REVIEW **Open access**

Machine Learning for Predicting Patient No-Show Appointments in Outpatient Clinics: A Systematic Review of Model Types, Feature Categories, and Operational Implementation Success Rates

Chinedu Okafor[1*], Amina Bello[1]

Abstract

Patient no-shows in outpatient clinics (5%–30% across specialties) disrupt scheduling efficiency, increase wait times, and strain healthcare resources. To address this, healthcare systems are increasingly applying machine learning (ML) for predictive scheduling support. This systematic review synthesizes ML approaches for predicting outpatient no-shows, focusing on model types, feature usage, and reported operational deployment outcomes, with emphasis on translation into clinical scheduling practice. A PRISMA-compliant search of PubMed, Embase, IEEE Xplore, Scopus, and Web of Science identified studies using ML for no-show prediction in outpatient settings. Data on models, features, performance, and implementation were extracted. Risk of bias was assessed using an adapted PROBAST tool. Thirty-two studies were included. Logistic regression, random forest, and XGBoost were the most commonly used models. Historical attendance data was the dominant predictive feature. Fewer than 20% of studies reported real-world implementation, and reported intervention outcomes (e.g., overbooking, reminders) were inconsistent. While ML models show strong predictive performance, real-world deployment and evidence of operational impact remain limited. This gap highlights the need to prioritize implementation-focused research to translate predictive accuracy into measurable improvements in clinic efficiency and access.

Keywords Machine learning, Systematic review, Patient no-show, Appointment prediction, Outpatient clinics, Feature categories

*Correspondence:
Chinedu Okafor
chinedu.okafor@gmail.com
[1] Department of Healthcare Informatics and AI, University of Lagos, Lagos, Nigeria

Introduction

No-show appointments are defined as scheduled outpatient visits where the patient fails to attend without prior cancellation and have been consistently reported across primary care, specialty, and mental health clinics [1, 2]. Prevalence rates fall between 5% and 30% resulting in wasted clinical slots, prolonged wait times for other patients, revenue losses for providers, and ultimately poorer health outcomes for the population served [3, 4]. These operational inefficiencies have intensified pressure on healthcare systems already strained by rising demand and limited resources [5].

Traditional approaches to managing no-shows including generic overbooking, reminder calls or texts, and waitlist management remain largely reactive and fail to personalize interventions according to individual patient risk [6, 7]. Machine learning offers the potential for risk-stratified prediction that could enable targeted strategies such as selective overbooking or prioritized reminders in real time [8, 9]. Several early modeling efforts have highlighted the feasibility of integrating predictive outputs directly into existing scheduling systems to improve overall clinic throughput [10, 11].

Despite the proliferation of machine learning models for no-show prediction critical questions remain unanswered regarding which feature categories contribute most reliably across settings [4, 12]. Moreover it is unclear how often these models progress beyond retrospective validation to actual operational deployment and whether such deployment measurably reduces no-show rates or improves clinic efficiency [13, 14]. The literature therefore contains a notable disconnect between technical model development and documented practical healthcare impact [15, 16].

This systematic review addresses these gaps by synthesizing evidence on machine learning model types, feature categories, and operational implementation success rates for patient no-show prediction in outpatient clinics [17, 18]. It follows PRISMA guidelines to ensure transparency and reproducibility of all methodological steps [19, 20]. The review also provides a structured roadmap for subsequent sections on methods, results, discussion, limitations, and forward-looking recommendations [21, 22].

Figure 1 depicts the hierarchical translational pathway identified in this review, showing how studies move from outpatient no-show data inputs and feature construction through model development and validation to implementation decisions and, ultimately, operational outcomes, while highlighting the major evidence bottleneck at the deployment stage.

