American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Ethical Use of Predictive Scores in Bedside Nursing

Prof. Wang Wenru

Professor, National University of Singapore (NUS), Singapore

Abstract

Predictive scores are increasingly incorporated into bedside care to support early recognition of clinical deterioration, prioritization of nursing surveillance, escalation of care, and allocation of clinical attention. Early warning scores offer a structured method for combining physiological observations into an interpretable risk signal, while newer electronic and algorithmic systems can integrate substantially larger clinical datasets. Their apparent objectivity, however, creates important ethical questions concerning false reassurance, excessive alerts, subgroup bias, incomplete observations, automation dependence, transparency, accountability, and the continuing role of professional nursing judgment. This simulation-based study develops an ethical framework for bedside use of predictive scores in adult inpatient nursing. A hypothetical dataset of 1,000 patient observation episodes was constructed to examine how changing escalation thresholds could alter sensitivity for clinical deterioration and the proportion of encounters generating alerts. Because no real patient records were supplied, all numerical results are synthetic and intended exclusively for methodological demonstration. Under the simulated model, a low escalation threshold produced 95% sensitivity but generated alerts in 49% of encounters, whereas a substantially higher threshold reduced alert burden to 11% but lowered sensitivity to 51%. The findings illustrate an ethical trade-off rather than an optimal clinical threshold.

Research available through 2020 demonstrates that early warning scores can discriminate risk but that their methodological quality and real-world implementation performance vary. A 2020 systematic review found important methodological weaknesses across many early warning score studies, while implementation research has shown that frequent low-value alerts can be ignored by frontline staff. Evidence from nursing literature also indicates that nurses use observation, experience, intuition, and pattern recognition in identifying deterioration, supporting a model in which predictive scores augment rather than replace bedside assessment. The study concludes that ethical deployment requires local validation, transparent escalation protocols, subgroup performance auditing, preservation of nurse-initiated escalation, documentation of overrides, and continuous monitoring for patient-safety consequences.

Keywords: predictive scores; bedside nursing; clinical deterioration; early warning scores; nursing ethics; algorithmic bias; clinical decision support; patient safety; nursing judgment; predictive analytics

1. Introduction

Predictive scoring has become an important feature of contemporary inpatient surveillance because physiological deterioration may develop before a catastrophic clinical event becomes obvious. Early warning systems combine routinely collected observations such as respiratory rate, oxygen saturation, temperature, blood pressure, pulse, and level of consciousness to create standardized risk signals that can prompt reassessment or escalation. The National Early Warning Score was developed to standardize this approach and demonstrated discrimination for outcomes including cardiac arrest, unexpected intensive-care admission, and death in its original evaluation. Such systems are attractive because they translate multiple observations into a common language that can support communication between nurses, physicians, rapid-response teams, and other clinicians.

Predictive scores nevertheless operate within a considerably more complex clinical environment than their numerical appearance suggests. A score depends on the observations entered into it, the timing and accuracy of measurement, the population in which it was developed, the outcome it was trained or calibrated to predict, and the threshold chosen for action. Ludikhuize and colleagues found that vital-sign documentation was frequently incomplete even among patients who later experienced serious adverse events, demonstrating that the quality of a predictive score cannot exceed the quality and completeness of its underlying observations. A highly sophisticated scoring system may therefore produce misleading reassurance when clinically important inputs are missing, delayed, incorrectly measured, or entered without appropriate contextual interpretation.

Nursing introduces an additional ethical dimension because bedside nurses observe patients continuously across repeated interactions and often recognize subtle patterns that are not fully represented by numerical variables. Odell, Victor, and Oliver's systematic review found that intuition was important in nurses' recognition of deterioration and that physiological observations were often used to validate an existing sense that something was wrong. Jensen, Skår, and Tveit's review subsequently found that early warning and rapid-response systems influenced nursing competence in assessment, referral, communication, and management of deteriorating patients, although effects were not uniformly straightforward. The ethical issue is therefore not whether nurses should use predictive scores, but how those scores should interact with professional judgment.

