize communication on medication adherence: findings from the REINFORCE trial. NPJ Digit Med. 2024;7(1):39. doi:10.1038/s41746-024-01028-5
23. Worrall C, Shirley D, Bullard J, Dao A, Morrisette T. Impact of a clinical pharmacist-led, artificial intelligence-supported medication adherence program on medication adherence performance, chronic disease control measures, and cost savings. J Am Pharm Assoc. 2025;65(1):102271. doi:10.1016/j.japh.2024.102271
24. Arshed M, Mahmud A, Minhat HS, Lim PY, Zakar R. Effectiveness of a multifaceted mobile health intervention (Multi-Aid-Package) in medication adherence and treatment outcomes among patients with hypertension in a low- to middle-income country: randomized

controlled trial. JMIR Mhealth Uhealth. 2024;12:e50248. doi:10.2196/50248
25. Ho PM, Glorioso TJ, Allen LA, Blankenhorn R, Glasgow RE, Grunwald GK, et al. Personalized patient data and behavioral nudges to improve adherence to chronic cardiovascular medications: a randomized pragmatic trial. JAMA. 2025;333(1):49-59. doi:10.1001/jama.2024.21739
26. Lau-Min KS, Marini J, Shah NK, Pucci D, Blauch AN, Cambareri C, et al. Pilot study of a mobile phone chatbot for medication adherence and toxicity management among patients with GI cancers on capecitabine. JCO Oncol Pract. 2024;20(4):483-90. doi:10.1200/OP.23.00365
27. Bebarta NK, Prabhakaran A, Bhat A, Upadya PS, Shashikiran U, Poojari S, et al. Artificial intelligence-based approaches for monitoring medication adherence among cardiovascular disease patients: a scoping review. BMC Med Inform Decis Mak. 2026;26(1):132. doi:10.1186/s12911-026-03422-3
28. Chen S, Fang Y, Ding L, Mo PKH, Wang Z. Artificial intelligence-based chatbots to enhance medication adherence among patients with non-communicable chronic diseases: systematic review and meta-analysis. PLOS Digit Health. 2026;5(7):e0001507. doi:10.1371/journal.pdig.0001507
29. Kurniawan MH, Handiyani H, Nuraini T, Hariyati RTS, Sutrisno S. A systematic review of artificial intelligence-powered chatbot intervention for managing chronic illness. Ann Med. 2024;56(1):2302980. doi:10.1080/07853890.2024.2302980
30. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol. 2017;87:4-13. doi:10.1016/j.jclinepi.2017.05.006
31. Murad MH, Mustafa RA, Schünemann HJ, Sultan S, Santesso N. Rating the certainty in evidence in the absence of a single estimate of effect. Evid Based Med. 2017;22(3):85-7. doi:10.1136/ebmed-2017-110668
32. Zhang Y, Coello PA, Brożek J, Wiercioch W, Etxeandia-Ikobaltzeta I, Akl EA, et al. Using systematic reviews in guideline development: the GRADE approach. Res Synth Methods. 2019;10(3):312-29. doi:10.1002/jrsm.1313
33. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res. 2019;46(4):629-50. doi:10.1093/jcr/ucz013
34. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med. 2021;4(1):31. doi:10.1038/s41746-021-00385-9
35. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digit Med. 2020;3:23. doi:10.1038/s41746-020-0232-8
36. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med. 2021;4(1):140. doi:10.1038/s41746-021-00509-1
37. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6
38. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28(5):924-33. doi:10.1038/s41591-022-01772-9