medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Risk prediction tools for pressure injury occurrence: An umbrella review of systematic reviews reporting model development and validation methods

Bethany Hillier[1,2]

Katie Scandrett[1]

April Coombe[1,2]

Tina Hernandez-Boussard[3]

Ewout Steyerberg[4]

Yemisi Takwoingi[1,2]

Vladica Velickovic[5,6]

Jacqueline Dinnes[1,2*]

Affiliations

[1] Department of Applied Health Sciences, College of Medicine and Health, University of Birmingham, Edgbaston, Birmingham, UK

[2] NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK

[3] Department of Medicine, Stanford University, Stanford, CA USA

[4] Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

[5] Evidence Generation Department, HARTMANN GROUP, Heidenheim, Germany

[6] Institute of Public Health, Medical, Decision Making and Health Technology Assessment, UMIT, Hall, Tirol, Austria

Email addresses

b.hillier@bham.ac.uk (BH); k.e.scandrett@bham.ac.uk (KS); a.r.coombe@bham.ac.uk (AC); boussard@stanford.edu (THB); e.w.steyerberg@lumc.nl (ES); y.takwoingi@bham.ac.uk (YT); vladica.velickovic@hartmann.info (VV)

* Corresponding author: j.dinnes@bham.ac.uk (JD)

Keywords

Development, internal, external validation, prediction, prognostic, pressure injury, ulcer, overview

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

ABSTRACT

Background

Pressure injuries (PIs) place a substantial burden on healthcare systems worldwide. Risk stratification of those who are at risk of developing PIs allows preventive interventions to be focused on patients who are at the highest risk. The considerable number of risk assessment scales and prediction models available underscore the need for a thorough evaluation of their development, validation and clinical utility.

Our objectives were to identify and describe available risk prediction tools for PI occurrence, their content and development and validation methods used.

Methods

The umbrella review was conducted according to Cochrane guidance. MEDLINE, Embase, CINAHL, EPISTEMONIKOS, Google Scholar and reference lists were searched to identify relevant systematic reviews. Risk of bias was assessed using adapted AMSTAR-2 criteria. Results were described narratively. All included reviews contributed to build a comprehensive list of risk prediction tools.

Results

We identified 32 eligible systematic reviews only seven of which described the development and validation of risk prediction tools for PI. Nineteen reviews assessed the prognostic accuracy of the tools and 11 assessed clinical effectiveness. Of the seven reviews reporting model development and validation, six included only machine learning models. Two reviews included external validations of models, although only one review reported any details on external validation methods or results. This was also the only review to report measures of both discrimination and calibration. Five reviews presented measures of discrimination, such as area under the curve (AUC), sensitivities, specificities, F1 scores and G-means. For the four reviews that assessed risk of bias assessment using the PROBAST tool, all models but one were found to be at high or unclear risk of bias.

Conclusions

Available tools do not meet current standards for the development or reporting of risk prediction models. The majority of tools have not been externally validated. Standardised and rigorous approaches to risk prediction model development and validation are needed.

Registration

The protocol was registered on the Open Science Framework (https://osf.io/tepyk).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

INTRODUCTION

Pressure injuries (PI) carry a significant healthcare burden. A recent meta-analysis estimated the global burden of PIs to be 13%, two-thirds of which are hospital-acquired PIs (HAPI).[1] The average cost of a HAPI has been estimated as $11k per patient, totalling at least $27 billion a year in the United States based on 2.5 million reported cases.[2] Length of hospital stay is a large contributing cost, with patients over the age of 75 who develop HAPI having on average a 10-day longer hospital stay compared to those without PI.[3]

PIs result from prolonged pressure, typically on bony areas like heels, ankles, and the coccyx, and are more common in those with limited mobility, including those who are bedridden or wheelchair users. PIs can develop rapidly, and pose a threat in community, hospital and long-term care settings. Multicomponent preventive strategies are needed to reduce PI incidence[4] with timely implementation to both reduce harm and burden to healthcare systems.[5] Where preventive measures fail or are not introduced in adequate time, PI treatment involves cleansing, debridement, topical and biophysical agents, biofilms, growth factors and dressings[6 7 8], and in severe cases, surgery may be necessary.[5 9]

A number of clinical assessment scales for assessing the risk of PI are available (e.g. Braden[10 11], Norton[12], Waterlow[13]) but are limited by reliance on subjective clinical judgment. Statistical risk prediction models may offer improved accuracy over clinical assessment scales, however appropriate methods of development and validation are required.[14 15 16] Although methods for developing risk prediction models have developed considerably,[14 15 17 18] methodological standards of available models have been shown to remain relatively low.[17 19-22] Machine learning (ML) algorithms to develop prediction models are increasingly commonplace, but these models are at similarly high risk of bias[23] and do not necessarily offer any model performance benefit over the use of statistical methods such as logistic regression.[24] Methods for systematic reviews of risk prediction model studies have also improved,[25-27] with tools such as PROBAST (Prediction model Risk of Bias Assessment Tool)[28] now available to allow critical evaluation of study methods.

Although several systematic reviews of PI risk assessment scales and risk prediction models for PI (subsequently referred to as risk prediction tools) are available[29-38], these have been demonstrated to frequently focus on single or small numbers of scales or models, use variable review methods and show a lack of consensus about the accuracy and clinical effectiveness of available tools.[39] We conducted an umbrella review of systematic reviews of risk prediction tools for PI to gain further insight into the methods used for tool development and validation, and to summarise the content of available tools.

METHODS

Protocol registration and reporting of findings

We followed guidance for conducting umbrella reviews provided in the Cochrane Handbook for Intervention Reviews.[40] The review was reported in accordance with guidelines for Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[41] (see Appendix 1), adapted for risk prediction model reviews as required. The protocol was registered on the Open Science Framework (https://osf.io/tepyk).

Literature search

Electronic searches of MEDLINE, Embase via Ovid and CINAHL Plus EBSCO from inception to June 2024 were developed, tested and conducted by an experienced information specialist (AC),

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

employing well-established systematic review and prognostic search filters[42-44] combined with specific keyword and controlled vocabulary terms relating to PIs. Additional simplified searches were undertaken in EPISTEMONIKOS and Google Scholar due to the more limited search functionality of these two sources. The reference lists of all publications reporting reviews of prediction tools (systematic or non-systematic) were reviewed to identify additional eligible systematic reviews and to populate a list of PI risk prediction tools. Title and abstract screening and full text screening were conducted independently and in duplicate by two of four reviewers (BH, JD, YT, KS). Any disagreements were resolved by discussion or referral to a third reviewer.

Eligibility criteria for this umbrella review

Published English-language systematic reviews of risk prediction models developed for adult patients at risk of PI in any setting were included. Reviews of clinical risk assessment tools or models developed using statistical or ML methods were included, both with or without internal or external validation. The use of any PI classification system[6 45-47] as a reference standard was eligible. Reviews of the diagnosis or staging of those with suspected or existing PIs or chronic wounds, reviews of prognostic factor and predictor finding studies, and models exclusively using pressure sensor data were excluded.

Systematic reviews were required to report a comprehensive search of at least two electronic databases, and at least one other indicator of systematic methods (i.e. explicit eligibility criteria, formal quality assessment of included studies, sufficient data presented to allow results to be reproduced, or review stages (e.g. search screening) conducted independently in duplicate).