The emergence of machine-learning prediction adds further concerns. Char, Shah, and Magnus argued that implementation of machine learning in clinical care raises ethical issues concerning data, interpretation, professional responsibility, and the consequences of embedding algorithmic recommendations into practice. Wiens and colleagues likewise emphasized that responsible health-care machine learning requires careful definition of clinical problems, appropriate data, external and prospective evaluation, and ongoing monitoring after deployment. Obermeyer and colleagues demonstrated a particularly important example of algorithmic bias: a widely used health-management algorithm produced racial disparities because health-care cost was used as a proxy for health need. These findings show that predictive accuracy alone does not establish ethical acceptability.

The present study therefore evaluates bedside predictive scores through the combined perspectives of nursing practice, patient safety, predictive performance, and health-care ethics. It proposes that predictive scores should operate as structured decision-support signals rather than autonomous decision authorities. A simulated threshold analysis is used to illustrate the tension between detecting deterioration and creating excessive alert burden. The manuscript additionally develops a practical governance framework covering local validation, subgroup fairness, missing data, alert design, nursing override, escalation responsibility, transparency, and documentation. Since no empirical patient dataset was provided, every numerical result is synthetic and should be interpreted as a methodological illustration rather than evidence supporting a particular bedside scoring threshold.

2. Objectives of the Study

2.1 To Evaluate Ethical Risks Associated With Bedside Predictive Scores

The first objective is to identify the principal ethical risks created when predictive scores influence bedside nursing surveillance and escalation. These risks include false-negative predictions that may delay recognition, false-positive alerts that can increase unnecessary escalation or alert fatigue, incomplete or inaccurate inputs, inappropriate use outside the population for which a score was validated, and excessive reliance on the score when bedside assessment indicates a different clinical picture. Systematic evidence indicates that early warning scores vary considerably in methodological quality and may perform less reliably than clinicians assume when development and validation methods are weak.

The ethical analysis also includes questions of autonomy and professional responsibility. Nurses remain accountable for observing, assessing, communicating, and escalating concerns within their professional scope even when a digital system assigns apparently reassuring risk. A score should therefore add information to nursing assessment rather than function as a procedural barrier preventing escalation.

2.2 To Examine the Balance Between Sensitivity and Alert Burden

The second objective is to illustrate the ethical consequences of changing the escalation threshold of a predictive score. A low threshold may identify a greater proportion of patients who subsequently deteriorate, but it may also create large numbers of alerts among patients who will not experience the targeted event. Bedoya and colleagues reported 175,357 electronic best-practice alerts during their implementation study and found that predictive performance varied substantially between hospital settings; the system was generally ignored by frontline nursing staff and did not appreciably alter the study's defined clinical outcomes.

Conversely, increasing the threshold to reduce unnecessary alerts can decrease sensitivity and increase the chance that a deteriorating patient will initially remain below the automated escalation level. Threshold selection is therefore a normative as well as statistical decision because it determines how false positives and false negatives are distributed across patients, nurses, and clinical resources.

2.3 To Develop a Nurse-Centered Governance Framework

The third objective is to formulate a governance model in which predictive scores are integrated with professional nursing assessment, clinical escalation policy, and ongoing ethical audit. WHO's 2018 consultation on big data and artificial intelligence identified concerns including privacy, governance, clinical decision-support use, and decisions made from probabilistic, imperfect, or flawed data. These concerns apply even when the bedside system is a conventional score rather than a complex artificial-intelligence model.

The proposed framework consequently requires transparent information about what the score predicts, which variables it uses, where it was validated, how missing data are handled, and what action is expected at different levels of risk. It also preserves an explicit route through which nurses can escalate concern irrespective of the calculated score.

3. Review of Literature

3.1 Early Warning Scores and Detection of Clinical Deterioration

Odell, Victor, and Oliver reviewed research concerning nurses' detection and management of deteriorating ward patients and identified the continuing importance of nursing intuition and clinical observation alongside physiological measurement. Their findings are particularly relevant to predictive scoring because the earliest indication of deterioration may be a change in appearance, behavior, interaction, work of breathing, functional status, or overall clinical pattern that is not yet fully reflected in an aggregated numerical score. A score-centered model that suppresses such concern would therefore remove rather than enhance a potentially valuable source of clinical information.

