Review Article

Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods

Ahmed Ben Ali[1]*, Samira Gharbi[1], Nabil Jebali[2]

[1]Department of Hospital Pharmacy and Medication Safety, Faculty of Pharmacy, University of Tunis, Tunis, Tunisia. [2]Department of Algorithm Evaluation in Clinical Pharmacy, Faculty of Pharmacy, University of Sfax, Sfax, Tunisia.

Abstract

Hospital-pharmacy algorithms are used to identify atypical prescriptions, prioritize pharmacist review, filter medication alerts, and estimate medication-related risk. Their credibility depends on how tasks, labels, comparators, validation, and workflow consequences are evaluated. To review systematically the clinical tasks and evaluation methods used for hospital-pharmacy medication-safety algorithms published from 2017 through 2025. This PRISMA 2020-aligned systematic methodological review searched major biomedical and multidisciplinary databases and citation chains. Eligible Q1 journal studies evaluated an algorithm relevant to hospital medication safety or pharmacist medication review. Duplicate independent screening, structured extraction, PROBAST-informed appraisal, and narrative synthesis examined task definition, data source, reference standard, validation, comparator, performance metrics, calibration, uncertainty, interpretability, workflow, safety, equity, and reproducibility.

Sixty-three records were identified; eight duplicates were removed, 55 records were screened, and 40 reports underwent full-text assessment. Seventeen reports were excluded, leaving 23 studies for qualitative synthesis and none for quantitative synthesis. Studies addressed prescription-error detection, review prioritization, atypical-order identification, alert optimization, and adverse-event risk. Retrospective single-centre evaluation predominated; external validation, calibration analysis, prospective workflow comparison, subgroup error assessment, and reproducibility reporting were uncommon. The literature shows technical diversity but uneven methodological credibility. Discrimination or alert yield alone does not establish clinical usefulness, safety, equity, or implementation readiness. Stronger reference standards, independent validation, calibrated thresholds, prospective workflow testing, and explicit error-consequence analysis are required.

Keywords: Systematic methodological review, Digital pharmacy, Artificial intelligence, Pharmacy practice, Medication safety, Evidence synthesi

Introduction

Artificial intelligence in pharmacy spans medication-order screening, clinical decision support, prescription prioritization, natural-language processing, and operational applications. Reviews nevertheless describe substantial variation in clinical purpose, technical method, and evidence maturity [1, 2]. In hospitals, algorithms may act before pharmacist review, during alert assessment, as a second screen for atypical orders, or as risk-stratification tools. These functions are not interchangeable merely because they share an artificial-intelligence label.

The predicted endpoints also differ. Studies may target statistical outliers, medication errors, pharmacist interventions, adverse drug events, high-risk interactions, or clinician responses to alerts. An unusual order is not necessarily erroneous; an intervention can reflect local practice; and alert acceptance does not demonstrate patient benefit. Medication-alert optimization studies have been concentrated in academic hospitals, with limited external validation and operational implementation [3]. Clinical decision support is therefore better understood as a sociotechnical intervention whose effects depend on data quality, timing, interface design, professional roles, and organizational response [4].

Evaluation methods determine which claims are defensible. Random splitting addresses internal performance but not

Address for correspondence: Ahmed Ben Ali, Department of Hospital Pharmacy and Medication Safety, Faculty of Pharmacy, University of Tunis, Tunis, Tunisia.
ahmed.benali@fpharm.u-tunis.tn
Received: 16 July 2025; **Accepted:** 21 November 2025

How to cite this article: Ben Ali A, Gharbi S, Jebali N. Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods. Arch Pharm Pract. 2025;16(4):54-63. https://doi.org/10.51847/DK74gof42o

temporal drift or geographic transportability. Accuracy may conceal class imbalance, and high discrimination may coexist with poor precision or calibration. Historical labels can reproduce documentation, staffing, or intervention patterns. Likewise, fewer displayed alerts do not necessarily mean less medication-related harm.

This review examined hospital-pharmacy medication-safety algorithms published from 2017 through 2025. It asked which clinical tasks were addressed; how populations, datasets, reference standards, comparators, and outcomes were defined; which validation strategies were used; and how calibration, uncertainty, interpretability, workflow, safety, equity, and reproducibility were evaluated. The purpose was to describe the evidence and derive bounded evaluation priorities without treating publication frequency or technical performance as proof of clinical effectiveness.

Review Design, Questions, and Protocol

Protocol and Registration

The review was designed as a systematic methodological review and reported according to PRISMA 2020 [5]. PRISMA explanation and elaboration guidance informed reporting of eligibility, selection, synthesis, and deviations [6]. The review follows PRISMA 2020 and reports its search using PRISMA-S [5, 7]. Prediction-model studies were appraised across PROBAST domains for participants, predictors, outcomes, analysis, and applicability [8]. A written protocol specified the questions, eligibility criteria, extraction fields, appraisal, and narrative synthesis before final inclusion. No prospective registry record was available; this was retained as a limitation.

Review Questions

The primary question was how hospital-pharmacy medication-safety algorithms were clinically framed and methodologically evaluated. Secondary questions addressed intended task and user, data and reference standard, development and validation, comparator, discrimination and calibration, threshold and uncertainty, interpretability, workflow integration, safety, equity, and reproducibility. Pooled treatment effectiveness, community-pharmacy automation without a hospital safety task, inventory optimization, and disease prediction without a medication-decision function were outside scope.

