PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Predicting Pressure Injury Risk in Hospitalised Patients using Machine Learning with Electronic Health Records: A U.S. Multilevel Cohort Study
AUTHORS	Padula, William; Armstrong, David; Pronovost, Peter; Saria, Suchi

VERSION 1 – REVIEW

REVIEWER	Alderden, Jenny Boise State University School of Nursing
REVIEW RETURNED	13-Dec-2023

GENERAL COMMENTS	I am grateful for the opportunity to review this manuscript, which focuses on developing a risk prediction model for serious, reportable pressure injuries (stages 3, 4, and unstageable). Below are my detailed observations and queries: Definition Clarity: The manuscript's definition of pressure injuries, central to the study, is not immediately clear in the text but rather confined to Table 1 (stages 3, 4, and unstageable only). A direct explanation within the manuscript would enhance reader comprehension. Updated Pressure Injury Definitions: The authors the 2015 PS103 definition but should consider reporting on the revised PS103 pressure injury definition as of 2021, which is quite different: “Stage III or IV pressure ulcers or unstageable (secondary diagnosis) per 1,000 discharges among surgical or medical patients ages 18 years and older. Excludes stays less than 3 days; cases with a principal stage III or IV (or unstageable) pressure ulcer diagnosis; cases with a secondary diagnosis of stage III or IV pressure ulcer (or unstageable) that is present on admission; obstetric cases; severe burns; exfoliative skin disorders.” See: https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/v2020/TechSpecs/PSI_03_Pressure_Ulcer_Rate.pdf Inclusion Criteria and Risk Factors: The manuscript exhibits a discrepancy in the inclusion criteria. While the 2015 PS103 definition (lines 18-22) excludes spinal cord injuries (hemiplegia, paraplegia, quadriplegia), the risk score developed includes spinal cord injury (Table 2 and line 30). This inconsistency necessitates clarification. Malnutrition Assessment: The study cites malnutrition as an important component of the risk model but lacks any established indicators of malnutrition. The assumption linking hyperlipidemia to malnutrition (hyperlipidemia = protective) is unconventional and not supported by /[standard dietary guidelines, requiring further elaboration. Mobility Assessment: Which variable is used to indicate immobility? I see ‘bed confinement’ but that variable only has 49 people (~0.002% of the study sample). Data Timeframe: The specific period from which the EHR data were sourced is not mentioned. Detailing the exact years is crucial, especially to understand the context of preventive measures during that time and the unique pressure injury risk factors associated with COVID-19 since 2020.

	Vasopressor Reference Accuracy: The manuscript refers to 'vasopressins' in line 30 and Table 2. It should be noted that vasopressin is a specific drug, a second-line vasopressor, and this reference may require correction for accuracy. Target Leakage: Using ICD discharge diagnosis codes in the predictive model (line 51, Table 2) could introduce target leakage if applied at the end of hospitalization, where the coder is aware of the patient's entire hospitalization narrative. Paradoxical Findings: The relationships in Table 2 present several paradoxes. For instance, stool cultures, typically indicative of diarrhea (a known risk factor for pressure injuries), are oddly shown as protective, whereas urinalysis is a risk factor. The clinical distinction between electrolyte (protective) and phosphate (risk factor) replacements is unclear. Moreover, the exceptionally high reported odds ratio for bed confinement (given only 49 patients were coded with this condition out of approximately 21,000) and spinal cord injury (just 99/21000 patients, odds ratio of 183131.00) call for a more in-depth examination of the data. Is it possible that the features are highly correlated and thus not well suited to parametric analysis? I only see one type of correlation — autocorrelation of risk—addressed, not correlation between predictor variables. Most pressure injury prediction models are cross sectional and thus I would like to commend the authors for considering pressure injury risk as a time-varying problem.