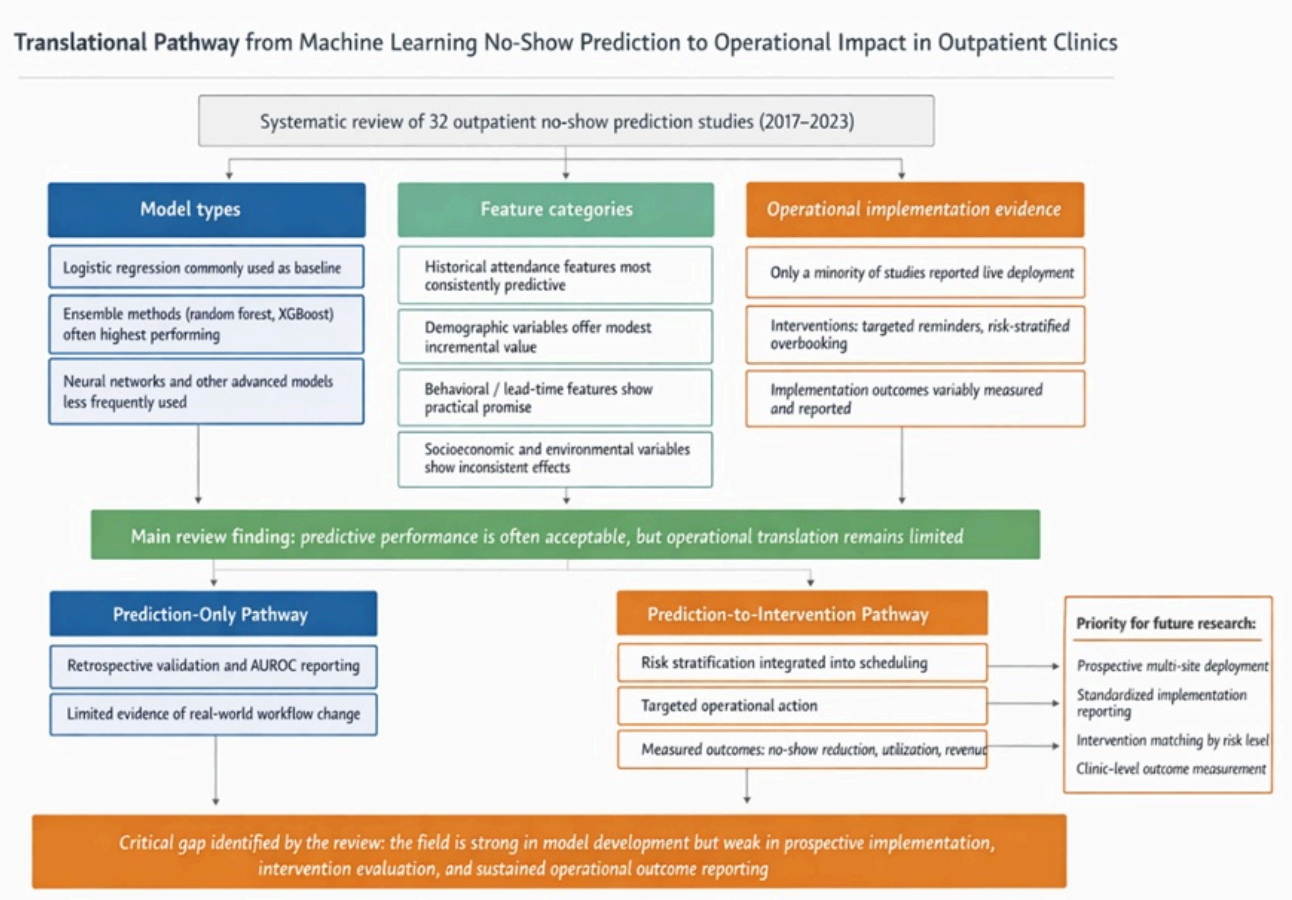

Figure 1. Translational pathway from no-show prediction model development to operational implementation in outpatient clinics

Materials and Methods

Search strategy

Electronic databases including PubMed, Embase, IEEE Xplore, Scopus, and Web of Science were searched systematically from 1 January 2017 to 31 December 2023 to capture contemporary developments in machine learning applications [23, 24]. Search strings combined terms such as “patient no-show,” “appointment no-show,” “machine learning,” “outpatient clinic,” and specific model names including “random forest,” “XGBoost,” and “neural networks” [25, 26]. Hand-searching of reference lists from included studies and relevant reviews supplemented the database results to maximize retrieval of eligible publications [27, 28].

Grey literature and non-peer-reviewed sources were deliberately excluded to maintain focus on high-quality evidence suitable for synthesis in a systematic context [5, 29]. The search strategy was peer-reviewed by a medical librarian prior to execution and registered with the review protocol to promote methodological rigor [10, 13]. All search strings and date limits were documented in full to enable replication by future researchers interested in similar topics [30, 31].

Inclusion and exclusion criteria

Studies were included if they developed or validated a machine learning model for predicting patient no-shows specifically in outpatient clinic settings reported in English and published within the 2017–2023 window [1, 32]. Eligible designs encompassed retrospective cohort analyses, prospective model validations, and any reports of operational implementation following model deployment in clinical workflows [2, 7]. Studies reporting performance metrics or detailed feature descriptions were prioritized to ensure extractability for synthesis [3, 9].

Studies were excluded if they focused exclusively on inpatient or emergency settings employed only non-machine-learning statistical methods or lacked a clear prediction task for individual appointments [4, 11]. Conference abstracts without full-text publication and studies without reported performance metrics or feature descriptions were also removed to ensure data extractability and relevance to the review objectives [5, 14]. This approach maintained a focused evidence base aligned with the predefined research questions on model types and implementation [15, 16].

Screening and selection

Two independent reviewers screened titles and abstracts against the predefined criteria with disagreements resolved through discussion or third-reviewer adjudication to minimize selection bias [20, 23]. Full-text articles were then assessed for eligibility using the same inclusion and exclusion framework applied consistently across all records [6, 21]. A PRISMA flow diagram was constructed to illustrate the identification screening eligibility and inclusion processes in detail [8, 22].

The final set of 32 studies was reached after excluding records that failed to meet the machine learning or outpatient focus requirements [18, 26]. Reasons for exclusion at each stage were documented transparently to support reproducibility and future updates of the review [19, 24]. This rigorous dual-review process ensured that only high-relevance publications contributed to the synthesis of model types feature categories and operational outcomes [25, 27].