Eligibility Criteria and Definitions

Eligible empirical reports were peer-reviewed Q1 journal articles published from January 2017 through December 2025, had a verifiable DOI, and evaluated an algorithm relevant to hospital medication safety or pharmacist medication review. Development, internal, temporal, geographic, external, prospective, and implementation evaluations were eligible. Methodology papers supported review methods but were not counted as empirical studies. Non-journal material, commentaries, adjacent topics, and reports lacking extractable task or evaluation data were excluded.

Hospital pharmacy included prescription verification, medication review, reconciliation, monitoring, alert assessment, stewardship-related medication choice, and prevention or detection of medication harm. Eligible algorithms transformed medication, order, laboratory, clinical-text, patient, or clinician-action data into a score, classification, anomaly signal, ranking, or recommendation. Fixed rules were retained only as comparators or hybrid components.

Information Sources and Search

MEDLINE/PubMed, Scopus, and Web of Science Core Collection were searched, supplemented by backward and forward citation checking. Search concepts combined hospital or clinical pharmacy; medication safety, error, adverse drug event, prescription review, alert, and decision support; and artificial intelligence, machine learning, deep learning, natural-language processing, anomaly detection, prediction, validation, workflow, implementation, calibration, uncertainty, and equity. Database-specific subject headings and free-text variants were used. The publication window was 2017–2025. DOI-first deduplication was followed by normalized title, author, and year comparison.

Eligibility, Definitions, and Information Sources

Study Selection and Extraction

Two reviewers independently screened titles, abstracts, and full texts; disagreements were resolved by consensus or adjudication. Each record retained a unique identifier, source, DOI, duplicate status, screening decision, retrieval status, full-text decision, exclusion reason, and synthesis status. Records, reports, and studies were distinguished.

Extraction covered setting, population, intended user, workflow position, algorithm class, input modality, sample size, unit of analysis, outcome prevalence, label construction, reference standard, missing data, comparator, validation, metrics, calibration, thresholds, uncertainty, interpretability, subgroup analysis, prospective evaluation, safety outcomes, implementation, and reproducibility. Development, validation, and analytical choices were separated because they create different risks of bias and applicability [9].

Task, Dataset, Outcome, and Comparator Framework

Medication errors, adverse drug events, preventability, potential harm, and realized harm were treated as distinct constructs [10]. The review-derived task taxonomy comprised prescription-error detection, anomaly detection, pharmacist-review prioritization, alert optimization, high-alert medication screening, adverse-event risk prediction, intervention prediction, dose-error detection, and medication-choice support. Medication-related harm is an important inpatient safety problem, but its prevalence does not validate an algorithmic surrogate [11]. Structured and unstructured inputs were extracted separately because their provenance, missingness, and reference standards differ [12].

Validation was classified as resubstitution, cross-validation, random internal holdout, temporal, geographic, or independent external validation. Outcomes included discrimination, precision, sensitivity, specificity, predictive values, alert volume, intervention yield, workload, process measures, safety, and patient outcomes. Comparators were usual review, no model, fixed rules, existing decision support, conventional statistics, or alternative machine-learning methods.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for eligibility, definitions, and evidence-extraction framework for the review Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods

Table 1. Eligibility, definitions, and evidence-extraction framework for the review Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods.

Review element	Operational definition	Eligibility or decision rule	Data field	Rationale	Bias or ambiguity risk	Planned synthesis use
Setting and task	Hospital medication-use activity involving pharmacy review, monitoring, reconciliation, alerts, stewardship, or harm prevention	Include when output could alter a hospital medication decision or pharmacist workload	Country; service; population; user; workflow stage	Preserves article-specific scope	Hospital care may be reported without meaningful pharmacy involvement	Stratify by setting, user, and task
Algorithm	Learned, statistical, or hybrid method producing a score, class, anomaly, ranking, or recommendation	Fixed rules included only as comparators or hybrid components	Model family; input; output; threshold	Separates evaluated algorithms from automation	"AI" may describe simple rules or regression	Compare technical approaches
Safety construct	Error, adverse event, preventability, potential harm, realized harm, or explicit surrogate [10]	Extract separately; do not merge unlike endpoints	Definition; severity; denominator; assessor	Prevents false equivalence	Intervention or unusualness may be mistaken for harm	Outcome-specific synthesis
Reference standard	Expert review, pharmacist intervention, corrected order, coded event, clinician action, or explicit label	Label construction must be identifiable	Source; blinding; adjudication; agreement	Determines what the model learns [9]	Historical actions may encode local practice	Appraise label credibility
Data and validation	Structured or text data [12]; internal, temporal, geographic, or independent external testing	Random splitting is not external validation	Source; time window; missingness; site; split	Supports leakage and transportability assessment	Local coding and repeated dataset use may inflate performance	Rank validation depth
Comparator and metrics	Usual care, rules, decision support, statistical baseline, or alternative model; discrimination, calibration, and classification measures	Require matched evaluation data and extract prevalence, denominator, and threshold when reported	Comparator; AUROC; AUPRC; sensitivity; specificity; PPV; NPV; calibration	Performance depends on baseline and decision context	AUROC or accuracy may conceal low precision or miscalibration	Compare metric completeness
Workflow and boundaries	Prospective use, workload, safety, subgroup errors, implementation, and reproducibility	Absence of reporting is unknown, not evidence of safety or equity	Study phase; user response; subgroup results; code/data availability	Separates technical output from clinical consequence	Implementation may be confounded; demographic inclusion may be mistaken for fairness	Identify gaps and decision boundaries