REVIEWER	Seth, Puneet McMaster University, Family Medicine
REVIEW RETURNED	22-Dec-2023

GENERAL COMMENTS	This is a well-designed study on a relevant topic, and the paper is additionally very well written. It represents important work in making improvements to both patient lives as well as to healthcare system sustainability. There are two points however I feel should be considered for incorporation into the Discussion: 1. The underlying premise of the study supposes that there is risk stratification or some kind of alert system to notify clinical staff of the need for further monitoring for specific patients. It is worthwhile to note that the current reality of care delivery with EHRs involves a great deal of user dissatisfaction and "alert fatigue" based on information flowing from the EHR. How this risk stratification gets implemented is arguably as important as the accuracy and validity of the tool itself. Furthermore, note that in order for such a tool to work, data entry needs to be standardized into the EHR, which may not be happening and the requirement of which may further contribute to clinical staff burnout. The authors need not dive into design considerations, but at least mention the important of thoughtful integration of such risk scales without further overwhelming clinical staff. 2. The Braden Scale itself has been demonstrated to have greater predictive validity in Caucasian patients (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8363405/). While the 2nd limitation briefly touches this subject ("Second, the retrospective nature of accessing these clinical data makes it difficult to verify the accuracy of Braden scores and other clinical data with respect to what was the true condition of the patient being observed."), it is worth pointing how that such models inherently risk amplifying bias in clinical decision making and may ingrain systems that result in poorer outcomes for many ethnic populations.

REVIEWER	Khanh Le, Nguyen Quoc Taipei Medical University

REVIEW RETURNED	27-Dec-2023

GENERAL COMMENTS	1. The study selects multilevel modeling methods with predictors determined by random forests. A detailed rationale for choosing this specific combination over other machine learning approaches is crucial for transparency and reproducibility. 2. The study relies on data from two academic medical center databases, but the representativeness of these datasets to the broader population is not explicitly addressed. The external validity of the developed algorithm might be compromised if the dataset is not diverse or representative of various healthcare settings. 3. The inclusion and exclusion criteria, particularly those related to Patient-Safety Indicator #3 (PSI03), may introduce biases. The exclusion of patients with certain conditions and the requirement for a minimum length-of-stay could impact the generalizability of the predictive model to a broader patient population. 4. While the use of the Braden Scale is common and aligns with clinical guidelines, the study doesn't explicitly discuss any limitations or challenges associated with this specific risk assessment measure. 5. The study describes the use of logistic regression and random forests for feature selection. However, the criteria for selecting specific features and the potential impact of feature choices on model performance are not explicitly discussed. 6. The use of multilevel logistic regression introduces complexity. The rationale for choosing a mixed-effects model over fixed-effects and the assumption of linearity in subject-specific risk trajectories should be carefully justified. Additionally, the choice of covariance structures and the handling of non-linearity should be thoroughly explained and tested. 7. The study uses sensitivity, specificity, ROC curve, and AUROC for model evaluation. While these are common metrics, the study could benefit from discussing the clinical implications of these metrics and their relevance in the context of pressure injury prevention. 8. Cross-validation is employed, but it would be essential to discuss measures taken to avoid overfitting, especially with the mention of 20,000 iterations in the feature selection process. 9. More references on machine learning model applying in biomedical informatics should be added to attract a broader readership i.e., PMID: 33848577, PMID: 37120403. 10. The financial impact analysis is based on assumptions about the algorithm's compliance and its impact on resource utilization. The study should transparently discuss these assumptions and acknowledge potential uncertainties, considering that real-world compliance may vary. 11. A more comprehensive explanation of the financial impact and the assumptions behind the cost savings estimation is necessary. 12. While the study links pressure injury risk to issues like mobility, nutrition, and moisture management, it assumes a high-risk trajectory for certain patients based on limited factors. 13. The study could delve deeper into the potential biases and inaccuracies introduced by the retrospective nature of accessing clinical data, affecting the accuracy of Braden scores and other clinical information. 14. The study does not elaborate on conducting external validation or potential challenges associated with generalizing the model to other healthcare institutions.