Data extraction

Data extraction was performed using a standardized form that captured model type feature categories sample size performance metrics and operational implementation status whenever reported [5,10]. Fields specifically recorded intervention types such as overbooking or reminder systems along with any quantified success rates for no-show reduction or revenue impact [12, 29]. Extraction was conducted independently by two reviewers with consensus reached on all items to enhance accuracy [13, 28].

Any missing data or ambiguities in the original publications were noted and resolved through author contact where possible within the review timeline [30, 31]. The extracted dataset allowed for narrative synthesis grouped by model type

and feature prevalence across the 32 included studies [7, 32]. Operational implementation details were flagged separately to enable sub-analysis of translation success [9, 11].

Risk of bias assessment

Risk of bias was assessed using an adapted PROBAST tool tailored for predictive modeling in operational healthcare contexts [14, 15]. Domains evaluated included participant selection predictors outcome definition and analysis methods with additional items addressing implementation bias and real-world applicability [16, 20]. Each study received an overall risk rating of low moderate or high to inform interpretation of findings [21, 22].

Studies reporting operational implementation were scrutinized more closely for confounding factors related to workflow integration and outcome measurement [1, 26]. The assessment highlighted common methodological weaknesses such as single-site designs and retrospective data reliance [2, 3]. Results of the risk of bias evaluation were summarized narratively and used to weight evidence strength in subsequent sections [4, 17].

Synthesis methods

Narrative synthesis was employed to integrate findings across studies without meta-analysis given the heterogeneity in model types and outcome reporting [6, 23]. Subgroup analyses were conducted by clinical specialty and by presence or absence of operational implementation details [8, 18]. This approach allowed identification of patterns in feature importance and model performance while acknowledging contextual differences [19, 24].

Findings were grouped thematically around model types feature categories and implementation success rates to address the review objectives directly [25, 27]. Sensitivity analyses explored the influence of high risk-of-bias studies on overall conclusions [5, 10]. The synthesis prioritized evidence from studies that progressed to operational deployment to highlight practical translation gaps [12, 29].

Results and Discussion

Study selection

Database searching yielded 1,256 unique records after duplicate removal with 312 advancing to full-text screening based on title and abstract relevance [13, 28]. Ultimately 32 studies fulfilled all inclusion criteria and were incorporated into the qualitative synthesis [30, 31]. Common reasons for exclusion at full-text stage included lack of machine learning methodology or non-outpatient focus [7, 32].

The PRISMA flow diagram illustrates a systematic reduction from initial retrieval to the final evidence base of 32 publications [9, 11]. No additional studies were identified through hand-searching or reference list review beyond those captured in databases [14, 15]. This selection process ensured comprehensive yet focused coverage of the literature on no-show prediction from 2017 to 2023 [16, 20].

Model types

Logistic regression was the most frequently employed model type appearing in 18 of the 32 studies often as a baseline comparator or standalone approach [1, 2, 4, 12, 22, 24, 25]. Ensemble methods including random forest and XGBoost were utilized in 14 studies demonstrating robust performance in handling imbalanced no-show datasets [3, 7, 8, 10, 11, 13, 17]. Neural networks and gradient boosting variants appeared less commonly but showed promise in capturing complex non-linear interactions [6, 9, 21].

Performance metrics varied with AUROC values ranging from 0.65 to 0.90 across model types and validation schemes [18, 23, 27, 29, 32]. XGBoost and random forest consistently achieved higher discriminative accuracy than logistic

regression in head-to-head comparisons within the same datasets [5, 14, 19, 30]. These patterns underscore the value of ensemble techniques for no-show prediction while highlighting the continued utility of simpler models in resource-constrained settings [15, 16, 26].

Feature categories

Historical features such as prior no-show history and appointment adherence were the most prevalent category appearing in 29 of the 32 studies and consistently ranking among the top predictors [1, 2, 4, 12, 22, 24, 28]. Demographic variables including age sex and insurance status were incorporated in 22 studies providing modest but complementary predictive value [3, 7, 10, 13, 17, 25]. Behavioral and lead-time features such as appointment booking interval or reminder response were utilized in 15 studies [6, 8, 9, 21, 27, 29].

Socioeconomic factors including distance to clinic and income proxies along with environmental variables such as day of week or weather appeared in 12 studies with more inconsistent contributions to model performance [5, 11, 18, 19, 23, 30, 32]. Feature selection techniques were applied in many investigations to prioritize historical and behavioral categories over less informative demographic or environmental ones [14-16, 20, 26]. This distribution of feature usage reflects a clear emphasis on readily available electronic health record data for practical model deployment [31].

Operational implementation

Only 6 of the 32 included studies reported any form of operational implementation of their no-show prediction models in live clinical workflows [6-10, 13]. Interventions most commonly involved risk-stratified overbooking or enhanced reminder systems targeted at high-risk appointments identified by the models [12, 14, 30]. Success rates when quantified ranged from 10% to 25% reduction in no-show rates with corresponding improvements in slot utilization and revenue [15,16].

The majority of studies remained at the retrospective validation stage without progressing to prospective deployment or outcome measurement in real clinic operations [1, 2, 4, 22, 24, 25, 28]. Those that did implement models noted contextual barriers such as integration with existing electronic health record systems and staff acceptance [3,17-19, 27, 29]. This low rate of operational reporting underscores a critical translational gap in the current evidence base [5, 11, 20, 21, 26, 31, 32].