Data extraction and quality assessment

Data extraction forms (Appendix 3) were developed using the CHARMS checklist (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) and Cochrane Prognosis group template.[48 49] One reviewer extracted data concerning: review characteristics, model details, number of studies and participants, study quality and results. Extractions were independently checked by a second reviewer. Where discrepancies in model or primary study details were noted between reviews, we accessed the primary model development publications where possible.

The methodological quality of included systematic reviews was assessed using AMSTAR-2 (A Measurement Tool to Assess Systematic Reviews)[50], adapted for systematic reviews of risk prediction models (Appendix 4). Quality assessment and data extraction were conducted by one reviewer and checked by a second (BH, JD, KS), with disagreements resolved by consensus. Our adapted AMSTAR-2 contains six critical items, and limitations in any of these items reduce the overall validity of a review.[50]

Synthesis methods

Reviews were considered according to whether any information concerning model development and validation was reported. This specifically refers to reporting methods of model development or validation, and/or the presentation of measures of both discrimination and calibration. This is in contrast to evaluations of prognostic accuracy, where models are applied at a binary threshold (e.g., for high or low risk), and present only discrimination metrics with no further consideration of model performance. Available data were tabulated, and a narrative synthesis provided.

All risk prediction models identified are listed in Appendix 5 Table S4, including those for which no information about model development or validation was provided at systematic review level. Risk prediction models were classified as ML-based or non-ML models, based on how they were classified

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

in included systematic reviews, including cases where models such as logistic regression were treated as ML-based models. Where possible, the predictors included in the tools were extracted at review level and categorised into relevant groups in order to describe the candidate predictors associated with risk of PI. No statistical synthesis of systematic review results was conducted.

Reviews reporting results as prognostic accuracy (i.e. risk classification according to a binary decision) or clinical effectiveness (i.e. impact on patient management and outcomes) are reported elsewhere.[39] Hereafter, the term clinical utility is used to encompass both accuracy and clinical effectiveness.

RESULTS

Characteristics of included reviews

Following de-duplication of search results, 7200 unique records remained, of which 118 were selected for full text assessment. We obtained the full text of 111 publications of which 32 met all eligibility criteria for inclusion (see Figure 1). Seven reviews reported details about model development and internal validation[36 37 51-55], two of which also considered external validation[52 54]; 19 reported accuracy data[29 31-35 38 54 56-66]; and 11 reported clinical effectiveness data.[30 56 58 61 66-72] One review[54] reported both model development and accuracy data, and four reviews reported both accuracy and effectiveness data.[56 58 61 66]

Table 1 provides a summary of systematic review methods for all 32 reviews according to whether or not they reported any tool development methods (see Appendix 5 for full details). The seven reviews reporting prediction tool development and validation were all published within the last six years (2019 to 2024) compared to reviews focused on the clinical utility of available tools (published from 2006 to 2024). Reviews focused on model development methods almost exclusively focused on ML-based models (all but one[60] of the seven reviews limited inclusion to ML models), and frequently did not report study eligibility criteria related to study participants or setting (Table 1). In comparison, only two reviews (8%) concerning the clinical utility of models included ML-based models,[38 54] but more often reported eligibility criteria for population or setting: hospital settings (n = 3),[33 38 54] or surgical settings (n=8),[34 61 63 64 70 31], hospital or acute settings (n=2)[67 71], long-term care settings (n=2)[29 35] or the elderly (n=1).[60]

On average, reviews about tool development included more studies than reviews of clinical utility (median 22 compared to 15), more participants (median 408,504 compared to 7,684) and covered more prediction tools (median 21 compared to 3) (Table 1). Ten reviews (38%) about clinical utility included only one risk assessment scale, whereas reviews of tool development included at least 3 different risk prediction models. The PROBAST tool for quality assessment of prediction model studies was used in 57% (n=4) of tool development reviews[37 52-54], whereas validated test-accuracy specific tools such as QUADAS were used less frequently (10/26, 38%) in reviews of clinical utility. Two reviews of tool development did not report any quality assessment of included studies (29%), compared to 4 (15%) of reviews of clinical utility. Meta-analysis was conducted in two of seven (29%) reviews of tool development compared to more than half of reviews of clinical utility (15, 58%).

Methodological quality of included reviews

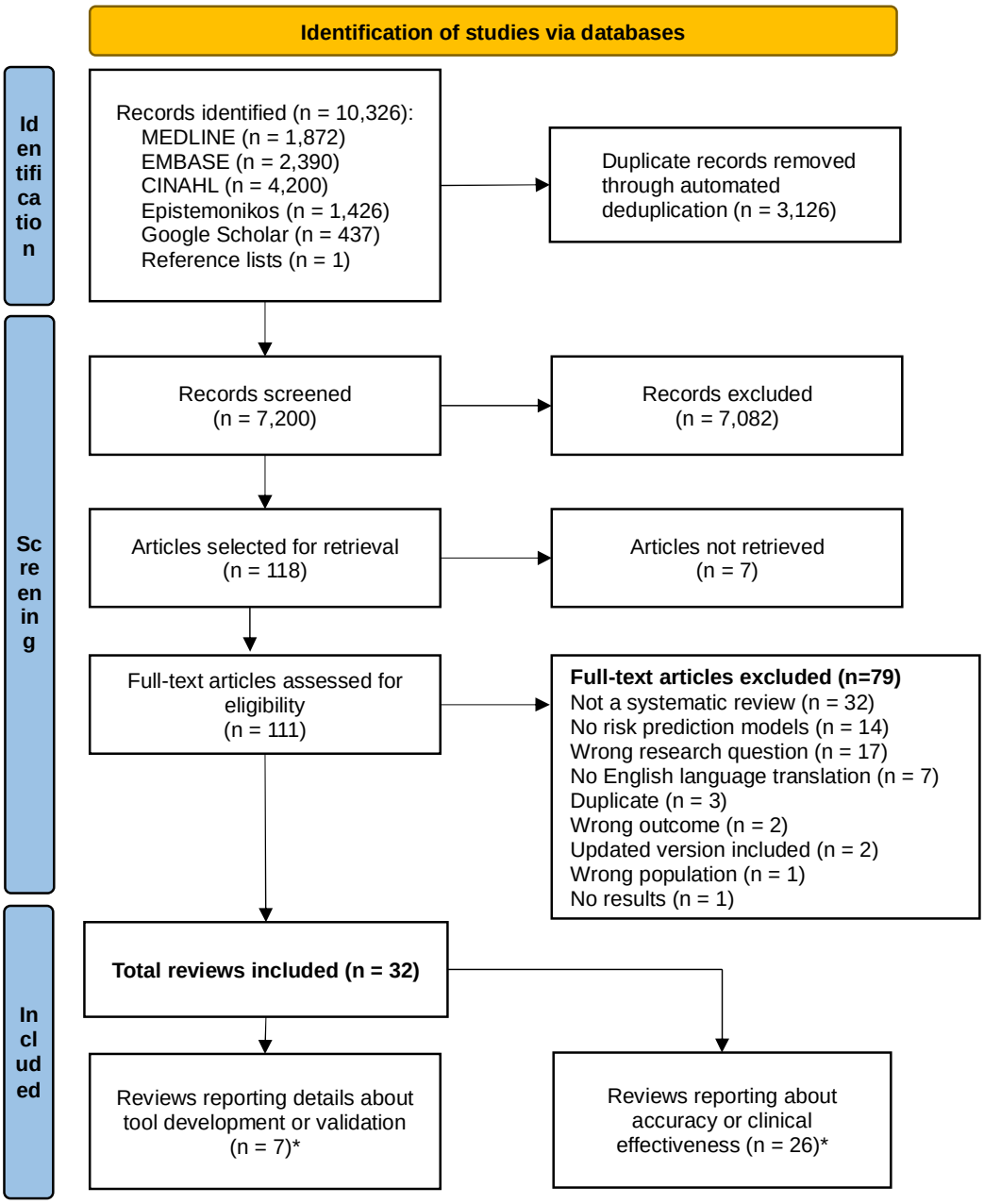
The quality of included reviews was generally low (Table 2; Appendix 5 for full assessments). The majority of reviews (71% (5/7) reviews on tool development and 78% (18/23) reviews on clinical utility) partially met the AMSTAR-2 criteria for the literature search (i.e. searched two databases, reported search strategy or key words, and justified language/publication restrictions), with only three (two reviews[56 72] on clinical utility, and one review[54] on both tool development and clinical