Search, Selection, Extraction, Appraisal, and Synthesis

Search and Selection

Results were imported into a record-level log and deduplicated before screening. Machine-assisted methods could support prioritization and auditability but did not replace reviewer judgment [13]. Automation was limited to organizational support; human adjudication governed eligibility decisions [14]. One primary reason was assigned to each full-text exclusion so that PRISMA totals remained reconcilable.

Appraisal and Applicability

PROBAST-informed appraisal considered participant selection, predictor availability, reference-standard validity, sample size, event counts, missing data, leakage, feature selection, overfitting, validation, comparator fairness, threshold specification, calibration, and reporting. Applicability was judged separately for population, hospital setting, task, intended user, and workflow. Prospective or implementation studies were also examined for controls, secular trends, alert exposure, uptake, co-interventions, and outcome ascertainment. Author terms such as "validated," "safe," or "effective" were not accepted without corresponding design evidence.

Synthesis and Subgroup Comparisons

Because tasks, labels, datasets, comparators, and outcomes were heterogeneous, evidence was synthesized narratively and in tables rather than pooled. Studies were grouped by clinical task and compared by data source, reference standard, algorithm class, validation level, comparator, and outcome. Metric selection was interpreted against class imbalance, decision thresholds, and unequal false-positive and false-negative consequences [15]. Retrospective model performance was kept separate from prospective impact, generalizability, and safe deployment [16].

Planned comparisons included adult versus pediatric or neonatal populations; general versus specialist services; prescription-level versus patient-level prediction; structured versus text data; supervised versus unsupervised or hybrid approaches; internal-only versus temporal, geographic, or external validation; and offline versus prospective evaluation. Missing subgroup analysis was recorded as an evidence gap, not equivalent performance.

Protocol Deviations

The absence of prospective registration was the principal protocol limitation. No post-screening change was made to the publication window, hospital-pharmacy scope, empirical eligibility criteria, task taxonomy, or appraisal domains. Quantitative synthesis was not undertaken because incompatible tasks and outcomes had been designated for methodological narrative synthesis. Clarifications introduced during extraction were applied consistently and labelled as review operationalization.

Search and Selection Results

Study-Selection Flow

The searches identified 63 records. Eight duplicates were removed, leaving 55 records for title-and-abstract screening. Fifteen records were excluded at that stage, and 40 reports were sought and retrieved. All 40 reports underwent full-text assessment. Seventeen were excluded: 16 were review, reporting, methodology, or background articles rather than eligible empirical evaluations, and one addressed explainability without evaluating a hospital-pharmacy medication-safety algorithm. Twenty-three reports representing 23 studies entered qualitative synthesis; none entered quantitative synthesis.

The eligible evidence included retrospective outlier screening of medication orders [17]. A smaller implementation-oriented subset evaluated probabilistic inpatient alerting within clinical workflow [18]. Other studies developed risk-prioritization models intended to direct pharmacist review toward prescriptions more likely to contain medication errors [19]. Unsupervised approaches also identified unusual medication patterns without conventional binary labels [20].

Figure 1 presents the completed review-selection pathway in a black-and-white black-and-white prisma 2020 flow diagram for study selection, documenting identification, deduplication, screening, eligibility, exclusions with reasons, and final inclusion using the values approved in the Part 1 PRISMA Record-Accounting Audit.

