International Journal of Current Sciences and Engineering Intelligence
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: *https://ijcsi.com*
Volume 01, Issue 01, September 2026
OPEN ACCESS
E-ISSN :

Capacity Aware Prioritization of Clinical Decision Support Alerts

Atharv Menon[1], Akash Chavan[2]
[1,2] *Students, Rabindranath Tagore University, Raisen*

Article Info

Article History:
Published: 14 Nov 2026

Publication Issue:
Volume 01, Issue 01
September 2026

Page Number:
27-33

Corresponding Author:
Atharv Menon

Abstract:
Clinical prediction systems can generate more alerts than a review team can process, making available staff time an essential part of evaluation. This paper develops a transparent analytical model for prioritizing a hypothetical routine alert queue under a fixed review budget. Four constructed profiles differ in actionable finding probability, case count, and expected review duration. A probability threshold, risk ranking, findings maximization, and findings maximization with a protected review requirement are compared. With 240 minutes available, risk ranking reviews 20 cases and identifies eight expected actionable findings. Unrestricted optimization reviews 75 cases and identifies 11.75 expected findings, but excludes the highest probability profile. Requiring eight protected reviews from that profile yields 56 reviews and 10.40 expected findings. A threshold applied without a capacity constraint requires 780 minutes and is infeasible. Sensitivity analysis shows that underestimated review times and miscalibrated probabilities can substantially reduce the apparent advantage. All results are calculated from synthetic inputs and do not demonstrate improved clinical outcomes. The contribution is a reproducible operational framework that makes workload, objective choice, and protection rules explicit before any prospective clinical evaluation.
Keywords: Clinical decision support; alert workload; probability calibration; capacity constraints; operational evaluation; health informatics.

1 Introduction

Clinical decision support systems can assist professionals by identifying records that may warrant additional attention. Their practical value depends on more than the accuracy of a predictive model. A review team needs time to inspect the record, resolve conflicting information, communicate with the responsible clinician, and document the action taken. If the alert volume exceeds available capacity, some alerts will remain unreviewed or receive delayed attention. A threshold chosen from a prediction curve can therefore create a workload that the service cannot actually complete.

Decision curve analysis connects prediction with consequences by evaluating net benefit across decision thresholds [1]. Its interpretation depends on the clinical meaning of the trade-offs represented by those thresholds [2]. Calibration is equally important because a probability used in a decision should correspond to the event frequency it claims to represent [3]. These principles motivate evaluating alert policies in relation to their intended use. They do not remove the need to represent the staff time required to perform the action triggered by an alert.

A further complication is that high probability does not necessarily mean short review time, large treatment benefit, or urgent need. A complex record may require substantially more work than a straightforward record with a lower predicted probability. Optimizing only the expected number of actionable findings can consequently favor many simple cases over fewer difficult ones. This may be efficient under a narrow operational objective while conflicting with clinical priorities. The objective and any protection requirements must therefore be stated explicitly rather than inferred from a model's score.

This paper examines these issues using a constructed routine review queue. It compares feasible and infeasible policies under a common time budget and quantifies the effect of protecting a subset of higher probability cases. The analysis does not identify a disease, prescribe a treatment threshold, or decide access to emergency care. Urgent escalation

Copyright © 2026 The Author(s) : This is an open access article under the CC BY license (http://creativecommons.org/licenses/by-nc/4.0/)

belongs to a separate clinical pathway and is outside the modeled queue. The contribution is a reproducible demonstration of how workload and objective choice alter an alert policy, together with a framework for the evidence needed before clinical use.

2 Detailed Literature Review

Vickers and Elkin introduce decision curve analysis as a method for comparing prediction based strategies through their clinical consequences [1]. Vickers, Van Calster, and Steyerberg provide guidance for interpreting these curves [2]. Their work explains why discrimination alone is insufficient for evaluating an operational decision. A model may separate higher and lower risk records while still supporting an unsuitable action threshold if the associated benefits, harms, and workload do not match the intended service.

Van Calster and colleagues describe calibration as a central weakness in predictive analytics [3]. Steyerberg and colleagues set out a framework for assessing prediction model performance through several complementary measures [4]. These contributions support treating ranking, probability accuracy, and decision usefulness as distinct properties. A score can rank cases correctly but systematically overestimate their event probabilities. Such an error matters directly when expected findings are multiplied by probability estimates to allocate a limited resource.