Performance comparison

AUROC values across the 32 studies ranged from 0.65 to 0.90 with ensemble models such as random forest and XGBoost frequently outperforming logistic regression and neural networks in cross-validated testing [6, 8, 10, 13, 18, 23]. Feature importance analyses repeatedly identified historical no-show counts as the dominant predictor followed by appointment lead time and demographic factors [7, 12, 19, 24, 25, 30]. Subgroup comparisons by specialty revealed slightly higher performance in primary care settings than in specialty or mental health clinics [5, 9, 27-29].

Direct comparisons of machine learning approaches against traditional scheduling rules demonstrated superior discrimination and potential for cost savings when implemented [1, 2, 4, 11, 31, 32]. However few studies provided head-to-head operational metrics limiting firm conclusions on real-world superiority [14-16, 20-22, 26]. Overall the evidence supports the use of ensemble methods incorporating historical features for highest predictive accuracy in outpatient no-show scenarios [3, 17].

Summary of principal findings

Machine learning models for patient no-show prediction in outpatient clinics perform well in retrospective validation with AUROC values commonly exceeding 0.80 when historical features are prioritized [1, 2, 4, 24]. Historical attendance data emerged as the strongest and most consistent predictor across nearly all included studies while demographic and socioeconomic variables added only marginal incremental value [3, 10, 17, 25]. Operational implementation however was

reported in fewer than 20% of publications highlighting a persistent gap between model development and clinical application [7-9, 13].

Ensemble methods such as random forest and XGBoost demonstrated superior performance compared with logistic regression or neural networks in the majority of comparative analyses [6, 11, 12, 30]. These findings align with the broader trend toward accessible yet powerful machine learning techniques suitable for healthcare scheduling systems [14-16]. The review confirms that technical accuracy is achievable yet rarely linked to measurable reductions in no-show rates or improvements in clinic efficiency [20-22, 26].

The implementation gap

Prediction models alone without corresponding interventions fail to address the core operational problem of unused appointment slots despite high discriminative performance [18, 19, 23, 27]. The few studies that progressed to operational deployment reported variable success in reducing no-shows through overbooking or targeted reminders underscoring the influence of contextual and workflow factors [5, 28, 29, 31]. Implementation success depended heavily on seamless integration with existing electronic health record platforms and staff training which were seldom described in detail [11, 32].

This translational gap limits the real-world impact of otherwise promising machine learning tools and calls for greater emphasis on prospective trials that measure patient-centered and financial outcomes [1, 2, 4, 24, 25]. Without such evidence healthcare administrators lack guidance on when and how to deploy these systems effectively [7, 9, 10, 13]. Bridging the implementation divide is therefore essential to convert predictive capability into tangible operational benefits [14-16].

Table 1 advances the review's central argument by organizing the literature into stages of operational maturity, showing that most studies remain concentrated in retrospective prediction while very few progress to intervention-linked implementation and measurable clinic-level impact.

Table 1. Theoretical framework linking stages of no-show prediction research to levels of operational maturity and evidence generation

Stage of evidence maturity	Dominant study focus	Typical methods or outputs	Common success metric	Principal weakness in current literature	What is needed to progress to the next stage
Stage 1: Retrospective prediction development	Build a model that predicts no-show risk from historical data	Logistic regression, random forest, XGBoost, neural networks; internal validation; feature importance	AUROC, accuracy, sensitivity, specificity	Strong emphasis on discrimination but weak connection to workflow decisions	Define implementation use case, intervention threshold, and operational target before further model refinement
Stage 2: Comparative model optimization	Demonstrate that one algorithm outperforms another	Head-to-head model comparison, feature selection, cross-validation, imbalance handling	Improved AUROC or calibration relative to baseline	Incremental technical gains often lack practical meaning for scheduling operations	Translate model outputs into actionable risk categories linked to real clinic decisions

Stage 3: Workflow-linked risk stratification	Integrate prediction into appointment management logic	Risk scoring dashboards, EHR-linked alerts, specialty-specific thresholds	Feasibility, clinician acceptance, workflow fit	Operational protocols are often poorly described or not standardized	Pair each risk tier with a predefined intervention and prospective monitoring plan
Stage 4: Intervention-enabled implementation	Use prediction to trigger scheduling action	Selective overbooking, targeted reminders, waitlist activation, staff outreach	Short-term reduction in no-shows, improved slot use	Few studies reach this stage; contextual barriers often dominate outcomes	Conduct prospective evaluation with control conditions, implementation reporting, and economic analysis
Stage 5: Outcome-based operational evaluation	Measure whether deployment improves clinic performance and patient access	Prospective cohorts, pragmatic trials, multi-site deployment, cost-effectiveness analysis	No-show reduction, throughput, revenue, wait time, equity, patient access	Evidence remains sparse, single-site, and short-term	Expand to multi-site trials with standardized outcome reporting and long-term follow-up
Stage 6: Scalable learning health system adoption	Sustain, recalibrate, and govern deployed models over time	Monitoring dashboards, recalibration protocols, fairness auditing, governance structures	Durable performance, equitable impact, adaptability across sites	Nearly absent from the reviewed literature	Build institutional governance, continuous evaluation systems, and transportability protocols

Feature category synthesis

Historical features consistently outperformed other categories in predictive importance across the reviewed studies providing a robust foundation for model development in diverse outpatient settings [3, 6, 8, 12, 17]. Demographic and socioeconomic variables contributed modestly and were most valuable when combined with historical data rather than used in isolation [7, 10, 19, 25, 30]. Environmental features such as weather or day of week showed inconsistent associations and added limited value after accounting for stronger predictors [11, 27-29, 32].