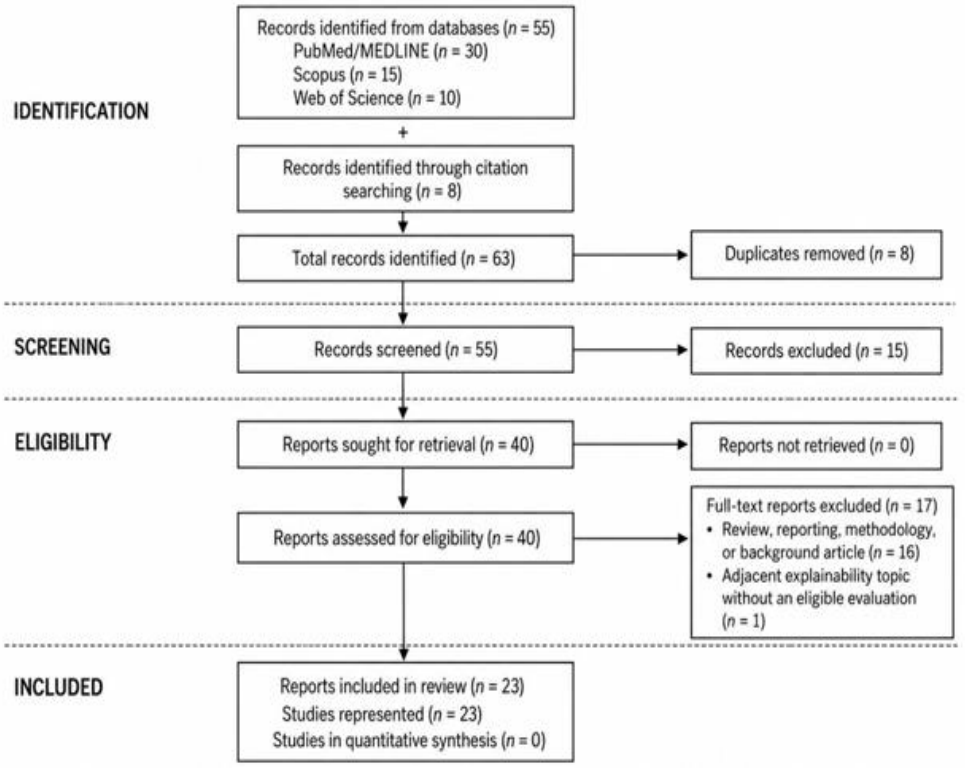

Figure 1. Black-and-White PRISMA 2020 Flow Diagram for Study Selection

Evidence-Base Characteristics

Clinical Tasks Addressed

The evidence base covered prescription-error detection, atypical-order identification, pharmacist-review prioritization, alert optimization, dose-error screening, adverse-event risk prediction, and medication-choice support. Pharmacist-perception research was uncommon but demonstrated that acceptance and perceived usefulness are distinct from technical performance [21]. Some models used historical pharmacist actions as outcome labels, which made the task operationally relevant while risking reproduction of local intervention practices [22].

Hybrid systems combined rule-based drug-utilization review with deep learning to identify errors not captured by fixed rules [23]. Alert-fatigue models commonly predicted clinician responses or estimated alert relevance rather than measuring medication-related harm directly [24]. The unit of analysis varied among prescriptions, orders, alerts, patients, and pharmacist interventions, limiting direct comparison across studies.

Data Sources and Reference Standards

Most models used routinely collected electronic health-record, prescribing, laboratory, medication-administration, or pharmacy-intervention data. Reference standards included pharmacist review, corrected prescriptions, clinician actions, coded adverse events, expert adjudication, or historical intervention records. These standards represented different constructs and were not treated as interchangeable.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for characteristics and classification of the evidence included in the review Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods.

Table 2. Characteristics and classification of the evidence included in the review Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods.

Evidence category	Study or source design	Population or setting	AI system or intervention	Comparator or reference standard	Outcome or construct	Validation level	Key limitation
Medication-order outlier detection	Retrospective evaluation	Hospital prescriptions	Statistical or unsupervised anomaly detection	Expert or pharmacist assessment	Potential prescribing error or atypicality [17]	Internal	Outlier status is not equivalent to clinical error
Inpatient alerting	Retrospective development with workflow evaluation	Inpatient medication orders	Probabilistic decision support	Existing review or observed alert response	Error interception and adverse-event prevention process [18]	Internal and implementation-oriented	Uncontrolled workflow effects may reflect concurrent changes
Pharmacist-review prioritization	Retrospective prediction	Hospital prescription review	Supervised risk ranking	Pharmacist-identified medication error	High-risk prescription [19]	Internal	Local review practices shape labels
Unsupervised error detection	Retrospective development	Medication-order datasets	Unsupervised learning	Expert-labelled or rule-derived unusual orders	Potential medication error [20]	Internal	Thresholds may not represent calibrated risk
Human-factors evaluation	Mixed-methods perception study	Hospital pharmacists	Atypical-order identification tool	Pharmacist perception and usability assessment	Acceptability and perceived usefulness [21]	Local user evaluation	Perception does not establish safety or effectiveness
Intervention prediction	Retrospective prediction	Inpatient pharmacy orders	Machine-learning classification	Historical pharmacist actions	Probability of pharmacist intervention [22]	Internal	Intervention is a practice-dependent proxy
Hybrid rule and learning systems	Retrospective model evaluation	Pediatric hospital prescriptions	Rule-based screening plus deep learning	Conventional drug-utilization review	Prescription-error prediction [23]	Internal	Added yield may not translate into patient benefit
Alert-fatigue reduction	Retrospective development and validation	Disease–medication alerts	Machine-learning filtering	Clinician response or alert relevance	Suppressed or prioritized alerts [24]	Internal	Response prediction is not a direct safety outcome

Internal, Temporal, Geographic, and External Validation

Large internal datasets supported medication-alert filtering, but internal holdout performance did not establish transportability across hospitals [25]. Prospective daily-practice evaluation provided stronger evidence about whether low-priority prescriptions could be excluded from pharmacist review, although findings remained dependent on the local workflow and case mix [26]. Neonatal-intensive-care evaluation illustrated both the value of population-specific

modelling and the restricted generalizability of specialist models [27].

Comparators included conventional statistics, fixed rules, existing decision support, alternative machine-learning algorithms, usual pharmacist review, or no active comparator. Random holdout and cross-validation predominated. Temporal and geographic testing were less frequent, and clearly independent external validation was uncommon. Reported metrics emphasized discrimination, sensitivity, specificity, accuracy, predictive values, or alert yield; metric sets were not sufficiently uniform for pooling.