	15. The study could explore the potential impact of these unobserved factors on the model's performance and discuss strategies for addressing these limitations in future iterations. 16. While the study highlights the improvement in sensitivity of prescriptive care, it would be beneficial to elaborate on how this sensitivity is measured and the potential implications for patient outcomes. Additionally, discussing the balance between sensitivity and specificity in the context of prescriptive care would add nuance to the interpretation. 17. The study suggests that pressure injury prevention and the integration of the predictive algorithm with EHRs may be the starting point for health systems. However, the challenges and considerations related to the seamless integration of the algorithm into existing EHR systems are not addressed. 18. While the study uses clinical judgment alongside the predictive algorithm, it could elaborate on the role of clinicians in the decision-making process. 19. The study suggests starting a technology-driven quality improvement program with predictive analytics, but it does not provide guidance on the challenges and considerations in implementing such programs.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

4. *Definition Clarity: The manuscript's definition of pressure injuries, central to the study, is not immediately clear in the text but rather confined to Table 1 (stages 3, 4, and unstageable only). A direct explanation within the manuscript would enhance reader comprehension.*

We have included a definition of “pressure injury” consistent with the NPIAP revised staging system (Edsberg et al., JWOCN 2016):

Methods, Study Population: “A pressure injury is caused by damage to the skin and underlying tissue as a result of pressure and shearing forces; pressure injuries are measured in stages from 1-4, including unstageable, to describe advancement from low-stage pressure injury denoted by bruising and erythemia, to high-stage pressure injuries that include open ulcerations that reach bone and muscle fascia.”

5. *Updated Pressure Injury Definitions: The authors the 2015 PS103 definition but should consider reporting on the revised PS103 pressure injury definition as of 2021, which is quite different: “Stage III or IV pressure ulcers or unstageable (secondary diagnosis) per 1,000 discharges among surgical or medical patients ages 18 years and older. Excludes stays less than 3 days; cases with a principal stage III or IV (or unstageable) pressure ulcer diagnosis; cases with a secondary diagnosis of stage III or IV pressure ulcer (or unstageable) that is present on admission; obstetric cases; severe burns; exfoliative skin disorders.”*
See: https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/v2020/TechSpecs/PSI_03_Pressure_Ulcer_Rate.pdf

Thank you for this comment. We agree and have updated the manuscript to reflect the latest version of PSI03 (see citation #23). We have re-written this section as you have presented it from PSI03 definition online from AHRQ.

Methods, Study Population: “PSI03 inclusion criteria for hospital-acquired pressure injury cases are as follows: stage 3, 4 or unstageable pressure injury not present-on-admission (not as a primary or secondary diagnosis: ICD9 = 707.x; ICD10 = L89.xx) per 1,000 discharges among surgical and medical patients ages 18 years and older; excludes stays less than 3 days and certain diagnosis or procedural codes that place them at predisposed risk to skin injury: certain skin conditions (MDC 9); pregnancy (MDC 14); spinal cord injury (hemiplegia, paraplegia, or quadriplegia); spina bifida; pedicled graft or debridement; and a transfer between different facilities.”

6. *Inclusion Criteria and Risk Factors: The manuscript exhibits a discrepancy in the inclusion criteria. While the 2015 PS103 definition (lines 18-22) excludes spinal cord injuries (hemiplegia, paraplegia, quadriplegia), the risk score developed includes spinal cord injury (Table 2 and line 30). This inconsistency necessitates clarification.*

We made some minor modifications to the inclusion criteria for our study cohort. These modifications are consistent with previous research that we cited (see citation #10). We have added clarity to the modifications from PSI03 in the Methods to support the resulting information in Table 2.

Methods, Study Population: “We modified some specifications of the PSI03 criteria to develop a predictive model that was inclusive of some high-risk patient populations, as consistent with prior informatics research on this patient population. For instance, we included patients that were hospitalized for at least 5 days and obtained 2 or more risk scores during admission. We also included patients with spinal cord injury in the study cohort since providers could benefit from a greater understanding of the association between risk and pressure injury, based on previous findings linking these outcomes.”