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

utility) meeting all criteria for 'Yes' (i.e. searching grey literature and reference lists, with the search conducted within 2 years of publication). Twenty-two reviews (69%) conducted study selection in duplicate (5/7 (71%) of reviews about tool development and 17/26 (65%) of clinical utility reviews). Conflicts of interest were reported in all seven tool development reviews and 77% of clinical utility reviews (20/26). Reviews scored poorly on the remaining AMSTAR-2 items, with around 50% or fewer reviews meeting the stipulated AMSTAR-2 criteria. Nine reviews (28%) used an appropriate method of quality assessment of included studies and provided itemisation of judgements per study. No review scored 'Yes' for all AMSTAR-2 items in either category.

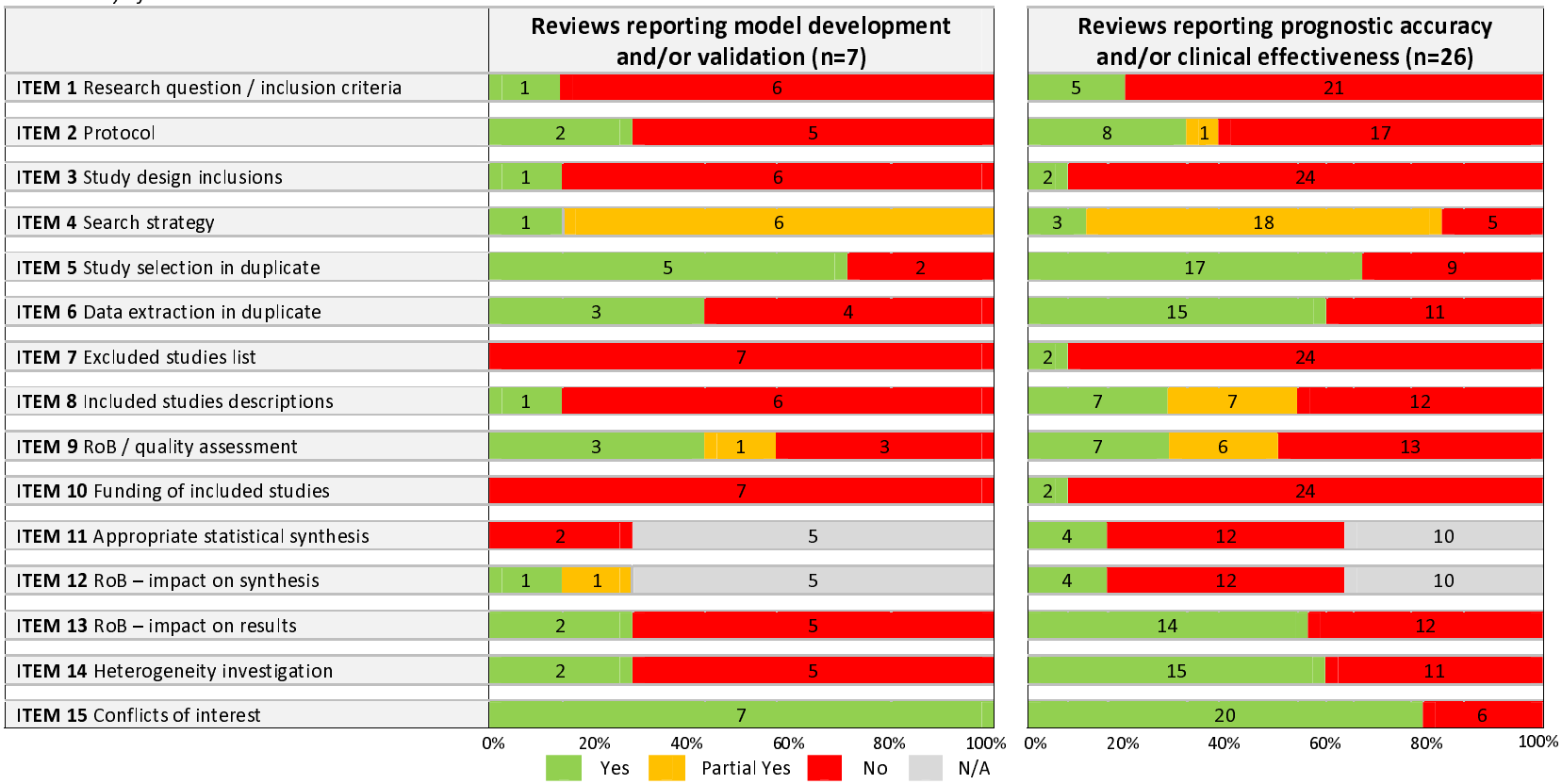
Figure 1. PRISMA[41] flowchart: identification, screening and selection process



*List of full-text articles excluded, with reasons, is given in Appendix 5. *Note that one review[54] is included in both.*

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

208 *Table 2. Summary of AMSTAR-2 assessment results*



AMSTAR – A MeaSurement Tool to Assess systematic Reviews; Item 1 – Adequate research question/ inclusion criteria?; Item 2 – Protocol and justifications for deviations?; Item 3 – Reasons for study design inclusions?; Item 4 – Comprehensive search strategy?; Item 5 – Study selection in duplicate?; Item 6 – Data extraction in duplicate?; Item 7 – Excluded studies list (with justifications)?; Item 8 – Included studies description adequate?; Item 9 – Assessment of RoB/quality satisfactory?; Item 10 – Studies' sources of funding reported?; Item 11 – Appropriate statistical synthesis method?; Item 12 – Assessment of impact of RoB on synthesised results?; Item 13 – Assessment of impact of RoB on review results?; Item 14 – Discussion/investigation of heterogeneity?; Item 15 – Conflicts of interest reported?; N/A – not applicable; RoB – risk of bias. Further details on AMSTAR items are given in Appendix 4, and results per review are given in Appendix 5. Note that where AMSTAR-2 assessment was applied to overlapping reviews (n=3) for prognostic accuracy and clinical effectiveness separately, and resulted in differing judgements for each review question, the judgements for the prognostic accuracy review question are displayed here for simplicity.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Findings

Of the 32 reviews, 26 reviews focused on the clinical utility (accuracy or effectiveness) of prediction tools. These clinical utility reviews provided no details about the development or validation of included models (except for one review[54]), and gave only limited detail about setting and study design (see Appendix 5). Reviews reporting the accuracy of prediction tools largely treated the tools as diagnostic tests to be applied at a single threshold (e.g., for high or low risk) and they did not focus on the broader aspects of prognostic model performance, such as calibration and the temporal relationship between prediction and the outcome, PI occurrence. These reviews included a total of 70 different prediction tools, predominantly derived by clinical experts, as opposed to empirically-derived models (that is, with statistical or ML methods). The methodology underlying their development is not always explicit, with scales in routine clinical usage apparently based on epidemiological evidence and clinical judgment about predictors that may not meet accepted principles for the development and reporting of risk prediction models. The most commonly included tools were the Braden[10 11] (included in 21 reviews), Waterlow[13] (n=14 reviews), Norton[12] (n=11 reviews), and Cubbin and Jackson scales[97 98] (n=8 reviews).