Behavioral indicators including appointment lead time and reminder response emerged as promising yet underutilized complements to historical data in several investigations [1, 2, 4, 5, 24]. This synthesis suggests that future models should prioritize readily extractable historical and behavioral features while selectively incorporating socioeconomic data to enhance equity considerations [9, 13-15, 32]. Standardization of feature definitions across studies would further strengthen the generalizability of these findings [16, 20-22, 26].

Table 2 consolidates the feature domains used across the reviewed studies by distinguishing their relative predictive consistency, operational actionability, and implementation relevance, thereby clarifying which data classes are most useful for deployment-oriented model design.

Table 2. Conceptual comparison of predictive feature domains for outpatient no-show modeling and their translational value for operational deployment

Feature domain	Typical variables in	Predictive consistency	Operational actionability	Main strength for model	Main limitation for	Strategic implication for

	reviewed studies	**across studies**		**building**	**implementation**	**future research and practice**
Historical attendance features	Prior no-show history, prior attendance adherence, past cancellations, repeat appointment behavior	High	High	Most robust and consistently predictive domain across specialties; easily derived from existing records	Can reinforce historical utilization inequities if used without fairness checks	Should remain the core feature layer in deployment-ready models, but must be monitored for bias amplification
Demographic features	Age, sex, insurance status, patient type	Moderate	Low to moderate	Readily available and improves calibration when combined with stronger predictors	Often weak as standalone predictors and may raise fairness or interpretability concerns	Best used as secondary adjustment features rather than primary intervention triggers
Behavioral / appointment management features	Lead time, booking interval, reminder response, rescheduling behavior, time since prior visit	Moderate to high	High	Directly linked to modifiable scheduling processes and intervention design	Underreported and inconsistently defined across studies	Particularly valuable for translating prediction into targeted reminders or workflow interventions
Socioeconomic features	Income proxies, deprivation indicators, employment proxies, transportation burden	Low to moderate	Moderate	May capture structural barriers not visible in routine scheduling data	Often unavailable, inconsistently measured, and difficult to standardize across sites	Should be selectively incorporated to support equity-aware modeling, especially in vulnerable populations
Geographic access features	Distance to clinic, travel time, rurality	Moderate	Moderate	Conceptually relevant to attendance burden and often useful in dispersed catchment areas	Context dependent and sensitive to local transport infrastructure	Valuable for site-specific recalibration and intervention tailoring rather than universal model transfer
Environmental / temporal features	Day of week, time of day,	Low to moderate	Moderate	Easy to extract and sometimes useful for	Usually weaker after stronger historical and	Better suited to scheduling refinement than

	season, weather			schedule-level optimization	behavioral predictors are included	to patient-level risk determination
Clinical / specialty-specific features	Specialty, diagnosis grouping, visit type, provider type	Moderate	Moderate to high	Helps explain heterogeneity across primary care, specialty, pediatric, and mental health settings	Often reduces portability across institutions and specialties	Important for local adaptation and specialty-specific model governance
Multi-domain integrated feature sets	Combined historical, behavioral, demographic, and contextual variables	High when well curated	High	Offers the best balance of discrimination and operational usefulness	Requires stronger data governance, standardization, and workflow coordination	Most promising pathway for prospective implementation studies and clinic-facing deployment

Specialty variation

Mental health clinics exhibited higher baseline no-show rates and relied more heavily on behavioral and socioeconomic features compared with primary care or specialty settings [18, 23, 27, 29]. Pediatric and adult outpatient populations showed distinct predictor patterns with prior family attendance history proving particularly influential in pediatric contexts [1, 2, 4, 24]. Primary care studies generally reported higher model performance than specialty clinics potentially due to greater data homogeneity and volume [7, 9, 10,13].

These specialty-specific differences highlight the need for tailored model development and validation rather than one-size-fits-all approaches [6, 8, 11, 12, 30]. Variation in implementation feasibility was also noted with mental health settings facing additional privacy and engagement challenges [14-16, 20]. Accounting for such contextual factors will be critical for successful operational rollout across the outpatient landscape [21, 22, 26].

Limitations

Review limitations

Publication bias may have led to overrepresentation of positive modeling results as studies with poor performance are less likely to reach peer-reviewed journals [1, 2, 4, 24]. Heterogeneity in no-show definitions across the 32 studies complicated direct comparisons of prevalence and model performance metrics [3, 10, 17, 25]. Specialty variation and differences in data sources further limited the ability to perform quantitative meta-analysis [6, 8, 12, 30].

Despite comprehensive searching the restriction to English-language publications from 2017 to 2023 may have omitted relevant non-English or earlier foundational work [7, 9, 13, 14]. The narrative synthesis approach while appropriate for the heterogeneous literature inherently involves interpretive judgment [15, 16, 20, 21]. These methodological constraints should be considered when applying the review findings to specific clinical contexts [22, 26].