Calibration, Uncertainty, and Interpretability

Unsupervised overdose and underdose detection produced anomaly scores whose thresholds required clinical interpretation and could not be assumed to represent calibrated probabilities of error [28]. Adverse-drug-reaction prediction studies compared neural networks with conventional models, but discrimination was more consistently described than calibration or decision consequences [29]. Comparisons between machine learning and conventional statistical methods were most informative when reference standards, thresholds, and calibration were reported together [30].

Prospective workflow evidence remained limited. Several systems were technically integrated or tested in daily practice, but few evaluations used concurrent controls or designs capable of separating algorithm effects from staffing, learning, secular trends, or associated process changes. Interpretability was usually feature-based or score-based rather than evaluated as an

Safety, Equity, and Error Analysis

Older-inpatient prediction studies demonstrated the importance of population-specific error analysis and the limits of transferring results to other age groups or services [31]. Self-intercepted medication orders provided pragmatic labels but omitted errors that prescribers did not recognize or correct [32]. Antibiotic-resistance prediction illustrated a medication-choice task in which false-positive and false-negative errors could affect both individual treatment and stewardship goals [33].

Few studies reported subgroup calibration, differential false-negative rates, or error consequences across demographic or clinical groups. Inclusion of diverse patients was therefore not treated as evidence of equitable performance. Code availability, executable models, complete feature definitions, and reusable preprocessing procedures were also inconsistent.

Table 3 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for methodological quality, applicability, and evidence gaps in the review Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods.

Table 3. Methodological quality, applicability, and evidence gaps in the review Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods.

Appraisal domain	Observed evidence pattern	Methodological concern	Applicability consequence	Direction of potential bias	Evidence needed	Interpretive boundary
Participants and setting	Predominantly local hospital cohorts	Convenience sampling and restricted case mix	Limited transportability	Performance may be optimistic in similar local data	Multicentre and geographic validation	Local accuracy does not establish generalizability
Labels and reference standards	Pharmacist actions, corrected orders, expert review, or coded events	Labels may encode documentation and practice behavior [22, 32]	Model may reproduce local decisions rather than errors	Apparent agreement may be inflated	Blinded adjudication and agreement reporting	Predicted intervention is not synonymous with harm prevention
Analysis	Cross-validation and random holdout were common	Leakage, repeated tuning, imbalance, and threshold selection	Estimates may not persist prospectively	Discrimination may be overstated	Locked models and temporally separated testing	Internal validation is not external validation
Calibration and uncertainty	Incompletely reported	AUROC may coexist with poor precision or miscalibration [28, 29]	Thresholds may not support safe triage	Rare events can inflate negative performance	Calibration plots, recalibration, and decision analysis	A score is not automatically a decision probability
Workflow evaluation	Limited prospective or controlled comparison	Uptake, workload, and secular effects may confound findings [26]	Operational benefit remains uncertain	Before–after effects may favor implementation	Concurrent controls and prespecified workflow outcomes	Implementation does not prove validated benefit
Safety and equity	Sparse subgroup and error-consequence analysis	Unequal errors may remain hidden [31, 33]	Use outside studied groups is uncertain	Aggregate performance can mask subgroup harm	Subgroup calibration and consequence-weighted errors	Absence of reported disparity is not evidence of equity

Reproducibility	Inconsistent code, preprocessing, and model access	Independent replication is constrained	Findings may depend on undisclosed local choices	Selective reporting may favor positive results	Executable code, data dictionaries, and external replication	Publication does not establish reproducibility

Discussion: Principal Findings and Interpretation

Principal Methodological Findings

The literature was technically diverse but methodologically concentrated in retrospective, single-centre evaluation. High-alert medication screening showed how specialization could improve apparent task performance while narrowing applicability [34]. Adverse-event prediction demonstrated that high discrimination could coexist with low precision when outcomes were uncommon [35]. Operational systems provided alert-yield and intercepted-error information not captured by AUROC alone, but these process outcomes did not establish patient benefit [36].

The principal evidence boundary was between predicting a local label and improving medication decisions. A model may reproduce pharmacist interventions, detect atypicality, or reduce displayed alerts without demonstrating fewer preventable adverse events. Independent validation, calibration, threshold analysis, prospective comparison, and subgroup error assessment were too limited to support universal clinical or implementation claims.

Figure 2 presents the original conceptual synthesis for methodological evaluation map for hospital-pharmacy medication-safety algorithms, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