Transparent reporting and risk of bias assessment are necessary for judging model evidence. The original TRIPOD statement specifies reporting expectations for prediction model studies [5], while PROBAST provides a structured approach to assessing bias and applicability [6]. TRIPOD+AI updates reporting guidance for regression and machine learning models [7], and PROBAST+AI extends quality and applicability assessment [8]. These instruments do not certify a clinical system as effective. They help readers determine what was studied, how it was evaluated, and whether essential information is missing.

External validation can reveal problems that are not visible in development results. Wong and colleagues evaluate a widely implemented proprietary sepsis prediction model in hospitalized patients [9]. Henry and colleagues develop a targeted early warning score for septic shock [10], while Adams and colleagues report a prospective multisite study after implementation of the TREWS system [11]. These studies concern different designs and settings, so their reported results should not be treated as interchangeable. Their relevance here is the need to distinguish model development, local validation, implementation, and patient outcome evaluation.

Sendak and colleagues describe the implementation of a sepsis deep learning technology in routine care [12]. Their operational focus illustrates that integrating a prediction system requires a defined workflow, responsibility, and communication structure. Sittig and Singh's sociotechnical model similarly places health information technology within an interacting organizational system [13]. These perspectives support treating review capacity as part of the intervention rather than an inconvenience to be considered only after model development.

Ancker and colleagues study workload, complexity, repeated alerts, and alert fatigue [14]. Sutton and colleagues review the benefits, risks, and implementation strategies of clinical decision support [15]. These contributions motivate attention to the work created by alerts and the context in which professionals respond. The current model simplifies that work to a review duration, which is useful for transparent arithmetic but does not capture interruption cost, cognitive burden, handoffs, or the possibility that review effort changes during a busy shift.

Fairness is another concern when a model becomes a resource allocation rule. Obermeyer and colleagues show how a healthcare algorithm's target can introduce racial bias [16]. Rajkomar and colleagues discuss fairness in machine learning for health equity [17]. These studies caution that an apparently neutral objective can reproduce undesirable differences. The synthetic profiles in this paper are workload categories rather than demographic groups, but a real implementation would need to examine whether review duration, missing data, and outcome labels act as proxies for unequal access or documentation quality.

Precision and recall provide useful workload-related descriptions, particularly when actionable events are uncommon. Saito and Rehmsmeier explain the value of precision-recall analysis for imbalanced classification [18]. Calibration methods such as those studied by Guo and colleagues address a different component of probability quality [19]. Finally, constrained optimization provides tools for allocating limited resources under explicit objectives [20]. This focused narrative synthesis supports combining these perspectives rather than expecting one metric to summarize the entire clinical workflow. It does not pool clinical effect estimates or establish that the numerical scenario is representative of a hospital.

3 Detailed Methodology & Framework

3.1 Synthetic queue and assumptions

The scenario contains 200 independent routine alerts divided into four profiles. Profile H contains 50 cases, each with actionable finding probability 0.40 and review duration 12 minutes. Profile M contains 30 cases with probability 0.25 and duration six minutes. Profile L contains 70 cases with probability 0.15 and duration three minutes. Profile V contains 50 cases with probability 0.05 and duration two minutes. The total expected number of actionable findings in the full queue is 40.5. These are constructed inputs, not patient data or measured clinical rates.

An actionable finding means that record review identifies information requiring the predefined routine follow-up workflow. It does not mean that the patient benefits from a treatment, that the condition would otherwise be missed, or that an adverse event is prevented. The model assumes one review per case and no interaction between reviews. Review durations are deterministic averages used for capacity accounting. There is no queue arrival process, repeated alert suppression, or waiting-time model. Available capacity is 240 minutes for the entire constructed queue.

3.2 Objective and protection requirement

The optimization chooses an integer number of cases from each profile. Its objective is the sum of selected cases multiplied by their actionable finding probabilities. Total selected review time must not exceed 240 minutes, and the selected count from each profile cannot exceed the available cases. Equation 1 expresses this finite allocation problem. The decision variables are profile counts, not patient diagnoses. A second version requires at least eight H-profile reviews, representing an externally specified protection rule for a subset of routine cases.

$$max \sum_g p_g x_g ; \ \sum_g t_g x_g \ \leq \ 240$$

$$0 \ \leq \ x_g \ \leq \ n_g, \ x_g \ \in \ \mathbb{Z}; \ \ x_H \ \geq \ 8$$

Equation 1. Expected actionable findings subject to time and integer constraints; g indexes profiles. The final constraint applies only to the protected version.