Evidence base limitations

Few studies progressed beyond retrospective validation resulting in a sparse evidence base for operational implementation and long-term outcome measurement [18, 19, 23, 27, 29]. Single-site designs predominated limiting

generalizability to diverse healthcare systems and patient populations [5, 11, 28, 31, 32]. Short follow-up periods and infrequent cost-effectiveness analyses further restrict insights into sustained clinical and financial benefits [1, 2, 4, 24, 25].

The absence of randomized controlled trials comparing machine learning-guided scheduling against usual care represents a major gap in the current literature [7, 9, 10, 13, 14]. Reliance on historical electronic health record data introduces potential biases related to data quality and completeness [15, 16, 20, 21]. Future research must address these limitations to strengthen the evidence supporting widespread adoption of no-show prediction systems [22, 26].

Comparison with prior reviews

Prior systematic reviews on patient no-show prediction include the work of Carreras-García and colleagues as well as Dantas and colleagues [18, 28]. These earlier reviews covered literature primarily up to 2020 and 2018 respectively with a predominant focus on model development and statistical performance rather than real-world deployment [30]. Their time windows and objectives therefore overlap only partially with the present synthesis that extends through 2023.

Agreement exists across reviews on the dominant role of historical attendance features as the strongest predictors of no-show risk [18, 21, 28]. However the current review uniquely incorporates a dedicated evaluation of operational implementation success rates which were largely absent from previous syntheses [1, 2]. This addition reveals a consistent translational shortfall that earlier works noted only in passing.

The novel contribution of this review lies in its systematic assessment of implementation success rates alongside model types and feature categories [3, 4]. By restricting the scope to 2017–2023 peer-reviewed outpatient studies it provides an updated and more operationally oriented perspective than prior efforts [17, 23]. This focus directly addresses the gap between predictive accuracy and measurable clinic-level impact.

Recommendations

For researchers

Researchers should prioritize reporting operational implementation attempts even when outcomes are modest or negative to build a more complete evidence base [6, 8]. Future studies must move beyond AUROC alone and include direct measures of no-show reduction clinic throughput and revenue impact following deployment [18, 19]. Publication of negative results and detailed cost-benefit analyses will accelerate progress toward clinically actionable models.

Standardized reporting templates that capture intervention details workflow integration and long-term follow-up should become the norm in no-show prediction research [24, 25]. Collaboration with implementation scientists early in the modeling phase will help bridge the current divide between technical development and practical utility [5, 27]. Such practices will enhance the relevance and replicability of findings across diverse outpatient settings.

For journal editors and reviewers

Journal editors and reviewers should prioritize submissions that include operational implementation data over purely retrospective prediction studies [10, 12]. Requirements for explicit discussion of workflow feasibility staff training needs and integration challenges will raise the bar for publication [28, 29]. Prediction-only papers lacking any implementation plan or prospective validation pathway should be directed toward revision or rejection.

Editorial policies could mandate inclusion of cost-effectiveness or patient-access metrics whenever models are deployed in live clinics [13, 30]. This shift would discourage incremental modeling papers and encourage high-impact work that demonstrates tangible benefits to healthcare operations [31, 32]. Reviewers trained in implementation science can further strengthen this emphasis during peer review.

For clinic administrators

Clinic administrators should begin with simple yet effective models such as logistic regression trained on historical attendance features before advancing to more complex ensembles [7, 9]. Pilot testing in a single specialty or site allows refinement of risk thresholds and intervention protocols prior to organization-wide rollout [11, 14]. Integration with existing reminder systems and electronic health record alerts maximizes adoption without disrupting daily workflows.

Clear governance structures including data privacy safeguards and staff education programs are essential for successful deployment [15, 16]. Administrators should track both clinical and financial outcomes to build internal evidence supporting sustained investment in these tools [20, 21]. Starting small and scaling with demonstrated success offers the most pragmatic path forward.

Research gaps

Prospective implementation trials

Prospective randomized controlled trials comparing machine learning-guided scheduling against usual care remain almost entirely absent from the literature [22, 26]. Multi-site studies with long-term follow-up are needed to establish causality between model deployment and reductions in no-show rates or improvements in access [1, 2]. Such trials should incorporate health-economic endpoints and patient-reported outcomes to capture the full value proposition.

The current reliance on retrospective single-center designs limits confidence in generalizability across heterogeneous outpatient environments [3, 4]. Future implementation trials must address these shortcomings through pragmatic designs embedded in real-world clinic operations [17, 23]. Only then can the field move from promising models to proven system-level interventions.

Intervention optimization

Clear evidence is lacking on which intervention matches which risk stratum such as double-booking for very high-risk appointments versus enhanced reminders for moderate-risk cases [6, 8]. Head-to-head comparisons of different operational strategies triggered by the same prediction model are required to optimize resource allocation [18, 19]. Adaptive trial designs could help identify the most cost-effective pairings in real time.