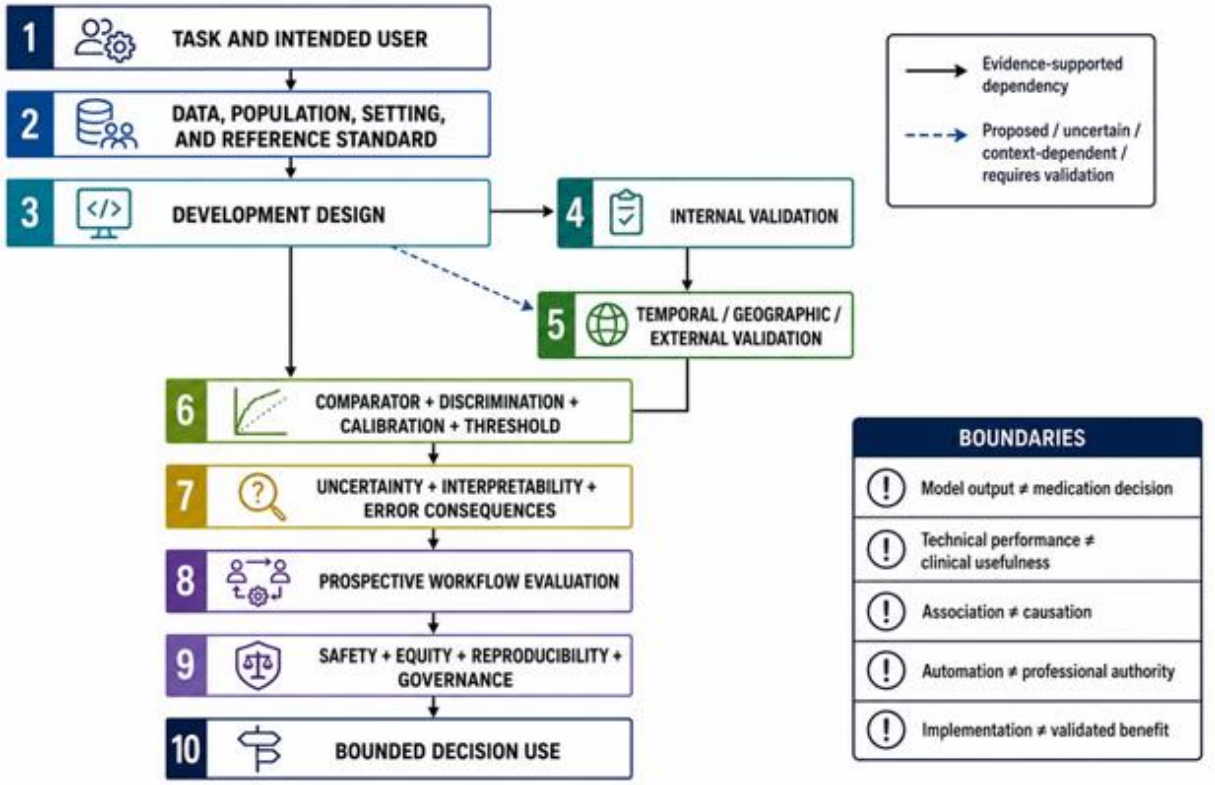

Figure 2. Methodological Evaluation Map for Hospital-Pharmacy Medication-Safety Algorithms

Consequences for Evidence Interpretation and Minimum Evaluation Requirements

Minimum credible evaluation should define the intended medication decision, intended user, reference standard, and error consequences before model development. It should use temporally or geographically independent data where feasible, report calibration and threshold-specific predictive values, compare against current review processes, and assess workload and missed-error consequences prospectively. Subgroup analyses should be prespecified when performance differences could alter safety or access.

Implications, Strengths, and Limitations

Recent medication-error risk-stratification models reinforce the need to examine local prevalence and independent validation before interpreting performance [37]. Pharmacist-led surgical-prescription optimization combined external testing with a prospective implementation phase, but before–

after outcome changes remained susceptible to confounding [38]. Very large prescription-review datasets produced precise local estimates while leaving geographic transportability, prospective utility, and equity unresolved [39].

The review's strengths were its task-specific scope, record-level accounting, explicit separation of reference standards and safety outcomes, and structured appraisal of validation and applicability. Limitations included heterogeneity that precluded quantitative synthesis, incomplete reporting within primary studies, variable definitions of medication error, and the absence of prospective protocol registration. Restriction to Q1 journal articles may also have excluded relevant evidence from other publication settings.

Table 4 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for integrated synthesis, uncertainty, and decision implications for the review Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods.

Table 4. Integrated synthesis, uncertainty, and decision implications for the review Medication-Safety Algorithms in Hospital Pharmacy between 2017 and 2025: A Systematic Review of Clinical Tasks and Evaluation Methods.

Synthesis domain	Evidence supporting the finding	Conflicting or absent evidence	Heterogeneity or overlap issue	Confidence boundary	Practice implication	Research priority
Task definition	Algorithms addressed identifiable pharmacy and medication-safety tasks [17, 19, 23]	Patient-safety outcomes were uncommon	Error, atypicality, intervention, and alert response overlap imperfectly	Moderate confidence in task diversity; low confidence in comparative effectiveness	Match outputs to a defined pharmacist decision	Consensus task and outcome definitions
Validation	Internal testing predominated [25, 27]	Independent external validation was uncommon	Validation terminology varied	Local performance cannot establish portability	Avoid transferring thresholds without recalibration	Multicentre temporal and geographic validation
Metrics and calibration	Discrimination was widely used [29, 30]	Calibration and decision analysis were incomplete	Prevalence and thresholds differed	AUROC alone is insufficient	Examine PPV, NPV, calibration, and workload at intended thresholds	Standard metric sets tied to clinical consequences
Workflow	Some daily-practice and implementation evaluations were reported [18, 26, 36]	Controlled prospective comparisons were sparse	Uptake and co-interventions differed	Process improvement is not equivalent to patient benefit	Preserve pharmacist override and escalation pathways	Prospective comparative workflow studies
Safety and equity	Specialized populations and error consequences were identifiable [31, 33]	Subgroup calibration and differential errors were rarely reported	Populations and harm definitions varied	Aggregate results cannot establish equitable safety	Limit use to evaluated populations and monitor errors	Consequence-weighted subgroup analysis
Reproducibility and governance	Large datasets and operational integration were feasible [38, 39]	Code, preprocessing, and independent replication were inconsistent	Local systems and labels limit reuse	Implementation is context-dependent	Require monitoring, documentation, and change control	Reproducible pipelines and post-implementation surveillance