The protection requirement illustrates how a clinical or service priority can constrain an efficiency objective. Eight is an assumed planning quantity, not a recommended clinical number. The rule is not a substitute for an urgent escalation process and does not establish that the remaining H-profile cases can safely wait. In an actual service, the permitted queue, protection criteria, and review window would need clinical approval and prospective evaluation. The scenario examines the arithmetic consequence of the constraint without claiming that the constraint is sufficient for safe care.

Figure 1 places capacity allocation after eligibility, calibration, and an independently defined escalation check. The feedback path records completed review time and adjudicated findings, allowing subsequent evaluation of both workload and probabilities. Routing a case to the routine queue does not remove the responsible clinician's ability to act through ordinary care. The diagram therefore describes decision support around a defined workflow rather than autonomous treatment assignment.

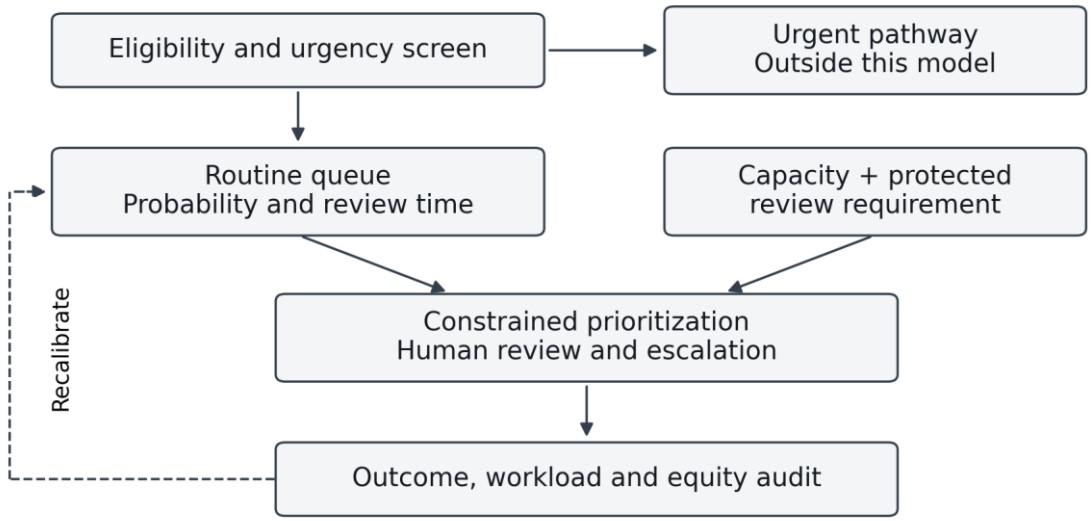


Figure 1. Routine alert review framework. Urgent cases follow a separate pathway; calibration, workload and equity audits are proposed safeguards.

3.3 Comparators and computational procedure

The unconstrained threshold comparator selects all profiles with probability at least 0.20, namely all H and M cases. It is retained to show whether a familiar score threshold creates feasible work, not to claim that an infeasible policy is a realistic operating baseline. The risk-ranked comparator reviews cases in decreasing probability order until the time budget is exhausted. With identical durations within H, it selects 20 H cases. The efficiency comparator maximizes expected findings under the time constraint, while the protected comparator adds the eight-case requirement.

An integer dynamic program evaluates feasible allocations and identifies the optimum for the stated objective. The implementation represents time in half-minute units so that the later duration sensitivity can be reproduced exactly. It does not rely on a simple probability-per-minute greedy rule, which is not generally optimal for indivisible cases. Although that ratio helps explain the chosen allocation, the reported solution is checked over the relevant discrete state space.

Evaluation reports review count, required time, expected findings, expected precision, and expected recall. Precision is expected findings divided by reviewed cases. Recall is expected findings divided by the expected findings in the complete queue. These are modeled ratios of expectations rather than estimates from realized patient outcomes. Sensitivity analysis increases every review duration by 50 percent and separately lowers L-profile probability from 0.15 to 0.10. No inferential significance test is appropriate because no sampled clinical observations were collected.