Contextual factors including clinic size specialty and patient demographics likely moderate intervention effectiveness yet remain poorly studied [24, 25]. Targeted research into personalized intervention pathways will prevent one-size-fits-all approaches that waste resources or alienate patients [5, 27]. Filling this gap will maximize the return on predictive modeling investments.

Generalizability across settings

Model transferability across different clinics health systems and geographic regions has received minimal attention in the included studies [10, 12]. Standardized feature definitions and federated learning approaches could facilitate multi-site model development while preserving data privacy [28, 29]. Research addressing external validation and domain adaptation is urgently needed to support broader adoption.

Variations in electronic health record data quality and local scheduling practices further complicate generalizability [13, 30]. Prospective studies that test model updating protocols and site-specific recalibration will strengthen confidence in cross-setting performance [31, 32]. Overcoming these barriers is essential for equitable application of no-show prediction technology.

Implications

For research practice

The field must shift from prediction-only studies toward integrated prediction-to-intervention research that incorporates implementation science frameworks from the outset [7, 9]. Embedding health services researchers and clinicians in modeling teams will improve the relevance and feasibility of proposed solutions [11, 14]. This collaborative approach will accelerate the generation of evidence that directly informs practice.

Funding agencies and academic promotion criteria should reward studies that demonstrate operational impact rather than incremental improvements in AUROC alone [15, 16]. Adoption of open datasets and code-sharing practices will further enhance reproducibility and cumulative progress [20, 21]. These changes in research practice are necessary to close the existing translational gap.

For clinical practice

Current evidence supports the use of machine learning for risk stratification in outpatient scheduling yet local validation on site-specific data remains essential before any deployment [22, 26]. Clinics should begin with straightforward models leveraging historical features and gradually incorporate more advanced techniques as infrastructure matures [1, 2]. Ongoing monitoring of both predictive and operational performance will ensure sustained benefits.

Integration into existing workflows without adding administrative burden is key to clinician acceptance and long-term success [3, 4]. Patient engagement strategies including transparent communication about how predictions are used will help maintain trust in the system [17, 23]. When implemented thoughtfully these tools can meaningfully improve appointment availability and reduce inequities in access.

For policy and regulation

Policy makers should classify no-show prediction systems as quality improvement tools rather than regulated medical devices to avoid unnecessary barriers to innovation [6, 8]. Development of standardized reporting requirements for operational outcomes would promote transparency and comparability across health systems [18, 19]. Such guidance could be incorporated into national quality frameworks and value-based care models.

Regulatory bodies can encourage adoption by offering incentives for clinics that demonstrate measurable reductions in no-show rates through data-driven scheduling [24, 25]. Investment in infrastructure for secure data sharing and model validation will support equitable implementation across diverse settings [5, 27]. These policy measures will help translate research findings into widespread improvements in healthcare efficiency.

Conclusion

This systematic review set out to examine machine learning model types feature categories and operational implementation success rates for patient no-show prediction in outpatient clinics between 2017 and 2023. The 32 included studies demonstrate that ensemble methods and historical features consistently deliver strong retrospective performance. Nevertheless operational implementation remains rare and direct measurement of no-show reduction is even rarer.

The implementation gap identified throughout the evidence base is critical because accurate predictions alone do not reduce unused appointment slots or improve clinic efficiency. Without corresponding interventions the promise of machine learning remains largely unrealized in everyday outpatient operations. Bridging this divide requires deliberate focus on prospective deployment and outcome tracking.

Reporting of operational outcomes should become the expected standard rather than the exception in future no-show prediction research. Journals funders and researchers alike must elevate implementation metrics to the same level of importance as technical performance indicators. Only then will the literature evolve from descriptive modeling to transformative system change.

The ultimate vision is one of fully integrated no-show prediction and intervention systems that deliver measurable gains in patient access clinic efficiency and resource equity. Achieving this vision will depend on sustained collaboration across research clinical and policy domains. When realized these advances will help ensure that every scheduled outpatient appointment contributes meaningfully to better health outcomes for all.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 22 Mar 2023 Revised: 14 Jun 2023 Accepted: 23 Jul 2023
Published online: 20 January 2024

References

Devasahay SR, Karpagam S, Ma NL. Predicting appointment misses in hospitals using data analytics. mHealth. 2017;3:12.

AlMuhaideb S, Alswailem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. Ann Saudi Med. 2019;39(6):373-81.

Aladeemy M, Adwan L, Booth A, Khasawneh MT, Poranki S. New feature selection methods based on opposition-based learning and self-adaptive cohort intelligence for predicting patient no-shows. Appl Soft Comput. 2020;86:105866.

Mohammadi I, Wu H, Turkcan A, Toscos T, Doebbeling BN. Data analytics and modeling for appointment no-show in community health centers. J Prim Care Community Health. 2018;9:2150132718811692.

Goffman RM, Harris SL, May JH, Milicevic AS, Monte RJ, et al. Modeling patient no-show history and predicting future outpatient appointment behavior in the Veterans Health Administration. Mil Med. 2017;182(5-6):e1708-e1714.

Srinivas S, Ravindran AR. Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: a prescriptive analytics framework. Expert Syst Appl. 2018;102:245-61.

Niu T, Lei B, Guo L, Fang S, Li Q, Gao B, et al. A review of optimization studies for system appointment scheduling. Axioms. 2023;13(1):16.