The seven systematic reviews that reported detailed information about model development and validation included 70 prediction models, 48 of which were unique to these seven reviews. Between three[51] and 35[36] model development studies were included; one review[52] also included eight external validation studies and another review[54] included one external validation study. Electronic health records (EHRs) were used for model development in all studies in one review[37] and for the majority of models (>66%) in the remaining reviews, where reported.[51 54 55 53] Three reviews[52 54 55] reported the use of prospectively or retrospectively collected data. No review included information about the thresholds used define whether a patient is at risk of developing PIs. Five reviews included detail about the predictors included in each model.

The largest review[36] reported that logistic regression was the most commonly reported modelling approach (20/35 models), followed by random forest (n=18), decision tree (n=12) and support vector machine (n=12) approaches. Logistic regression was also the most frequently used approach in three other reviews (18/23[55], 16/21[52] and 15/22[53]). Primary studies frequently compared the use of different ML methods using the same datasets, such that 'other' ML methods were reported with little to no further detail (e.g. 19 studies in the review by Dweekat and colleagues[36]).

Approaches to internal validation were not well reported in the primary studies. One review[52] found no information on internal validation for 76% (16/21) of studies; with re-sampling reported in two and tree-pruning, cross-validation and split sample reported in one study each. Another review[36] reported finding no information about internal validation for 20% of studies (7/35) and the use of cross-validation (n=10), split sample (n=10) techniques, or both (n=8) for the remainder. Cross-validation was used in more than half (12/22) of studies in another.[53]

Only one review reported details on methods for selection of model predictors[52]: 29% (6/21) selected predictors by univariate analysis prior to modelling and 9 used stepwise selection for final model predictors; 11 (52%) clearly reported candidate predictors, and all 21 clearly reported final model predictors. Another review[54] stated that feature selection (or predictor selection) was performed improperly and that some studies used univariate analyses to select predictors, but further details were not provided. One review[52] reported 15 models (71%) with no information about missing data, and only two using imputation techniques (imputation using another data set, and multiple imputation by chained equations). Another review[54] reported 7 models (39%) with no

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

information about missing data, missing data excluded or negligible for 4 models (22%), and single or multiple imputation techniques used for 5 (28%) and 3 (17%) models, respectively.

Model performance measures were reported by three reviews[37 52 53], all of which noted considerable variation in reported metrics and model performance including C-statistics (0.71 to 0.89 in 10 studies[53]), F1 score (0.02 to 0.99 in 9 studies[53]), G-means (0.628 to 0.822 in four studies[37]), and observed versus expected ratios (0.97 to 1 in 3 studies[52]). Four reviews[37 53-55] reported measures of discrimination associated with included models. Across reviews, reported sensitivities ranged between 0.04 and 1, specificities ranged between 0.69 and 1, and AUC values ranged between 0.50 and 1.

Shi and colleagues[52] included eight external validations using data from long-term care (n=4) or acute hospital care (n=4) settings (Appendix 5 Table S5). All were judged to be at unclear (n=4) or high (n=4) risk of bias using PROBAST. Model performance metrics for five models (TNH-PUPP[89], Berlowitz 11-item model[99], Berlowitz MDS adjustment model[90], interRAI PURS[88], Compton ICU model[94]) included C-statistics between 0.61 and 0.9 and reported observed versus expected ratios were between 0.91 and 0.97. The review also reported external validation studies for the 'SS scale'[100] and the prePURSE study tool[91], but no model performance metrics were given. A meta-analysis of C-statistics and O/E ratios was performed, including values from both development and external validation cohorts (Table 3). Parameters related to model development were not consistently reported: C-statistics ranged between 0.71 and 0.89 (n = 10 studies); observed versus expected ratios ranged between 0.97 and 1 (n=3 studies).

Pei and colleagues[54] reported that one[81] (1/18, 6%) of the model development studies included in their review also conducted an external validation. However, review authors presented accuracy metrics that originated from the internal validation, as opposed to the external validation (determined from inspection of the primary study). Additionally, no details on external validation methods and no measures of calibration were presented. Pei and colleagues[54] judged this study to be of high risk of bias using PROBAST, as with the majority of studies (16/18, 89%) included in their review. More detailed information about individual models, including predictors, specific model performance metrics and sample sizes, is presented in Appendix 5.

Included tools and predictors

A total of 124 risk prediction tools were identified (Table 4); 111 tools were identified from the 32 included systematic reviews and 13 were identified from screening the reference lists of literature reviews that used non-systematic methods that were considered during full text assessment. Full details obtained at review-level are reported in Appendix 5 Table S4.

Tools were categorised as having been developed with (60/124, 48%) or without (64/124, 52%) the use of ML methods (as defined by review authors). Prospectively collected data was used for model development for 21% of tools (26/124), retrospectively collected data for 41% (51/124), or was not reported (47/124). Information about the study populations was poorly reported, however study setting was reported for 112 prediction tools. Twenty-seven tools were reported to have been developed in hospital inpatients, and 22 were developed in long-term care settings, rehabilitation units or nursing homes or hospices. Where reported (n=100), sample sizes ranged from 15[101] to 1,252,313.[102] The approach to internal validation used for the prediction tools (e.g. cross-validation or split sample) was not reported at review-level for over two thirds of tools (83/124, 67%).

We could extract information about the predictors for only 66 of the 124 tools (Table 5 and Appendix 5). The most frequently included predictor was age (33/66, 50%), followed by pre-disposing

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

diseases/conditions (32/66, 48%), medical treatment/care received (28/66, 42%) and mobility (27/66, 41%). Tools often (31/66, 47%) included multiple pre-existing conditions or comorbidities and multiple types of treatment or medication as predictors. Other common predictors include laboratory values, continence, nutrition, body-related values (e.g. weight, height, body temperature), mental status, activity, gender and skin assessment (27% to 35% of tools). Ten tools incorporated scores from other established risk prediction scales as a predictor, with eight including Braden[10 11] scores, one including the Norton[12] score and one including the Waterlow[13] score.