Ludikhuize and colleagues examined patients who experienced severe adverse events and retrospectively calculated Modified Early Warning Scores. Eighty-one percent had a score of at least three at some point during the preceding 48 hours, but vital-sign documentation was frequently incomplete. This study illustrates two distinct problems. First, physiological deterioration can often be detectable before a serious event. Second, a scoring system cannot function reliably when the observations required for calculation are missing or inconsistently obtained.

Smith and colleagues subsequently evaluated the National Early Warning Score and reported that it discriminated patients at risk of early cardiac arrest, unexpected intensive-care admission, and death. This supported standardized scoring as a clinically useful risk-stratification strategy, but later systematic evidence emphasized that predictive performance alone should not be confused with proven improvement in patient outcomes. Smith and colleagues' 2014 systematic review similarly found that early warning scores could demonstrate predictive ability while evidence concerning clinical implementation and outcomes was more limited.

3.2 Nursing Judgment, Escalation, and Alert Response

Jensen, Skår, and Tveit's integrative review found that early warning and rapid-response systems influenced nursing competence in three broad areas: assessment and care, referral and communication, and nurses' experiences of coping and mastery. Their synthesis also indicated that the impact was somewhat contradictory, reinforcing the point that structured scoring can support nursing practice without eliminating the relational, interpretive, and organizational dimensions of escalation.

Implementation evidence further demonstrates that technically valid scoring does not guarantee effective bedside use. Bedoya and colleagues evaluated the implementation of NEWS with an electronic best-practice alert in an academic and community hospital. They found poor performance characteristics in the deployed settings, extensive alert generation, frequent disregard by frontline nursing staff, and no appreciable improvement in their defined outcomes. Local refitting improved model performance. This study is particularly relevant ethically because repeated low-value alerts can alter clinician behavior toward subsequent warnings.

Gerry and colleagues' 2020 systematic review and critical appraisal examined early warning score development and validation studies and identified substantial methodological weaknesses. The authors cautioned that scores may not perform as well as expected and that poorly performing systems could adversely affect care. Their analysis demonstrates why bedside implementation should include independent validation rather than relying solely on reported development-study accuracy.

3.3 Fairness, Transparency, and Algorithmic Ethics

Ethical questions become especially visible when risk systems use complex models or proxies not apparent to clinicians. Char, Shah, and Magnus argued that machine-learning implementation requires attention to data quality, opacity, professional responsibility, and ethical consequences throughout clinical integration rather than only during model development. A predictive score may appear mathematically neutral while incorporating assumptions about which outcomes are important and which variables adequately represent clinical need.

Obermeyer and colleagues demonstrated this problem empirically in 2019. They found substantial racial bias in a widely used population-health algorithm because the system used health-care expenditure as a proxy for health need. Black patients had greater illness burden than White patients at equivalent algorithmic risk levels because historical expenditure did not represent health need equally across groups. Although this algorithm was not a bedside early warning score, the ethical lesson directly applies: a predictor can achieve apparently acceptable statistical performance while distributing benefit and error inequitably.

Wiens and colleagues therefore proposed a responsible machine-learning roadmap emphasizing problem formulation, representative data, meaningful evaluation, prospective testing, deployment monitoring, and attention to unintended consequences. WHO's earlier ethics consultation similarly emphasized privacy, governance, clinical decision support, and risks associated with decisions based on probabilistic and imperfect data. The research gap addressed by the present study is the translation of these broad ethical principles into a practical bedside nursing framework in which predictive scores influence observation, escalation, and distribution of nursing attention.

4. Research Methodology

4.1 Research Design and Simulated Clinical Dataset

The study adopts a simulation-based methodological design. A hypothetical dataset representing 1,000 adult inpatient observation episodes was generated to examine how changing a predictive-score escalation threshold could affect sensitivity and the proportion of patient encounters producing an alert. No real patient records, identifiable data, hospital information systems, or actual early warning score datasets were used.