4 Detailed Results and Analysis

4.1 Feasibility and policy comparison

The probability threshold selects 80 cases and requires 780 minutes, exceeding capacity by 540 minutes. Its expected 27.5 findings cannot be interpreted as performance achievable by the 240-minute service. Table 1 therefore labels this policy infeasible. This comparison illustrates an operational failure that may be missed when an alert threshold is assessed only through sensitivity or precision. A prediction threshold specifies whom to alert; it does not guarantee that the resulting work can be completed.

Table 1. Synthetic queue results with 240 minutes of review capacity.

Policy	Reviews	Minutes	Expected findings	Recall (%)	Within capacity?
Threshold ≥ 0.20	80	780	27.50	67.90	No

Policy	Reviews	Minutes	Expected findings	Recall (%)	Within capacity?
Risk ranked	20	240	8.00	19.75	Yes
Findings optimized	75	240	11.75	29.01	Yes
8 protected H reviews	56	240	10.40	25.68	Yes

Recall uses 40.5 expected actionable findings in the full queue as its denominator. The unrestricted threshold policy is infeasible.

Risk ranking selects 20 H-profile reviews, uses all 240 minutes, and produces eight expected findings. It achieves the highest precision among the feasible policies, 40 percent, but expected recall is only 19.75 percent. Findings optimization instead selects five M cases and all 70 L cases. It also uses 240 minutes but produces 11.75 expected findings, a 46.88 percent increase over risk ranking. Its precision falls to 15.67 percent because many more low probability cases are reviewed, while expected recall increases to 29.01 percent.

The unconstrained efficiency solution excludes all H-profile cases. This is not presented as an acceptable clinical recommendation. It demonstrates the consequences of choosing expected finding count as the entire objective when high probability cases take longer to review. The H profile generates approximately 0.033 expected findings per minute, compared with 0.042 for M and 0.050 for L. An efficiency objective therefore favors L even though each L case has a lower probability than each H case. Objective selection has ethical and clinical implications that arithmetic cannot resolve.

The protected policy selects eight H cases and 48 L cases. It reviews 56 records, identifies 10.40 expected findings, and achieves expected recall of 25.68 percent. Compared with unrestricted optimization, the protection constraint costs 1.35 expected findings. Compared with risk ranking, it gains 2.40 expected findings while preserving a specified subset of H reviews. This explicit trade-off is more informative than describing the constrained policy as simultaneously best on every metric. It remains a constructed example whose protection rule requires external justification.

4.2 Sensitivity and interpretation

Increasing all review durations by 50 percent reduces the protected solution to eight H cases and 21 L cases. The allocation consumes 238.5 minutes and produces 6.35 expected findings. The remaining 1.5 minutes cannot accommodate another case. This result shows why average review-time estimates need validation and why an apparently small unused capacity may be an unavoidable consequence of indivisible tasks. Real workflows would also require allowance for interruptions and uncertainty rather than planning to use every minute exactly.

If L-profile probability is actually 0.10, the original unrestricted solution produces 8.25 expected findings rather than 11.75. Its advantage over the risk-ranked policy's eight findings almost disappears. Holding the original allocation fixed, the L probability must exceed approximately 0.0964 for unrestricted optimization to outperform risk ranking on expected findings. This breakpoint follows from comparing 70 times the L probability plus 1.25 findings from M with eight findings from H. It is a sensitivity result, not a clinical threshold for an individual patient.

No policy comparison establishes improved survival, reduced complications, or better patient experience. Actionable findings may differ in severity and benefit, while review duration may reflect complex needs that should not be deprioritized. The model also excludes simultaneous arrivals, incomplete reviews, clinician overrides, and downstream treatment capacity. These limitations explain why retrospective prediction metrics and simple workload calculations cannot substitute for a prospective evaluation of the full intervention.

5 Future Enhancement

Future work should define a clinically justified routine review population and collect deidentified measurements of review duration, actionable findings, duplicate alerts, and escalation outcomes. A silent prospective evaluation could

compare projected workloads with actual operational capacity before any workflow changes are introduced. Calibration should be assessed over time and across relevant patient groups, with uncertainty around both event probabilities and review times propagated into the allocation decision.