Only one review[52] reported the presentation format of included tools, coded as 'score system' (n=11), 'formula equation' (n=3), 'nomogram scale' (n=2), or 'not reported' (n=6).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Table 4. Summary of tool characteristics, extracted at review-level

Tool characteristics	ML-based models (N=60, 48%)	Non-ML tools (N=64, 52%)	Total (N=124)
No. of included reviews[A] considered in			
0	0 (0)	13 (20)	13 (10)
1	31 (52)	23 (36)	54 (44)
2	6 (10)	9 (14)	15 (12)
>2	23 (38)	19 (30)	42 (34)
Development study details			
Median (range) year of publication	2020 (2000 – 2023)	1998 (1962 – 2015)	2008 (1962 – 2023)
Source of data			
Prospective	8 (13)	18 (28)	26 (21)
Retrospective	41 (68)	10 (16)	51 (41)
NS	11 (18)	36 (56)	47 (38)
Setting			
Hospital	16 (27)	11 (17)	27 (22)
Long-term care (incl. end-of-life and rehab)	8 (13)	14 (22)	22 (18)
Acute care (incl. surgical and ICU)	33 (55)	24 (38)	57 (46)
Mixed settings	1 (2)	1 (2)	2 (2)
Other	2 (3)	2 (3)	4 (3)
NS	0 (0)	12 (19)	12 (10)
Study population age			
Adults	36 (60)	34 (53)	70 (56)
Any	4 (7)	3 (5)	7 (6)
NS	20 (33)	27 (42)	47 (38)
Baseline condition			
PIs at baseline	1 (2)	0 (0)	1 (1)
No PIs at baseline	11 (18)	19 (30)	30 (24)
NS	48 (80)	45 (70)	93 (75)
Development methods			
Development method/algorithm[B]			
ML algorithms	48 (80)	0 (0)	48 (39)
Logistic regression	40 (67)	15 (23) [C]	55 (44)
Cox regression	0 (0)	5 (8)	5 (4)
Fine-Gray model	2 (3)	0 (0)	2 (2)
Clinical expertise	0 (0)	2 (3)	2 (2)
NS	0 (0)	44 (69) [D]	44 (35)
Internal validation method[B]			
Cross-validation	21 (35)	3 (5) [G]	24 (19)
Data splitting	28 (47)	0 (0)	28 (23)
Not done / NS	22 (37) [F]	61 (95)	83 (67)
Median (range) no. of final predictors[E]	7 (3 – 23)	8 (3 – 12)	7 (3 – 23)
Study cohort			
Median (range) total sample size	2674 (27 – 1252313)	285 (15 – 31150)	686 (15 – 1252313)
Median (range) number of events	207 (8 – 86410)	51 (9 – 1350)	98 (8 – 86410)
Median (range) proportion of events (% of sample size)	10.43% (0.42% – 80.00%)	14.84% (1.18% – 46.67%)	14.69% (0.42% – 80.00%)

Note that tools were categorised as ML or non-ML tools based on the descriptions from authors of the included systematic reviews that the tools were identified in. [A] *the 32 included systematic reviews;* [B] *tools use multiple methods, therefore total number not equal to N (100%);* [C] *one study also used discriminant analysis for model development;* [D] *many seemed to use clinical expertise, but development methods were not clearly reported;* [E] *counting of final predictors may vary between models: some* authors *may count individual factors, while others consider domains or subscales;* [F] *one review*[36] *implies 5 models did not implement internal validation;* [G] *'resampling' (not described further) was used for the development of 2 models; ML – machine learning; NS – not stated; ICU – intensive care unit; PI – pressure injury.*

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Table 5. Predictor categories and frequency (%) of inclusion in N=66 tools.

Predictor category	No. of tools predictor appears in
Age	33 (50)
Pre-disposing conditions	32 (48)
Receiving medical treatment/care	28 (42)
Mobility	27 (41)
Laboratory values	23 (35)
Continence	22 (33)
Nutrition	22 (33)
Body	21 (32)
Mental Status	21 (32)
Activity	21 (32)
Gender	21 (32)
Skin	18 (27)
General Health	14 (21)
Braden[10 11] score	8 (12)
Length of stay	8 (12)
Pressure injury	7 (11)
Surgery duration	6 (9)
Ability to ambulate	6 (9)
Medical unit, ward, visit	5 (8)
Ethnicity or place of birth	5 (8)
Friction, shear, pressure	3 (5)
Body position	3 (5)
Pain	3 (5)
Hygiene	2 (3)
Isolation	2 (3)
Smoking	2 (3)
Norton[12] or Waterlow[13] score	2 (3)
'Special' (not explained)	2 (3)

Figures are given as count (% out of 66 tools with information on predictors). Note that multiple predictors may fall within the same predictor category. For instance, the category 'skin' may encompass both 'skin moisture' and 'skin integrity', with the frequency count reflecting the entire predictor category rather than individual predictors.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

DISCUSSION

This umbrella review summarises data from 32 eligible systematic reviews of PI risk prediction tools. Quality assessment using an adaptation of AMSTAR-2 revealed that most reviews were conducted to a relatively poor standard. Critical flaws were identified, including inadequate or absent reporting of protocols (23/32, 72%), inappropriate statistical synthesis methods (13/17, 76%) and lack of consideration for risk of bias judgements when discussing review results (17/32, 53%). Despite the large number of risk prediction models identified, only seven reviews reported information about model development and validation, predominantly for ML-based prediction models. The remaining reviews reported the accuracy (sensitivity and specificity), or effectiveness of identified models. The studies included in the 'accuracy' reviews that we identified, typically reported a binary classification of participants as high or low risk of PI based on the risk prediction tool scores, rather than constituting external validations of models. For many (44/64, 69%) prediction tools that were developed without the use of ML, we were not able to determine whether reliable and robust statistical methods were used or whether models were essentially risk assessment tools developed based on expert knowledge. For nearly half (58/124, 47%) of the identified tools, predictors included in the final models were not reported. Details of study populations and settings were also lacking. It was not always clear from the reviews whether the poor reporting occurred at review level or in the original primary study publications.

Model development algorithms included logistic regression, decision trees and random forests, with a vast number of ML-based models having been developed in the last five years. Although logistic regression is considered a statistical approach[107], it does share some characteristics with ML methods.[108] Modern ML frameworks and libraries have streamlined the automation of logistic regression, including feature selection, hyperparameter optimisation, and cross-validation, solidifying its role within the ML ecosystem; however, logistic regression may still appear in non-ML contexts, as some developers continue to apply it using more traditional methods. Most (6/7, 86%) of our set of reviews reported the use of logistic regression as part of an ML-based approach, however this reflects the classifications used by included systematic reviews as opposed to our own assessment of the methods used in the primary studies, and may therefore be an overestimation of the use of ML models.

In contrast to logistic regression approaches, decision trees and random forests may not produce a quantitative risk probability. Instead, they commonly categorise patients into binary 'at risk' or 'not at risk' groups. Although the risk probabilities generated in logistic regression prediction models can be useful for clinical decision making, it was not possible to derive any information about thresholds used to define 'at risk' or 'not at risk', and for most reviews, it was unclear what the final model comprised of. This lack of transparency poses potential hurdles in applying these models effectively in clinical settings.

A recent systematic review of risk of bias in ML-developed prediction models found that most models are of poor methodological quality and are at high risk of bias.[23] In our set of reviews, of the four reviews that conducted a risk of bias assessment using the PROBAST tool, all models but one[103] were found to be at high or unclear risk of bias.[37 52-54] This raises significant concerns about the accuracy of clinical risk predictions. This issue is particularly critical in light of emerging evidence[104] on skin tone classification versus ethnicity/race-based methods in predicting pressure ulcer risk. These results underscore the need for developing bias-free predictive models to ensure accurate and equitable healthcare outcomes, especially in diverse patient populations.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Where the method of internal validation was reported, split-sample and cross-validation were the most commonly used techniques, however, detail was limited, and it was not possible to determine whether appropriate methods had been used. Although split-sample approaches have been favoured for model validation, more recent empirical work suggests that bootstrap-based optimism correction[105] or cross-validation[106] are preferred approaches. None of the included reviews reported the use of optimism correction approaches.