The simulation does not reproduce NEWS, NEWS2, MEWS, or any proprietary clinical prediction model. Threshold values are illustrative numerical categories used only to demonstrate an ethical trade-off. They should therefore not be used for clinical decision-making or interpreted as recommendations for any existing scoring system.

A future empirical investigation would require local validation using representative inpatient data, predefined clinical outcomes, appropriate ethics and information-governance review, evaluation of missing observations, subgroup analysis, and comparison with existing escalation practice. Gerry and colleagues' critical appraisal underscores the importance of robust model-development and validation procedures before predictive scores are relied upon in clinical practice.

4.2 Ethical Bedside Use Framework

The proposed framework separates the predictive output from the clinical decision. The score provides a risk signal, while the nurse performs bedside assessment and determines whether the patient requires additional observations, senior nursing review, medical assessment, rapid-response activation, or another form of escalation according to local policy. Nursing concern can therefore increase escalation intensity even when the numerical score remains low.

A second principle concerns transparency. Nurses should know what event the score predicts, the observation window, the meaning of the threshold, and the expected response. The presence of a numerical output should not obscure uncertainty. Char and colleagues and Wiens and colleagues both emphasize that clinical implementation introduces ethical responsibilities that extend beyond model accuracy.

A third principle requires local and subgroup audit. Bedoya and colleagues found that NEWS performance differed across hospital settings and improved after site-specific refitting, while Obermeyer and colleagues demonstrated that algorithmic error can be distributed inequitably when apparently reasonable proxies encode existing disparities. Hospitals should consequently evaluate predictive performance by relevant patient subgroups rather than relying only on a single overall accuracy statistic.

Table 1: Proposed Ethical Governance and Bedside Nursing Assessment Framework

Domain / Code	Proposed assessment or questionnaire item	Ethical purpose	Required governance response
Clinical judgment	Nurses may escalate clinical concern regardless of calculated score	Prevent automation from suppressing bedside concern	Explicit override/escalation pathway
Score transparency	Nurses know what outcome and time horizon the score predicts	Support informed interpretation	Accessible score documentation
Missing data	System identifies incomplete or outdated observations	Prevent false reassurance from incomplete inputs	Missing-data warning and reassessment
Alert burden	Frequency and clinical usefulness of alerts are monitored	Reduce alert fatigue	Periodic threshold and workflow audit

Domain / Code	Proposed assessment or questionnaire item	Ethical purpose	Required governance response
Fairness	Performance is compared across relevant patient subgroups	Detect unequal error distribution	Bias audit and corrective action
Accountability	Every alert has a defined professional response pathway	Avoid ambiguity over responsibility	Named escalation roles and time expectations
EPN-1	“I understand what clinical outcome the predictive score is intended to estimate.”	Score comprehension	1–5 Likert response
EPN-2	“I can escalate care when I am clinically concerned even if the predictive score is low.”	Professional autonomy	1–5 Likert response
EPN-3	“I understand the limitations and possible errors of the predictive score used on my unit.”	Risk awareness	1–5 Likert response
EPN-4	“Predictive alerts on my unit occur at a frequency that allows meaningful clinical response.”	Alert usability	1–5 Likert response
EPN-5	“I know who is accountable for reviewing and acting on a high-risk predictive alert.”	Responsibility clarity	1–5 Likert response

Note: The five questionnaire items are proposed for future empirical research and have not been psychometrically validated or administered to nurses.

4.3 Analytical Strategy

The primary simulation examines two outcomes across six illustrative escalation thresholds: sensitivity for the simulated deterioration outcome and percentage of encounters generating a predictive alert. Sensitivity was used because failure to identify deteriorating patients creates a direct patient-safety concern. Alert burden was used because highly frequent alerts can consume clinical attention and can undermine responsiveness when a large proportion of alerts are not actionable.

The model intentionally does not identify a mathematically "optimal" threshold. Threshold selection involves values and trade-offs, including tolerance for missed deterioration, available rapid-response resources, nursing workload, and clinical consequences of unnecessary escalation. These decisions should therefore combine statistical analysis with professional, organizational, and ethical evaluation.