Conclusion

Hospital-pharmacy medication-safety algorithms addressed clinically relevant tasks, including prescription-error detection, pharmacist-review prioritization, atypical-order identification, alert optimization, and medication-related risk prediction. Their evaluation, however, remained dominated by retrospective and frequently single-centre designs using heterogeneous labels, reference standards, comparators, and metrics.

The principal methodological weakness was not a lack of technical models but an incomplete connection between model performance and decision-relevant evidence. Internal discrimination could not establish transportability; alert reduction could not establish reduced harm; historical pharmacist actions could not serve automatically as unbiased error labels; and implementation without controlled evaluation could not establish causation.

Future studies should define the medication decision and error consequences prospectively, validate models temporally and geographically, report calibration and threshold-specific performance, compare systems with existing pharmacist workflows, evaluate missed-error and workload consequences, and examine subgroup performance. Reproducible reporting and governance are necessary throughout model updating and operational use.

The evidence therefore supports cautious methodological development rather than broad claims of effectiveness, safety, equity, regulatory acceptance, or deployment readiness. Hospital-pharmacy algorithms should be interpreted as context-dependent decision-support candidates whose clinical value requires independent, prospective, and consequence-aware evaluation.

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

References

1. Hatzimanolis J, Riley B, El-Den S, Aslani P, Zhou J, Chaar BB. Applications of artificial intelligence in current pharmacy practice: A scoping review. Res Social Adm Pharm. 2025;21(3):134-41. doi:10.1016/j.sapharm.2024.12.007
2. Ranchon F, Chanoine S, Lambert-Lacroix S, Bosson JL, Moreau-Gaudry A, Bedouch P. Development of artificial intelligence powered apps and tools for clinical pharmacy services: A systematic review. Int J Med Inform. 2023;172:104983. doi:10.1016/j.ijmedinf.2022.104983
3. Graafsma J, Murphy RM, van de Garde EMW, Karapinar-Çarkit F, Derijks HJ, Hoge RHL, et al. Artificial intelligence to optimize medication alerts generated by clinical decision support systems: A scoping review. J Am Med Inform Assoc. 2024;31(6):1411-22. doi:10.1093/jamia/ocae076
4. 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;3:17. doi:10.1038/s41746-020-0221-y
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. doi:10.1136/bmj.n71
6. 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. doi:10.1136/bmj.n160
7. 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
8. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-8. doi:10.7326/M18-1376
9. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. Ann Intern Med. 2019;170(1). doi:10.7326/M18-1377
10. Laatikainen O, Miettunen J, Sneck S, Lehtiniemi H, Tenhunen O, Turpeinen M. Medication-related adverse events in health care—what have we learned? A narrative overview of the current knowledge. Eur J Clin Pharmacol. 2022;78(2):159-70. doi:10.1007/s00228-021-03213-x
11. Bates DW, Levine DM, Salmasian H, Syrowatka A, Shahian DM, Lipsitz S, et al. The safety of inpatient health care. N Engl J Med. 2023;388(2):142-53. doi:10.1056/NEJMsa2206117
12. Wong A, Plasek JM, Montecalvo SP, Zhou L. Natural language processing and its implications for the future of medication safety: A narrative review of recent advances and challenges. Pharmacotherapy. 2018;38(8):822-41. doi:10.1002/phar.2151
13. van de Schoot R, de Bruin J, Schram R, Zahedi P, de Boer J, Weijdema F, et al. An open source machine learning framework for efficient and transparent systematic reviews. Nat Mach Intell. 2021;3(2):125-33. doi:10.1038/s42256-020-00287-7
14. Marshall IJ, Wallace BC. Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. Syst Rev. 2019;8(1):163. doi:10.1186/s13643-019-1074-9
15. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep. 2022;12(1):5979. doi:10.1038/s41598-022-09954-8
16. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
17. Schiff GD, Volk LA, Volodarskaya M, Williams DH, Walsh L, Myers SG, et al. Screening for medication errors using an outlier detection system. J Am Med Inform Assoc. 2017;24(2):281-7. doi:10.1093/jamia/ocw171
18. Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting. J Am Med Inform Assoc. 2019;26(12):1560-5. doi:10.1093/jamia/ocz135
19. Corny J, Rajkumar A, Martin O, Dode X, Lajonchère JP, Billuart O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. J Am Med Inform Assoc. 2020;27(11):1688-94. doi:10.1093/jamia/ocaa154
20. Santos HDP, Ulbrich AHDPS, Woloszyn V, Vieira R. DDC-Outlier: Preventing medication errors using unsupervised learning. IEEE J Biomed Health Inform. 2019;23(2):874-81. doi:10.1109/JBHI.2018.2828028
21. Hogue SC, Chen F, Brassard G, Lebel D, Bussières JF, Thibault M. Pharmacists' perceptions of a machine learning model for the identification of atypical medication orders. J Am Med Inform Assoc. 2021;28(8):1712-8. doi:10.1093/jamia/ocab071
22. Balestra M, Chen J, Iturrate E, Aphinyanaphongs Y, Nov O. Predicting inpatient pharmacy order interventions using provider action data. JAMIA Open. 2021;4(3). doi:10.1093/jamiaopen/ooab083
23. Lee S, Shin J, Kim HS, Lee MJ, Yoon JM, Lee S, et al. Hybrid method incorporating a rule-based approach and deep learning for prescription error prediction. Drug Saf. 2022;45(1):27-35. doi:10.1007/s40264-021-01123-6
24. Poly TN, Islam MM, Muhtar MS, Yang HC, Nguyen PAA, Li YJ. Machine learning approach to reduce alert fatigue using a disease medication-related clinical decision support system: Model development and validation. JMIR Med Inform. 2020;8(11). doi:10.2196/19489
25. Liu S, Kawamoto K, Del Fiol G, Fiszman M, Weir C, Reese T, et al. The potential for leveraging machine learning to filter medication alerts. J Am Med Inform Assoc. 2022;29(5):891-9. doi:10.1093/jamia/ocab292
26. Levivien C, Cavagna P, Grah A, Buronfosse A, Courseau R, Bézie Y, et al. Assessment of a hybrid decision support system using machine learning with artificial intelligence to safely rule out prescriptions from medication review in daily practice. Int J Clin Pharm. 2022;44(2):459-65. doi:10.1007/s11096-021-01366-4
27. Yalçın N, Kaşıkcı M, Çelik HT, Allegaert K, Demirkan K, Yiğit Ş, et al. Development and validation of a machine learning-based detection system to improve precision screening for medication errors in the neonatal intensive care unit. Front Pharmacol. 2023;14:1151560. doi:10.3389/fphar.2023.1151560
28. Nagata K, Nakamura M, Yamamoto N, Tsuji T, Watanabe H, Ishikawa T, et al. Detection of overdose and underdose prescriptions—an unsupervised machine learning approach. PLoS One. 2021;16(11). doi:10.1371/journal.pone.0260315
29. Imai S, Takekuma Y, Kashiwagi H, Miyai T, Kobayashi M, Iseki K, et al. Validation of the usefulness of artificial neural networks for risk prediction of adverse drug reactions used for individual patients in clinical practice. PLoS One. 2020;15(7). doi:10.1371/journal.pone.0236789
30. Van Laere S, Muylle KM, Dupont AG, Cornu P. Machine learning techniques outperform conventional statistical methods in the prediction of high risk QTc prolongation related to a drug-drug interaction. J Med Syst. 2022;46(12):100. doi:10.1007/s10916-022-01890-4