7. *Malnutrition Assessment: The study cites malnutrition as an important component of the risk model but lacks any established indicators of malnutrition. The assumption linking hyperlipidemia to malnutrition (hyperlipidemia = protective) is unconventional and not supported by /[standard dietary guidelines, requiring further elaboration.*

Thank you for this comment. We do not intend to make the assertion that hyperlipidemia is a proxy for malnutrition, and sorry if it may have been interpreted as such. We agree that malnutrition is best assessed using a validated tool, as consistent with the NPIAP SPIPP 2.0 Checklist. Unfortunately the EHR we worked with did not have composite measures of nutrition from one of these validated tools. Our assertion is that the lab orders (e.g. pre-albumin lipid panel (Table 2)) represent best available data in this EHR and could proxy for a provider’s desire to examine nutritional levels, that is all. We have edited the paper to make these assertions more clear.

Discussion, Limitations: Sixth, the NPIAP Standardized Pressure Injury Prevention Protocol (SPIPP Checklist) recommends the use of validated tools to measure nutrition and mobility levels. While such tools were not captured in the EHR used in this study, our study team and PURPOSE Advisory Council reviewed resulting predictors from the random forest that could be helpful proxies for nutrition (e.g. lab orders for pre-albumin and lipid panels) and mobility (e.g. bed confinement and spinal cord injury). Future research could improve upon this analysis by including validated tools for mobility and nutrition in predictive analytics.

8. *Mobility Assessment: Which variable is used to indicate immobility? I see ‘bed confinement’ but that variable only has 49 people (~0.002% of the study sample).*

See response to Comment #7.

9. *Data Timeframe: The specific period from which the EHR data were sourced is not mentioned. Detailing the exact years is crucial, especially to understand the context of preventive measures during that time and the unique pressure injury risk factors associated with COVID-19 since 2020.*

We have included the timeframe of the data in the Methods/Data Source, which pre-date COVID-19 pandemic to minimize risk of bias during that period: “An observational cohort of hospitalized patient EHRs over 5 years was used to obtain a substantial sample of calibration data for analytics from 2014 through 2019.”

10. *Vasopressor Reference Accuracy: The manuscript refers to ‘vasopressins’ in line 30 and Table 2. It should be noted that vasopressin is a specific drug, a second-line vasopressor, and this reference may require correction for accuracy.*

Thank you for catching this typo. We mean to refer broadly to “Vasopressor” throughout the paper, and have made this modification in the Methods and Table 2.

11. *Target Leakage: Using ICD discharge diagnosis codes in the predictive model (line 51, Table 2) could introduce target leakage if applied at the end of hospitalization, where the coder is aware of the patient's entire hospitalization narrative.*

Thank you for this comment. Our data contains diagnostic codes and other EHR data entered into patient charts in real-time. To your point, perhaps earlier diagnoses could bias later ones (e.g. beyond the primary and secondary diagnosis). We have included this as a possible limitation: “Fourth, the study uses codes entered into a patient EHR in real-time. There is potential bias in the entry of latter codes based on earlier diagnoses, prescriptions and orders.”

12. *Paradoxical Findings: The relationships in Table 2 present several paradoxes. For instance, stool cultures, typically indicative of diarrhea (a known risk factor for pressure injuries), are oddly shown as protective, whereas urinalysis is a risk factor. The clinical distinction between electrolyte (protective) and phosphate (risk factor) replacements is unclear. Moreover, the exceptionally high reported odds ratio for bed confinement (given only 49 patients were coded with this condition out of approximately 21,000) and spinal cord injury (just 99/21000 patients, odds ratio of 183131.00) call for a more in-depth examination of the data. Is it possible that the features are highly correlated and thus not well suited to parametric analysis? I only see one type of correlation — autocorrelation of risk—addressed, not correlation between predictor variables.*

Thank you for this comment. We agree there is a lot to unpack of these data. We tested for autocorrelation in the model, as well as non-parametric (i.e “quadratic”) random-effects, but neither of those tests returned information that would suggest independent covariance in the underlying fit of the model (see **Figure 2**). From an analytical standpoint, we are confident that these data “do not lie”, and therefore, could do more to explain potential directionality of these findings. Keeping in mind these are coefficients based on risk rather than actual PrI outcomes, we have included a new paragraph in the discussion to address these points in order.