In a future empirical study, discrimination should be supplemented with calibration, sensitivity, specificity, positive and negative predictive values, decision-curve analysis where appropriate, subgroup performance, missing-data patterns, alert-response times, override frequency, and patient outcomes. Model performance should also be reassessed after implementation because clinical behavior can change once a prediction system enters routine workflow. Wiens and colleagues explicitly emphasize the importance of monitoring systems after deployment rather than treating validation as a one-time event.

5. Analysis

5.1 Simulated Predictive Performance and Alert Burden

The synthetic results demonstrate an inverse relationship between escalation threshold and sensitivity. At the lowest illustrative threshold of 3, sensitivity was modeled at 95%, meaning that most simulated deterioration episodes generated an alert. However, 49% of all encounters also produced an alert. Such a threshold could protect against missed deterioration while simultaneously creating a substantial workload for bedside nurses and response teams.

At the illustrative threshold of 5, sensitivity declined to 85% while the percentage of encounters generating alerts declined to 32%. Raising the threshold to 8 reduced alert generation to 11% but also reduced sensitivity to 51%. In ethical terms, the higher threshold decreases interruption and resource use at the cost of leaving a greater proportion of deteriorating patients below the automated escalation threshold.

Table 2: Simulated Ethical Trade-off Across Predictive-Score Thresholds

Illustrative escalation threshold	Simulated sensitivity for deterioration	Encounters generating alerts	Simulated deterioration cases missed	Ethical interpretation
3	95%	49%	5%	Very high sensitivity with substantial alert burden
4	91%	40%	9%	High sensitivity with frequent alerts
5	85%	32%	15%	Intermediate trade-off
6	76%	24%	24%	Reduced alert burden with greater missed-risk potential

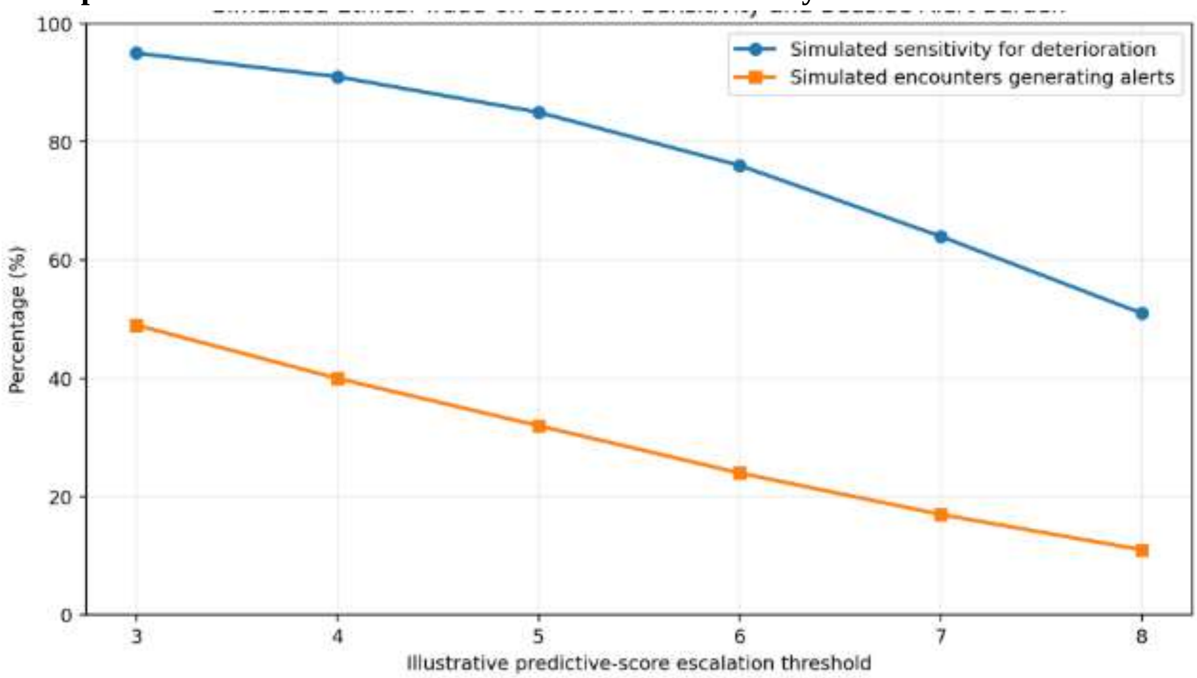
Illustrative escalation threshold	Simulated sensitivity for deterioration	Encounters generating alerts	Simulated deterioration cases missed	Ethical interpretation
7	64%	17%	36%	Low alert burden but substantial sensitivity loss
8	51%	11%	49%	Few alerts but ethically concerning missed deterioration