Only two reviews included external validations of previously developed models[52 54], however limited details of model performance were presented. External validation is necessary to ensure a model is both reproducible and generalisable[109 110], bringing the usefulness of the models included in these reviews into question. The PROGRESS framework suggests that multiple external validation studies should be conducted using independent datasets from different locations.[15] In the two reviews that included model validation studies[52 54], it is unclear whether these studies were conducted in different locations. Where reported, they were all conducted in the same setting as the corresponding development study. PROGRESS also suggests that external validations are carried out in a variety of relevant settings. Shi and colleagues[52] described four of eight validations as using 'temporal' data, which suggests that the validation population is largely the same as the development population but with use of data from different timeframes. This approach has been described as lying somewhere 'between' internal and external validation, further emphasising the need for well-designed external validation studies.[109]

Importantly, model recalibration was not reported for any external validations. Evidence suggests greater focus should be placed on large, well-designed external validation studies to validate and improve promising models (using recalibration and updating[111]), rather than developing a multitude of new ones.[15 18] Model validation and recalibration should be a continuous process, and this is something that future research should address. Following external validation, effectiveness studies should be conducted to assess the impact of model use on decision making, patient outcomes and costs.[15]

The effective use of prediction tools is also influenced by the way in which the model's output is presented to the end-user. Only one review[52] reported the presentation format of included tools, such as formula equations and nomograms. In conjunction with this, identifying and mitigating modifiable risk factors can help prevent PIs. Additional effort is needed in the development of risk prediction tools to extract predictors that are risk modifiers and provide end-users with this information, to make the predictions more interpretable and actionable.

Risk stratification in itself is not clinically useful unless it leads to an effective change in patient management. For instance, in high-risk groups, additional types of preventive interventions can be triggered, or default preventive measures can be applied more intensively (e.g., more frequent repositioning) based on the results of the risk assessment. While sensitivity and specificity are valid performance metrics, their optimisation must consider the cost of misclassification. Net benefit calculations, which can be visualised through decision curves,[112] provide a more reliable means of evaluating the clinical utility of risk assessment for PIs across a range of thresholds at which clinical action is indicated. These calculations can assist in providing a balanced use of resources while maximising positive health outcomes, such as lowering incidence of PI.

It is also important to assess whether the tool can improve outcomes with existing preventive interventions and whether it integrates well into clinical workflows (i.e., clinical effectiveness). A well-developed tool with good calibration and discrimination properties may be of limited value if these practical concerns are not addressed. Therefore, model developers should check the expected

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

value of prognosis and how the tool can guide prevention when employed in practice, before planning model development. If it's determined that there is no value in predicting certain outcomes – that brings into question whether the model should even be developed.[113]

Despite the advances in methods for developing risk prediction models, scales developed using clinical expertise such as the Braden Scale[10 11], Norton Scale[12], Waterlow Score[13] and Cubbin-Jackson Scales[97 98] are extensively discussed in numerous clinical practice guidelines for patient risk assessment, and are commonly used in clinical practice.[6 114] Although guidelines recognise their low accuracy, they are still acknowledged, while other risk prediction models are not even considered. This may be due to the availability of at least some clinical trials evaluating the clinical utility of scales.[39] Some scales, such as the Braden scale[10 11], are so widely used that they have become an integral component of risk assessment for PI in clinical practice, and have even been incorporated into EHRs. Their widespread use may impede the progress towards development, validation and evaluation of more accurate and innovative risk prediction models. Striking a balance between tradition and embracing advancements is crucial for effective implementation in healthcare settings and improving patient outcomes.

Strengths and limitations

Our umbrella review is the first to systematically identify and evaluate systematic reviews of risk prediction models for PI. The review was conducted to a high standard, following Cochrane guidance[40], and with a highly sensitive search strategy designed by an experienced information specialist. Although we excluded non-English publications due to time and resource constraints, where possible these publications were used to identify additional eligible risk prediction models. To some extent our review is limited by the use of AMSTAR-2 for quality assessment of included reviews. AMSTAR-2 was not designed for assessment of diagnostic or prognostic studies and, although we made some adaptations, many of the existing and amended criteria relate to the quality of reporting of the reviews as opposed to methodological quality. There is scope for further work to establish criteria for assessing systematic reviews of prediction models.

The main limitation, however, was the lack of detail about risk prediction models and risk prediction model performance that could be determined from the included systematic reviews. To be as comprehensive as possible in model identification, we were relatively generous in our definition of 'systematic', and this may have contributed to the often-poor level of detail provided by included reviews. It is likely, however, that reporting was poor in many of the primary studies contributing to these reviews. Excluding the ML-based models, more than half of available risk prediction scales or tools were published prior to the year 2000. The fact that the original versions of reporting guidelines for diagnostic accuracy studies[115] and risk prediction models[116] were not published until 2003 and 2015 respectively, is likely to have contributed to poor reporting. In contrast, the ML-based models were published between 2000 and 2023, with a median year of 2020. Reporting guidelines for development and validation of ML-based models are more recent[117 118], but aim to improve the reporting standards and understanding of evolving ML technologies in healthcare.

CONCLUSIONS

There is a very large body of evidence reporting various risk prediction scales, tool and models for PI which has been summarised across multiple systematic reviews of varying methodological quality. Only five systematic reviews reported the development and validation of models to predict risk of PIs. It seems that for the most part, available models do not meet current standards for the development or reporting of risk prediction models. Furthermore, most available models, including ML-based models have not been validated beyond the original population in which they were

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

developed. Identification of the optimal risk prediction model for PI from those currently available would require a high-quality systematic review of the primary literature, ideally limited to studies conducted to a high methodological standard. It is evident from our findings that there is still a lack of consensus on the optimal risk prediction model for PI, highlighting the need for more standardised and rigorous approaches in future research.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

460 *Table 1. Summary of included systematic review characteristics*

Review characteristics	Reviews on model development and validation (N=7)	Reviews on accuracy or clinical effectiveness (N=26)	All included reviews (N=32)
Median (range) year of publication	2022 (2019 – 2023)	2017 (2006 – 2024)	2019 (2006 – 2024)
Eligibility criteria			
Participants			
Adults only	2 (29) [A]	15 (58) [B]	16 (50) [A,B]
Any age	0 (0)	2 (8)	2 (6)
No age restriction reported	5 (71)	9 (35)	14 (44)
Presence of PI at baseline			
No PIs at baseline	0 (0)	6 (23)	6 (19)
NS	7 (100)	20 (77)	26 (81)
Setting			
Any healthcare setting	0 (0)	2 (8)	2 (6)
Hospital	3 (43)	3 (12)	5 (16)
Acute care (incl. surgical and ICU)	0 (0)	8 (31)	8 (25)
Hospital or acute care	0 (0)	2 (8)	2 (6)
Long-term care	0 (0)	2 (8)	2 (6)
Long-term, acute or community settings	0 (0)	1 (4)	1 (3)
NS	4 (57)	8 (31)	12 (38)
Risk assessment tools			
Any prediction tool or scale	0 (0)	9 (35)	9 (28)
Specified clinical scale(s)	0 (0)	12 (46)	12 (38)
ML-based prediction models	6 (86)	2 (8)	7 (22)
ML or statistical models	1 (14)	0 (0)	1 (3)
PI prevention strategies	0 (0)	1 (4)	1 (3)
NS	0 (0)	2 (8)	2 (6)
PI classification system			
Any	0 (0)	1 (4)	1 (3)
Accepted standard classifications	0 (0)	2 (8)	2 (6)
Several specified classification systems (NPUAP, EPUAP, AHCPR or TDCPS)	0 (0)	3 (12)	3 (9)
Other	0 (0)	1 (4)	1 (3)
NS	7 (100)	19 (73)	25 (78)
Source of data			
Prospective only	0 (0)	4.5 (17) [C]	4.5 (14) [C]
Prospective or retrospective	1 (14)	2.5 (10) [C]	3.5 (41) [C]
NS	6 (86)	19 (73)	24 (75)
Study design restrictions			
Yes	1 (14)	14 (54)	15 (47)
No	0 (0)	3 (12)	3 (9)
NS	6 (86)	9 (35)	14 (44)
Review methods			
Median (range) no. sources[D] searched	5 (2 – 9)	6 (2 – 14)	5 (2 – 14)
Publication restrictions:			
End date (year)			
2000-2009	0 (0)	3 (12)	3 (9)
2010-2019	1 (14)	16 (62)	17 (53)
2020-2023	6 (86)	7 (27)	12 (38)
Language			
English only	5 (71)	10 (38)	15 (47)
2 languages	1 (14)	3 (12)	3 (9)
>2 languages	0 (0)	3 (12)	3 (9)
No restrictions	0 (0)	4 (15)	4 (13)