The optimization objective should then be extended to incorporate clinically justified consequences, waiting-time limits, and protection requirements. These values must come from an accountable clinical process rather than from a convenient numerical example. Queueing models could examine arrivals, staffing changes, and deferred work across shifts. Fairness analysis should investigate whether documentation, language, or care complexity systematically changes access to review. A subsequent controlled implementation would be needed to determine whether the resulting workflow improves outcomes that matter to patients and clinicians.

6 Conclusion

Review capacity changes the interpretation of a clinical alert policy. In the specified synthetic queue, a simple probability threshold produces infeasible work, while risk ranking and findings optimization make different trade-offs between precision and the number of actionable findings. Adding a protected review requirement produces an intermediate allocation with a visible opportunity cost. These results demonstrate why the operational objective and its constraints should be reported alongside prediction performance.

The model's value is explanatory and reproducible. It does not validate a clinical prioritization rule or establish a benefit to patient outcomes. Miscalibrated probabilities and underestimated review durations substantially weaken its apparent advantages. A real application therefore requires locally justified objectives, independent urgent escalation, workload measurement, subgroup evaluation, and prospective assessment of the complete decision support workflow.

References

[1] Vickers, A. J., and Elkin, E. B. (2006). Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Medical Decision Making, 26(6), 565-574. doi: 10.1177/0272989X06295361

[2] Vickers, A. J., Van Calster, B., and Steyerberg, E. W. (2019). A simple, step-by-step guide to interpreting decision curve analysis. Diagnostic and Prognostic Research. doi: 10.1186/s41512-019-0064-7

[3] Van Calster, B., et al. (2019). Calibration: The Achilles heel of predictive analytics. BMC Medicine, 17, 230. doi: 10.1186/s12916-019-1466-7

[4] Steyerberg, E. W., et al. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology, 21(1), 128-138. doi: 10.1097/EDE.0b013e3181c30fb2

[5] Collins, G. S., Reitsma, J. B., Altman, D. G., and Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. BMJ, 350, g7594. doi: 10.1136/bmj.g7594

[6] Wolff, R. F., et al. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Annals of Internal Medicine, 170(1), 51-58. doi: 10.7326/M18-1376

[7] Collins, G. S., et al. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078378. doi: 10.1136/bmj-2023-078378

[8] Moons, K. G. M., et al. (2025). PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ, 388, e082505. doi: 10.1136/bmj-2024-082505

[9] Wong, A., et al. (2021). External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Internal Medicine, 181(8), 1065-1070. doi: 10.1001/jamainternmed.2021.2626

[10] Henry, K. E., Hager, D. N., Pronovost, P. J., and Saria, S. (2015). A targeted real-time early warning score (TREWScore) for septic shock. Science Translational Medicine, 7(299), 299ra122. doi: 10.1126/scitranslmed.aab3719

[11] Adams, R., et al. (2022). Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. Nature Medicine. doi: 10.1038/s41591-022-01894-0

[12] Sendak, M. P., et al. (2020). Real-World Integration of a Sepsis Deep Learning Technology Into Routine Clinical Care: Implementation Study. JMIR Medical Informatics, 8(7), e15182. doi: 10.2196/15182

[13] Sittig, D. F., and Singh, H. (2010). A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Quality and Safety in Health Care, 19(Suppl 3), i68-i74. doi: 10.1136/qshc.2010.042085

[14] Ancker, J. S., et al. (2017). Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Medical Informatics and Decision Making, 17, 36. doi: 10.1186/s12911-017-0430-8

[15] Sutton, R. T., et al. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. npj Digital Medicine, 3, 17. doi: 10.1038/s41746-020-0221-y

[16] Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. doi: 10.1126/science.aax2342

[17] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., and Chin, M. H. (2018). Ensuring Fairness in Machine Learning to Advance Health Equity. Annals of Internal Medicine, 169(12), 866-872. doi: 10.7326/M18-1990

[18] Saito, T., and Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLOS ONE, 10(3), e0118432. doi: 10.1371/journal.pone.0118432

[19] Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). On Calibration of Modern Neural Networks. Proceedings of ICML, PMLR 70, 1321-1330. Source

[20] Boyd, S., and Vandenberghe, L. (2004). Convex Optimization. Cambridge University Press. Source