31. Hu Q, Wu B, Wu J, Xu T. Predicting adverse drug events in older inpatients: A machine learning study. Int J Clin Pharm. 2022;44(6):1304-11. doi:10.1007/s11096-022-01468-7
32. King CR, Abraham J, Fritz BA, Cui Z, Galanter W, Chen Y, et al. Predicting self-intercepted medication ordering errors using machine learning. PLoS One. 2021;16(7). doi:10.1371/journal.pone.0254358
33. Lewin-Epstein O, Baruch S, Hadany L, Stein GY, Obolski U. Predicting antibiotic resistance in hospitalized patients by applying machine learning to electronic medical records. Clin Infect Dis. 2021;72(11). doi:10.1093/cid/ciaa1576
34. Wongyikul P, Thongyot N, Tantrakoolcharoen P, Seephueng P, Khumrin P. High alert drugs screening using gradient boosting classifier. Sci Rep. 2021;11(1):20132. doi:10.1038/s41598-021-99505-4
35. Langenberger B. Machine learning as a tool to identify inpatients who are not at risk of adverse drug events in a large dataset of a tertiary care hospital in the USA. Br J Clin Pharmacol. 2023;89(12):3523-38. doi:10.1111/bcp.15846
36. Chen CY, Chen YL, Scholl J, Yang HC, Li YCJ. MedGuard: A machine learning-based medication error detection system for improving medication safety in hospitals. Comput Methods Programs Biomed. 2024;243:107869. doi:10.1016/j.cmpb.2023.107869
37. Abdo A, Gallay L, Vallecillo T, Clarenne J, Quillet P, Vuiblet V, et al. A machine learning-based clinical predictive tool to identify patients at high risk of medication errors. Sci Rep. 2024;14(1):32022. doi:10.1038/s41598-024-83631-w
38. Li X, Yue X, Zhang L, Zheng X, Shang N. Pharmacist-led surgical medicines prescription optimization and prediction service improves patient outcomes—a machine learning based study. Front Pharmacol. 2025;16:1534552. doi:10.3389/fphar.2025.1534552
39. Johns E, Guendouz A, Dal Mas L, Beck M, Alkanj A, Gourieux B, et al. Using machine learning to predict pharmaceutical interventions during medication prescription review in a hospital setting. Am J Health Syst Pharm. 2025;82(22):1238-48. doi:10.1093/ajhp/zxaf089