Note: All values are synthetic and do not represent NEWS, NEWS2, MEWS, or any validated predictive model. No clinical threshold should be selected from this table.

The analysis demonstrates why sensitivity and workload cannot be considered independently. Bedoya and colleagues' implementation experience showed that extremely frequent alerts can be disregarded by staff when perceived clinical value is low. Conversely, Gerry and colleagues cautioned that predictive scores with methodological weaknesses may fail to identify deterioration as reliably as clinicians expect. Ethical threshold selection therefore requires consideration of both missed deterioration and the practical consequences of alert saturation.

5.2 Ethical Threshold Trade-off

Graph 1: Simulated Ethical Trade-off Between Sensitivity and Bedside Alert Burden



The graph demonstrates that reducing alert frequency is not ethically neutral. As the illustrative escalation threshold rises, workload generated by predictive alerts decreases, but sensitivity falls more sharply at higher thresholds. A system designed primarily to minimize interruptions may therefore fail at its fundamental patient-safety purpose if clinically important deterioration remains undetected.

The opposite extreme also creates risk. When nearly half of all patient encounters generate an alert, the distinction between routine monitoring and exceptional risk can become difficult to sustain operationally. Implementation research demonstrates that poorly performing electronic warning systems may generate many alerts while producing little observable benefit in clinical outcomes. The ethically desirable threshold is therefore unlikely to be determined by a universal number; it must depend on validated local performance, response capacity, patient population, and clinical consequences of different errors.

A further implication is that an automated threshold cannot replace individualized escalation. A patient below the formal threshold may still require urgent attention when the nurse detects a new pattern of confusion, respiratory distress, unusual behavior, skin changes, sudden functional decline, or another clinically meaningful change. Nursing literature indicates that experiential recognition and intuitive concern are important elements in detecting deterioration.

5.3 Simulated Governance Outcomes

A secondary hypothetical analysis considered how implementation safeguards could change the ethical profile of a predictive-score system. In the simulated governance model, nurses were permitted to override low scores when clinically concerned, missing observations generated visible warnings, and alerts were periodically reviewed for usefulness. The model assumed that these safeguards would reduce the chance that a patient was falsely classified as safe simply because the algorithmic threshold had not been reached.

The simulated framework also treated overrides as data for quality improvement rather than evidence of noncompliance. Frequent nurse escalation in patients receiving low automated scores could indicate valuable nursing recognition, a calibration problem, missing model variables, or a patient subgroup for whom the score performs poorly. Such patterns should be investigated rather than suppressed.

Likewise, repeated high-risk alerts that do not lead to clinically meaningful intervention should trigger review of the model or threshold. Bedoya and colleagues demonstrated that performance can vary substantially between settings and that local refitting may improve predictive behavior. The ethical governance cycle should therefore treat implementation as a continuously monitored clinical intervention.

6. Discussion

6.1 Predictive Scores Should Augment, Not Replace, Nursing Judgment

The most important finding of the present study is conceptual rather than numerical: a predictive score is ethically appropriate only when it supports rather than displaces bedside assessment. Nursing recognition of deterioration involves repeated observation, contextual knowledge, communication with patients and families, interpretation of subtle change, and comparison with the patient's previous condition. Odell and colleagues found that intuition can play an important role in recognizing deterioration and that nurses may use vital signs to confirm an existing clinical concern. A system that requires nurses to wait for a score threshold before escalating concern would therefore invert the appropriate relationship between professional judgment and decision support.