Discussion, Limitations: “Eighth, a number of unique findings in this predictive algorithm could be considered paradoxical. Stool culture orders were associated with reduced risk, perhaps because despite having risk of bacterial infection and diarrhea, ordering this lab within 5 days of admission places patients on a correct path to reduced pressure injury risk. Prescribed electrolyte replacement was protective, whereas phosphate replacement increased risk, perhaps differentiating the underlying risk factors of patients need one replacement compared to the other. Exceptionally high odds ratios

for bed confinement and spinal cord injury, despite small sample sizes, are worth additional cross-validation with other data samples given the small sample sizes of these data. While there is potential that some of these sets of predictors are highly correlated, our test for autocorrelation or non-parametric fit of the underlying covariance structure of the model returned no alternative findings."

Reviewer 2

13. *The underlying premise of the study supposes that there is risk stratification or some kind of alert system to notify clinical staff of the need for further monitoring for specific patients. It is worthwhile to note that the current reality of care delivery with EHRs involves a great deal of user dissatisfaction and "alert fatigue" based on information flowing from the EHR. How this risk stratification gets implemented is arguably as important as the accuracy and validity of the tool itself. Furthermore, note that in order for such a tool to work, data entry needs to be standardized into the EHR, which may not be happening and the requirement of which may further contribute to clinical staff burnout. The authors need not dive into design considerations, but at least mention the important of thoughtful integration of such risk scales without further overwhelming clinical staff.*

Thank you for this comment, we fully agree with these potential shortcomings of health system integration of these types of predictive models with EHRs. We have made mention of both points in the Discussion.

Discussion: "That being said, health systems should be prepared to adjust their culture to the use of these tools – overcoming barriers such as alarm fatigue brought on by risk notifications, and common data models to perform predictive analytics using the longitudinal data structure required by this predictive algorithm are critical to successful implementation."

14. *The Braden Scale itself has been demonstrated to have greater predictive validity in Caucasian patients (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8363405/). While the 2nd limitation briefly touches this subject ("Second, the retrospective nature of accessing these clinical data makes it difficult to verify the accuracy of Braden scores and other clinical data with respect to what was the true condition of the patient being observed."), it is worth pointing how that such models inherently risk amplifying bias in clinical decision making and may ingrain systems that result in poorer outcomes for many ethnic populations.*

Thank you for this very important point. Our lack of access to data on race/ethnicity in the EHRs used for this study potentially bias our calibration data as well. We have noted this concern in the **Limitations**: "Fifth, given known biases of the Braden Scale towards Caucasian skin tones, there may be inherent bias in this predictive algorithm based on its calibration data, unless future studies of this algorithm can control for race and ethnicity."

Reviewer: 3

15. *The study selects multilevel modeling methods with predictors determined by random forests. A detailed rationale for choosing this specific combination over other machine learning approaches is crucial for transparency and reproducibility.*

See response to Comment #2. In addition, we have written the following in the **Methods, Study Design**: "Multilevel modeling offered improved individual-specific prediction capabilities relative to other methods tested with the data at hand. This approach also offered transparency to clinicians who

may elect to use predictive analytics combined with clinical practice guidelines, such as validated risk assessment tools (e.g. the Braden Scale), in order to make more informed decisions about patient care."

16. *The study relies on data from two academic medical center databases, but the representativeness of these datasets to the broader population is not explicitly addressed. The external validity of the developed algorithm might be compromised if the dataset is not diverse or representative of various healthcare settings.*

See response to Comment #14.

17. *The inclusion and exclusion criteria, particularly those related to Patient-Safety Indicator #3 (PSI03), may introduce biases. The exclusion of patients with certain conditions and the requirement for a minimum length-of-stay could impact the generalizability of the predictive model to a broader patient population.*

See response to Comments #5 and #6.

18. *While the use of the Braden Scale is common and aligns with clinical guidelines, the study doesn't explicitly discuss any limitations or challenges associated with this specific risk assessment measure.*

See response to Comment #14.