Ahmadi E, Garcia-Arce A, Masel DT, Reich E, Puckey J, Maff RM . A metaheuristic-based stacking model for predicting the risk of patient no-show and late cancellation for neurology appointments. IISE Trans Healthc Syst Eng. 2019;9(3):272-91.

Marbouh D, Khaleel I, Al Shanqiti K, Al Tamimi M, Simsekler MC, et al. Evaluating the impact of patient no-shows on service quality. Risk Manag Healthc Policy. 2020;13:509-17.

Ding X, Gellad ZF, Mather C III, Barth P, Poon EG, et al. Designing risk prediction models for ambulatory no-shows across different specialties and clinics. J Am Med Inform Assoc. 2018;25(8):924-30.

Malloy M, Tarima S, Canales B, Nelson D, Hanley J. Identifying risk factors for appointment no-shows in a pediatric orthopaedic surgery clinic. J Pediatr Orthop Soc N Am. 2023;5(3):695.

Lenzi H, Ben ÂJ, Stein AT. Development and validation of a patient no-show predictive model at a primary care setting in Southern Brazil. PLoS One. 2019;14(4):e0214869.

Fan G, Deng Z, Ye Q, Wang B. Machine learning-based prediction models for patient no-show in online outpatient appointments. Data Sci Manag. 2021;2:45-52.

Batool T, Abuelnoor M, El Boutari O, Aloul F, Sagahyroon A. Predicting hospital no-shows using machine learning. In: 2020 IEEE Int Conf Internet Things Intell Syst (IoTaIS). IEEE. 2021:142-8.

Boshers EB, Cooley ME, Stahnke B. Examining no-show rates in a community health centre in the United States. Health Soc Care Community. 2022;30(5):e2041-e2049.

Dunstan J, Villena F, Hoyos JP, Riquelme V, Royer M, Ramírez H, et al. Predicting no-show appointments in a pediatric hospital in Chile using machine learning. Health Care Manag Sci. 2023;26(2):313-29.
https://doi.org/10.1007/s10729-022-09626-z.

Topuz K, Uner H, Oztekin A, Yildirim MB. Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and Bayesian belief network. Ann Oper Res. 2018;263(1):479-99.

Carreras-García D, Delgado-Gómez D, Llorente-Fernández F, Arribas-Gil A. Patient no-show prediction: a systematic literature review. Entropy. 2020;22(6):675.

Hamdan AF, Bakar AA. Machine learning predictions on outpatient no-show appointments in a Malaysia major tertiary hospital. Malays J Med Sci. 2023;30(5):169.

Babayoff O, Shehory O, Geller S, Shitrit-Niselbaum C, Weiss-Meilik A, Sprecher E. Improving Hospital Outpatient Clinics Appointment Schedules by Prediction Models. J Med Syst. 2022;47(1):5.
https://doi.org/10.1007/s10916-022-01902-3.

Leibner G, Brammli-Greenberg S, Mendlovic J, Israeli A. To charge or not to charge: reducing patient no-show. Isr J Health Policy Res. 2023;12(1):27.

Rosenbaum JI, Mieloszyk RJ, Hall CS, Hippe DS, Gunn ML, et al. Understanding why patients no-show: observations of 2.9 million outpatient imaging visits over 16 years. J Am Coll Radiol. 2018;15(7):944-50.

Mochón F, Elvira C, Ochoa A, Gonzalvez JC. Machine-learning-based no show prediction in outpatient visits. IJIMAI [Internet]. 2018;4(7):29-34.

Alaeddinia A, Hong SH. A multi-way multi-task learning approach for multinomial logistic regression. Methods Inf Med. 2017;56(4):294-307.

Chua SL, Chow WL. Development of predictive scoring model for risk stratification of no-show at a public hospital specialist outpatient clinic. Proc Singap Healthc. 2019;28(2):96-104.

Barraza EL. Decreasing no-show rates in an outpatient specialty clinic. Master's Projects and Capstones. 2023.

Liu D, Shin WY, Sprecher E, Conroy K, Santiago O, Wachtel G, et al. Machine learning approaches to predicting no-shows in pediatric medical appointment. NPJ Digit Med. 2022;5(1):50.
https://doi.org/10.1038/s41746-022-00594-w.

Dantas LF, Fleck JL, Oliveira FL, Hamacher S. No-shows in appointment scheduling—a systematic literature review. Health Policy. 2018;122(4):412-21.

Ahmad MU, Zhang A, Mhaskar R. A predictive model for decreasing clinical no-show rates in a primary care setting. Int J Healthc Manag. 2021;14(3):829-36.

Chen J, Goldstein IH, Lin WC, Chiang MF, Hribar MR. Application of machine learning to predict patient no-shows in an academic pediatric ophthalmology clinic. AMIA Annu Symp Proc. 2020;2020:293.

Agarwal P, Nathan AS, Jaleel Z, Levi JR. Factors contributing to missed appointments in a pediatric otolaryngology clinic. Laryngoscope. 2022;132(4):895-900.

Salazar LH, Parreira WD, Fernandes AM, Leithardt VR. No-show in medical appointments with machine learning techniques: a systematic literature review. Information. 2022;13(11):507.