This does not mean that intuitive nursing concern should be considered infallible. Structured scores can reduce some forms of inconsistency and provide a common language for communicating physiological risk. Smith and colleagues demonstrated that NEWS could discriminate patients at elevated risk of serious outcomes, supporting the value of standardized physiological assessment. Jensen and colleagues likewise found that early warning and rapid-response systems can contribute positively to nursing competence and referral processes. The ethical model is consequently additive: structured prediction and professional assessment should inform one another.

Bedside policies should make this relationship explicit. A high score should mandate the appropriate response defined by policy, but a low score should never prohibit escalation when nursing concern persists. Documentation can capture both numerical risk and clinical rationale, allowing organizations to learn from disagreements between algorithmic output and bedside judgment.

6.2 Fairness, Missing Data, and the Illusion of Objectivity

Predictive scores can appear neutral because they are represented numerically, yet every model embodies choices concerning input variables, target outcomes, thresholds, missing-data handling, and populations used for development. Obermeyer and colleagues' analysis of a widely used health-care algorithm demonstrated how a seemingly reasonable proxy—health-care expenditure—generated substantial racial bias because expenditure reflected unequal access and use rather than health need alone. The broader ethical lesson is that a model can be technically sophisticated and statistically accurate according to one metric while remaining unfair in clinically meaningful ways.

Bedside systems therefore require subgroup analysis. Performance should be examined where clinically and ethically relevant across age groups, sex, disease groups, chronic respiratory conditions, disability, language or communication barriers, and other characteristics for which physiology, documentation, or care patterns may differ. The objective is not to assume bias in advance but to test whether sensitivity, false-positive rates, calibration, or missed deterioration differ materially among populations.

Missing data create another important ethical problem. Ludikhuize and colleagues found incomplete vital-sign recordings in patients who later experienced serious deterioration. A digital score calculated from incomplete or outdated measurements may present a precise number even when the information supporting that number is inadequate. Systems should therefore distinguish low risk from insufficient information to estimate risk reliably. Displaying these states identically can create false reassurance.

6.3 Alert Fatigue, Accountability, and Continuous Governance

The simulated graph demonstrates that reducing false negatives generally requires more alerts, while reducing alerts can increase missed deterioration. This is an ethical resource-allocation problem because nurses possess limited attention and must simultaneously respond to medication administration, patient communication, documentation, procedures, monitoring, and acute clinical events. A predictive system that generates frequent low-value alerts can compete with rather than support patient care.

Bedoya and colleagues provide an instructive real-world example. Their implemented system generated a very large number of best-practice advisories, was generally ignored by frontline nurses, and produced no appreciable improvement in the study's defined outcomes. The finding does not imply that predictive alerts are ineffective in general. It demonstrates that alert frequency, local calibration, workflow fit, and clinical credibility are integral components of effectiveness.

Accountability must also remain explicit. Char and colleagues noted that algorithmic clinical decision support creates questions concerning professional and institutional responsibility, while Wiens and colleagues emphasized monitoring throughout the deployment lifecycle. Hospitals should therefore define who receives an alert, who must assess the patient, how quickly action is expected, how unresolved alerts are escalated, and what happens when the predictive system fails. Ethical governance is weakened when responsibility becomes diffused between the nurse, software vendor, physician, rapid-response team, and hospital administration.

7. Conclusion

Predictive scores can strengthen bedside nursing by organizing physiological information, standardizing communication, and identifying patients who may require closer observation or escalation. Research concerning NEWS and other early warning systems demonstrates meaningful predictive potential, while nursing literature supports their contribution to structured recognition and communication of clinical deterioration. Their value, however, depends on appropriate implementation rather than the mere presence of a numerical score in the electronic health record.

The present simulation illustrates the ethical tension inherent in escalation thresholds. Lower thresholds can preserve sensitivity while generating substantial alert burden, whereas higher thresholds can reduce interruptions at the cost of missing a greater proportion of deteriorating patients. No universal numerical threshold can be ethically derived from the simulation, and the presented values must not be used for clinical decision-making. Local validation, response capacity, patient population, and consequences of false-negative and false-positive classifications must determine how an actual score is governed.