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

NS	1 (14)	6 (23)	7 (23)
Quality assessment tool [E]			
PROBAST	4 (57)	1 (4) [F]	4 (13) [F]
QUADAS	0 (0)	2 (8)	2 (6)
QUADAS-2	0 (0)	8 (31)	8 (25)
JBI tools	1 (14)	3 (12)	4 (13)
CASP	0 (0)	2 (8)	2 (6)
Cochrane RoB tool	0 (0)	1 (4)	1 (3)
Other	0 (0)	6 (23)	6 (19)
None	2 (29)	4 (15)	6 (19)
Meta-analysis included	2 (29)	15 (58)	16 (50)
Method of meta-analysis (% of reviews incl. meta-analysis)			
Univariate RE/FE model (depending on heterogeneity assessment)	1 (50) [G]	2 (13) [G]	3 (19)
Univariate RE model	1 (50)	6 (40) [G]	6 (38) [G]
Hierarchical model (for DTA studies)	0 (0)	2 (13)	2 (13)
Unclear/NS	0 (0)	5 (33) [G]	5 (31) [G]
Volume of evidence			
Median (range) no. studies	22 (3 – 35)	15 (1 – 70)	17 (1 – 70)
Median (range) no. participants	408,504 (6,674 – 1,278,148)	7,684 (528 – 408,504)	11,729 (528 – 1,278,148)
Median (range) no. tools	21 (3 – 35)	3 (1 – 28)	4 (1 – 35)

Figures are number (%) of reviews, unless otherwise specified. [A] one review[55] specified restricting to "adult" populations, but only restricted by aged ≥14 years; [B] one review[60] restricted to aged >60 years; [C] one review [56] states either prospective or retrospective data eligible for Research Question 1, but prospective only for Research Question 2, hence 0.5 added to each category; [D] including databases, bibliographies or registries; [E] reviews may fall into multiple categories, therefore total number within domain not necessarily equal to N (100%); [F] one review[38] reported use of PROBAST in methods, but did not present any PROBAST results; [G] one review conducts univariate meta-analysis for a single estimate, e.g. c-statistic[52], AUC[62], RR[57], or OR.[58]

AHCPR – Agency for Health Care Policy and Research; CASP – Critical Appraisal Skills Programme; DTA – diagnostic test accuracy; EPUAP – European Pressure Ulcer Advisory Panel; FE – fixed effects; ICU – intensive care unit; JBI – Joanna Briggs Institute; ML – machine learning; NPUAP – National Pressure Ulcer Advisory Panel; NS – not stated; PI – pressure injury; PROBAST – Prediction model Risk of Bias Assessment; QUADAS (2) – Quality Assessment of Diagnostic Accuracy Studies (Version 2); RE – random effects; TDCPS – Torrance Developmental Classification of Pressure Sore.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

474 *Table 3. Results of reviews reporting model development and validation*

Review author (publication year)	DEV/ VAL (no. studies)	Setting of included studies; data sources	Model development algorithms	Internal validation methods	Brief description of study quality	Summary of model performance results
Barghouthi[55] (2023)	DEV (23)	Setting of included studies NS, but the review's inclusion criteria specified hospital settings Retrospective n=15; prospective n=5; both retrospective and prospective n=1; case-control study n=1; experimental study design n=1 EHRs n=20; international or national database n=3	LR n=18; RF n=13; DT n=5; NN n=5; SVM n=5; Fine-Gray Model n=2; KNN n=2; XGBoost n=2; Adaboost n=1; BART n=1; EBM n=1; Gaussian Naïve Bayes n=1; GB n=1; GBM n=1; LDA n=1; NB n=1	Split sample n=17; NS n=6	RoB assessed using JBI critical appraisal checklist for cohort studies, and only summary results provided. Only one domain was low RoB across all included studies, which was whether the participants were free from the outcome (PIs) at the start of the study. Domains with mostly high-risk (<50%) or moderate-risk (51-81%) results related to statistical analysis methods, follow-up time, dealing with confounding factors, and measurement of the exposure.	Only reported measures of discrimination: Accuracy ranged between 0.52 (ML Walther[73]) and 0.99 (ML Anderson[74]); Sensitivity ranged between 0.04 (ML Walther[73]) and 1 (ML Hu[75], ML Anderson[74]); Specificity ranged between 0.69 (ML Hyun[76], ML Nakagami[77]) and 1 (ML Cai[78], ML Walther[73]); PPV ranged between 0.01 (ML Nakagami[77]) and 1 (ML Cai[78]); NPV ranged between 0.08 (ML SPURS[79], ML Cramer[80]) and 1 (ML Hu[75], ML Anderson[74], ML Ladios-Martin[81]); AUC ranged between 0.50 (ML Cai[78]) and 1 (ML Hu[75], ML Cai[78])
Dweekat[36] (2023)	DEV (34); unclear (1)[A]	HAPI/CAPI n=32; SRPI n=2; detection of PI (effect on length of stay) n=1; nursing home residents n=2 Data sources NS	LR n=20; RF n=18; DT n=12; SVM n=12; MLP n=9; KNN n=4; LDA n=1; other n=19	CV n=10; split sample n=10; split sample and CV n=8; NS n=7	No RoB assessment	Results not reported; review focused on methods only
Jiang[37] (2021)	DEV (9)	ICU n=3; operating room n=2; acute care hospital n=1; oncology department n=1; end-of-life care n=1; mobility-related disabilities n=1 EHRs used in all models	DT n=5; LR n=3; NN n=2; SVM n=2; BN n=1; GB n=1; MTS n=1; RF n=1	Split sample n=4; NS n=9	RoB assessed using PROBAST. Overall RoB high for all predictive models. All models at high RoB in analysis domain.	Only reported measures of discrimination: F-score ranged between 0.377 (ML Su MTS[82]) and 0.670 (ML Su LR[82]); G-means ranged between 0.628 (ML Kaewprag BN[83]) and 0.822 (ML Su MTS[82]); Sensitivity ranged between 0.478 (ML Kaewprag[83]) and 0.848 (ML Yang[84]); Specificity ranged between 0.703 (ML Deng[85]) and 0.988 (ML Su LR[82])
Pei[54] (2023)	DEV (17); DEV+VAL (1)	*DEV* ICU n=4; hospitalised patients n=8; hospitalised patients	RF n=12; LR n=11; DT n=9; SVM n=8; NN n=5; MTS n=1;	CV n=1; Split sample n=5; split sample and CV	RoB assessed using PROBAST. Overall, 16/18 (88.9%) papers were at high RoB, 1 (5.6%) was at	Only reported measures of discrimination: *Summary AUC* 0.9449