19. *The study describes the use of logistic regression and random forests for feature selection. However, the criteria for selecting specific features and the potential impact of feature choices on model performance are not explicitly discussed.*

We ran multiple feature selection methods, and ran comparative tests of performance. Random forests out-performed other methods. To further answer your question, we included features in the final algorithm that were statistically significant. We have made mention of this in the **Methods**: "To develop the predictive algorithm, the model necessitated features that captured patient-specific measurements of risk for pressure injury, as well as weighted estimates (i.e. regression coefficients) using machine learning. Features that would serve as predictors were obtained using a variety of supervised methods (i.e. random forests, support vector machines, boosting, bagging, LASSO and Ridge regression). After testing these methods, we determined that random forests provided relevant predictors at reduced mean-squared error compared to other approaches. Predictors for the algorithm were derived from random forests applied to randomly selected groups of 1,000 observations at a time, and regressing 10 variables in each run. We simulated this process for 20,000 iterations, and maintained predictors in the final model that remained statistically significant ($p<0.05$)."

20. *The use of multilevel logistic regression introduces complexity. The rationale for choosing a mixed-effects model over fixed-effects and the assumption of linearity in subject-specific risk trajectories should be carefully justified. Additionally, the choice of covariance structures and the handling of non-linearity should be thoroughly explained and tested.*

Thank you for your comment. We performed a chi-square test of the log-likelihood ratios between FE and RE models, and rejected the FE model based on this test. This test is consistent with recommended comparative test according to Hedeker and Gibbons (see citation #21).

We tested multiple covariance structures, which were rejected in favor of independent covariance. No signs of autocorrelation. See response to Comment #12, and see Figure 2.

Methods: “LLR tests can be conducted to estimate the degree of improved fit between models, therefore we used chi-square tests of the LLRs to differentiate predictive performance between fixed- and random-effects models.21 We also used LLR to classify the degree of enhanced predictive accuracy of models with additional data.”

21. *The study uses sensitivity, specificity, ROC curve, and AUROC for model evaluation. While these are common metrics, the study could benefit from discussing the clinical implications of these metrics and their relevance in the context of pressure injury prevention.*

Thank you for your comment. We have included the following in the Discussion section in respoinse: “A predictive algorithm that can direct care efficiently, and with greater sensitivity, specificity and AUROC than standard risk measures and subjective clinical judgement may increase the likelihood that providers will follow-through on prevention guidelines when prescribed. Health systems should consider using predictive algorithms that introduce significant gains in AUROC, which represent a balance in gain of sensitivity and specificity, over the Braden Scale alone.”

22. *Cross-validation is employed, but it would be essential to discuss measures taken to avoid overfitting, especially with the mention of 20,000 iterations in the feature selection process.*

Thank you for this comment. We also performed regularization of these data prior to analytics following a previously published method (see citation #22). We have included this in our Method, which is an accepted approach to further protect against overfitting.

Methods, Model Development: “Given the large sample size, we first applied established normalization and regularization techniques to the data structure to protect this analysis for risk of overfitting, in addition to cross-validation.”

23. *More references on machine learning model applying in biomedical informatics should be added to attract a broader readership i.e., PMID: 33848577, PMID: 37120403.*

Thank you for these suggestions. We have included these refernces in the revision.

24. *The financial impact analysis is based on assumptions about the algorithm's compliance and its impact on resource utilization. The study should transparently discuss these assumptions and acknowledge potential uncertainties, considering that real-world compliance may vary.*

Thank you for this comment. The assumptions are now detailed in the methods, which are consistent with the assumptions of previously published economic models of pressure injury prevention now cited in this paper (see citations 29 and 30).

Methods, Financial Impact: “A typical hospitalized patient requires a routine that includes nursing time and material resources that can add up to $99.44 USD per patient per day on average. Thus, a large 500-bed hospital could spend up to $50,000 per day, and $18 million per year on this standard practice. Based on the assumption that a predictive algorithm could provide prescriptive follow-up for a select group of about only 25% of a typical patient population, we used an existing economic model in order to calculate the budget impact that this algorithm could save hospitals.[29] We assumed that the budget impact was based on 100% compliance with an algorithm-driven protocol, adjusting for the sensitivity and specificity of the algorithm. These economic modeling assumptions are consistent with previous economic evaluations on the cost-effectiveness of pressure injury prevention.”