Ethical bedside use also requires protection against automation bias. Nurses must retain the authority to escalate concern when the patient's clinical presentation conflicts with the score. Predictive systems should identify missing or outdated inputs, and institutions should analyze cases in which nursing concern correctly identifies deterioration despite a reassuring prediction. Such cases provide valuable information about model limitations and the continuing importance of professional assessment.

Finally, predictive-score governance must include fairness, transparency, accountability, alert monitoring, and post-implementation evaluation. Evidence concerning algorithmic bias demonstrates that models can distribute error unevenly even when overall performance appears acceptable, while real-world warning-score implementation demonstrates that poorly calibrated alert systems can be ignored by frontline clinicians. The safest and most ethically defensible role for predictive scoring in bedside nursing is therefore one of augmented clinical judgment: the score contributes structured evidence, the nurse contributes contextual assessment and professional reasoning, and the health-care organization remains responsible for ensuring that the combined system is validated, equitable, understandable, and continuously monitored.

References

1. Heyen, N. B., & Salloch, S. (2021). The ethics of machine learning-based clinical decision support: An analysis through the lens of professionalisation theory. *BMC Medical Ethics, 22*, 112. https://doi.org/10.1186/s12910-021-00679-3.
2. Seibert, K., et al. (2021). Application scenarios for artificial intelligence in nursing care: Rapid review. *Journal of Medical Internet Research, 23*(11), e26522. https://doi.org/10.2196/26522.
3. World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. World Health Organization.
4. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679.
5. Char, D. S., Abràmoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. *The American Journal of Bioethics, 20*(11), 7–17. https://doi.org/10.1080/15265161.2020.1819469.
6. Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A systematic review. *Social Science & Medicine, 260*, 113172. https://doi.org/10.1016/j.socscimed.2020.113172.
7. Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine, 15*(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689.
8. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7.
9. Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering, 2*, 719–731. https://doi.org/10.1038/s41551-018-0305-z.
10. Paulus, J. K., & Kent, D. M. (2020). Predictably unequal: Understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. *npj Digital Medicine, 3*, 99. https://doi.org/10.1038/s41746-020-0304-9.
11. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342.
12. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine, 169*(12), 866–872. https://doi.org/10.7326/M18-1990.
13. O'Reilly-Shah, V. N., Gentry, K. R., Walters, A. M., Zivot, J., Anderson, C. T., & Tighe, P. J. (2020). Bias and ethical considerations in machine learning and the automation of perioperative risk assessment. *British Journal of Anaesthesia, 125*(6), 843–846. https://doi.org/10.1016/j.bja.2020.07.040.
14. Grote, T., & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics, 46*(3), 205–211. https://doi.org/10.1136/medethics-2019-105586.

15. Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*, 310. https://doi.org/10.1186/s12911-020-01332-6.
16. McGaughey, J., Fergusson, D. A., Van Bogaert, P., & Rose, L. (2021). Early warning systems and rapid response systems for the prevention of patient deterioration on acute adult hospital wards. *Cochrane Database of Systematic Reviews*, *11*, CD005529. https://doi.org/10.1002/14651858.CD005529.pub3.
17. Smith, V., Kenny, L. C., Sandall, J., & Devane, D. (2021). Physiological track-and-trigger/early warning systems for use in maternity care. *Cochrane Database of Systematic Reviews*, *9*, CD013276. https://doi.org/10.1002/14651858.CD013276.pub2.
18. Sendak, M. P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., & Ratliff, W. (2020). A path for translation of machine learning products into healthcare delivery. *EMJ Innovations, 4*(1), 43–49.
19. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine*, *25*, 1337–1340. https://doi.org/10.1038/s41591-019-0548-6.
20. Braun, M., Hummel, P., Beck, S., & Dabrock, P. (2021). Primer on an ethics of AI-based decision support systems in the clinic. *Journal of Medical Ethics, 47*(12), e3. https://doi.org/10.1136/medethics-2020-106820.