25. *A more comprehensive explanation of the financial impact and the assumptions behind the cost savings estimation is necessary.*

We have provided a careful explanation of the economic findings to the best of our ability given the space constraints established by BMJ Open.

Results, Financial Impact: "The predictive algorithm offers improved economic efficiency, such that a hospital could recoup substantial savings on a weekly basis. Since a risk assessment can take anywhere from 5-15 minutes per patient, this could represent up to 250 labor-hours in a single 500-bed facility per day, and between 30,000 and 90,000 labor-hours per year. An average 500-bed hospital would spend up to $99.44 per patient per day on follow-up preventive tasks. For a hospital at 100% volume capacity, this represents a weekly investment of about $348,000, or $18 million annually. By comparison, a predictive algorithm that limits follow-up after the 5th time point reduces costs by over 48%; reduced follow-up after the 6th time point still can save as much as 42%. The 6% gain in economic efficiency between the 5th and 6th time point could represent a savings of about $975,000 annually without substantial losses in sensitivity or specificity."

26. *While the study links pressure injury risk to issues like mobility, nutrition, and moisture management, it assumes a high-risk trajectory for certain patients based on limited factors.*

See response to Comments #7 and #8.

27. *The study could delve deeper into the potential biases and inaccuracies introduced by the retrospective nature of accessing clinical data, affecting the accuracy of Braden scores and other clinical information.*

See response to Comment #11. In addition, we have included the following in the **Limitations**: "Second, the retrospective nature of accessing these clinical data makes it difficult to verify the accuracy of Braden scores and other clinical data with respect to what was the true condition of the patient being observed."

28. *The study does not elaborate on conducting external validation or potential challenges associated with generalizing the model to other healthcare institutions.*

See response to Comments #2, #7, and #14.

29. *The study could explore the potential impact of these unobserved factors on the model's performance and discuss strategies for addressing these limitations in future iterations.*

See response to Comment #2.

30. *While the study highlights the improvement in sensitivity of prescriptive care, it would be beneficial to elaborate on how this sensitivity is measured and the potential implications for patient outcomes. Additionally, discussing the balance between sensitivity and specificity in the context of prescriptive care would add nuance to the interpretation.*

Thank you for the suggestion. We have included the following paragraph in the **Methods**: "In this analysis, improved sensitivity represents an increase in the identification of patients at-risk of a pressure injury, given that risk actually does exist (i.e. true positive); specificity represents a reduction in the designation of at-risk patients who are not in fact considered at-risk (i.e. false positive). This balance in improved sensitivity and sensitivity leads to health system efficiency since clinicians would be providing follow-up care to patients that actually have a higher likelihood of pressure injury risk."

31. *The study suggests that pressure injury prevention and the integration of the predictive algorithm with EHRs may be the starting point for health systems. However, the challenges and considerations related to the seamless integration of the algorithm into existing EHR systems are not addressed.*

See response to Comment #13.

32. *While the study uses clinical judgment alongside the predictive algorithm, it could elaborate on the role of clinicians in the decision-making process.*

See response to Comment #7. We provided additional explanation of the PURPOSE Advisory Council's role in this process.

33. *The study suggests starting a technology-driven quality improvement program with predictive analytics, but it does not provide guidance on the challenges and considerations in implementing such programs.*

See response to Comment #13.

VERSION 2 – REVIEW

REVIEWER	Alderden, Jenny Boise State University School of Nursing
REVIEW RETURNED	20-Feb-2024

GENERAL COMMENTS	Thank you for the opportunity to review your interesting paper. I enjoyed reading your revision and I have no further questions or suggestions.

REVIEWER	Khanh Le, Nguyen Quoc Taipei Medical University
REVIEW RETURNED	22-Feb-2024

GENERAL COMMENTS	My previous concerns have been addressed.