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Review author (publication year)	DEV/ VAL (no. studies)	Setting of included studies; data sources	Model development algorithms	Internal validation methods	Brief description of study quality	Summary of model performance results
		awaiting surgery n=3; cancer patients n=1; end-of-life inpatients n=1 Retrospective n=14; prospective n=3 EHRs n=12; MIMIC-IV database n=1; CONCERN database n=1 *DEV+VAL* ICU n=1 Retrospective n=1 EHRs n=1	NB n=3; KNN n=2; MLP n=1; XGBoost n=2; BART n=1; LASSO n=1; BN n=1; ANN n=1; EN n=1; GBM n=1; Other[B] n=1	n=10; NS n=2	unclear RoB and only 1 (5.6%) was at low RoB. 14 (77.8%) studies were at high RoB in the analysis domain. The most common factors contributing to the high risk of bias in the analysis domain included an inadequate number of events per candidate predictor, poor handling of missing data and failure to deal with overfitting.	*Summary sensitivity* 0.79 (95% CI: 0.78, 0.80); N_{cases} = 19,893 *Summary specificity* 0.87 (95% CI: 0.88, 0.87); $N_{non\text{-}cases}$ = 388,611 *Summary likelihood ratios* PLR 10.71 (95% CI: 5.98, 19.19) NLR 0.21 (95% CI: 0.08, 0.50) *Pooled odds ratio* 52.39 (95% CI: 24.83, 110.55)
Ribeiro[51] (2021)	DEV (3)	SRPI cardiovascular n=2; SRPI critical care n=1 EHRs used in n=2 models	ANN n=1; RF n=1; XGBoost n=1	Split sample n=2; NS n=1	No RoB assessment	Only reported measures of discrimination: Accuracy ranged between 0.79 (ML Alderden [86]) and 0.82 (ML Chen[87]).
Shi[52] (2019)	DEV (21); VAL (7)	*DEV* General acute care hospital n=7; long-term care n=5; specific acute care (e.g. ICU) n=4; cardiovascular surgery n=2; trauma and burn centres n=1; rehabilitation units n=1; unclear n=1 Retrospective n=11; prospective n=10 *VAL* Long-term care n=3; specific acute care (e.g. ICU) n=2; general (acute care) hospital n=2 Retrospective n=4; prospective n=3	LR n=16; cox regression n=5; ANN n=1; C4.5 ML (DT induction algorithm) n=1; DA n=1; DT n=1; NS n=1	CV n=1; tree-pruning n=1; split sample n=1; re-sampling n=2; NS n=16	RoB assessed using PROBAST. *DEV* Overall RoB unclear for two models. Overall RoB high for the remaining 19 models. Analysis and outcome domains were mostly at high RoB. *VAL* Overall RoB unclear for three validation studies. Overall RoB high for the remaining four validation studies. Analysis and outcome domains were mostly at high RoB.	C-statistics[C] ranged between 0.61 (interRAI PURS[88]) and 0.90 (TNH-PUPP[89]); O/E ratios[C] ranged between 0.91 (Berlowitz MDS[90]) and 1.0 (prePURSE study tool[91]) *Pooled C-statistics*[C] TNH-PUPP[89]: 0.86 (95% CI 0.81–0.90), n=2 Fragmment scale[92]: 0.79 (95% CI 0.77–0.82), n=1[D] Berlowitz 11-item model[93]: 0.75 (95% CI 0.74–0.76), n=2 Berlowitz MDS model[90]: 0.73 (95% CI 0.72–0.74), n=2 interRAI PURS[88]: 0.65 (95% CI 0.60–0.69), n=3 Compton[94]: 0.81 (95% CI 0.78–0.84), n=2 *Pooled O/E ratios*[C] Berlowitz 11-item model[93]: 0.99 (95% CI 0.95–1.04), n=2

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Review author (publication year)	DEV/ VAL (no. studies)	Setting of included studies; data sources	Model development algorithms	Internal validation methods	Brief description of study quality	Summary of model performance results
						Berlowitz MDS[90]: 0.94 (95% CI 0.88–1.01), n=2
Zhou[53] (2022)	DEV (22)	SRPI n=3; ICU n=11; hospitalised n=6; rehabilitation centre n=1; hospice n=1 EHR n=18; MIMIC-III database n=4	LR n=15; RF n=10; DT n=9; SVM n=9; ANN n=8; BN n=3; XGBoost n=3; GB n=2; AdaBoost n=1; CANTRIP n=1; LSTM n=1; EN n=1; KNN n=1; MTS n=1; NB n=1	CV n=12; NS n=10	RoB assessed using PROBAST. Overall RoB unclear for five studies. Overall RoB high for 15 models. RoB not assessed in two studies due to use of unstructured data.	Only reported measures of discrimination: F1 score ranged between 0.02 (ML Nakagami[77]) and 0.99 (ML Song [2][95]); AUC ranged between 0.78 (ML Delparte[96]) and 0.99 (ML Song [2][95]); Sensitivity ranged between 0.08 (ML Cai[78]) and 0.99 (ML Song [2][95]); Specificity ranged between 0.63 (ML Delparte[96]) and 1 (ML Cai[78])

[A] Appears to be a model validation study but the review only included model development studies.
[B] Other includes: average perception, Bayes point machine, boosted DT, boosted decision forest, decision jungle and locally deep SVM. All reported for one study[81].
[C] Values from fixed-effects meta-analyses, pooling development and external validation study estimates together.
[D] One data source but included two C-statistic values (one for model development and one for internal validation) that were subsequently pooled.

AUC – area under curve; ANN – artificial neural network; BART – Bayesian additive regression tree; BN – Bayesian network; CAPI – community-acquired pressure injury; CANTRIP - reCurrent Additive Network for Temporal RIsk Prediction; CONCERN – Communicating Narrative Concerns Entered; CV – cross-validation; DEV – development; DOR – diagnostic odds ratio; DT – decision tree; EBM – explainable boosting machine; EHRs – electronic health records; EN – elastic net; GB(M) – gradient boosting (machine); HAPI – hospital-acquired pressure injury; ICU – intensive care unit; JBI – Joanna Briggs Institute; KNN – k-nearest neighbours; LASSO – least absolute shrinkage and selection operator; (L)DA – (linear) discriminant analysis; LSTM – long short-term memory; LR – logistic regression; MIMIC – Medical Information Mart for Intensive Care; ML – machine learning; MLP – multilayer perceptron; MTS – Mahalanobis-Taguchi system; N/A – not applicable; NB – naïve Bayes; NN – neural network; NLR – negative likelihood ratio; NS – not stated; O/E – observed vs expected; PI – pressure injury; PLR – positive likelihood ratio; PROBAST – Prediction model Risk of Bias ASsessment Tool; RF – random forest; RoB – risk of bias; SRPI – surgery-related pressure injury; SVM – support vector machine; VAL – validation; XGBoost – extreme gradient boosting

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data produced in the present work are contained in the manuscript and supplementary file.

Conflicting Interests

The authors of this manuscript have the following competing interests: VV is an employee of Paul Hartmann AG; ES and THB received consultancy fees from Paul Hartmann AG. VV, ES and THB were not involved in data curation, screening, data extraction, analysis of results or writing of the original draft. These roles were conducted independently by authors at the University of Birmingham. All other authors received no personal funding or personal compensation from Paul Hartmann AG and have declared that no competing interests exist.

Funding

This work was commissioned and supported by Paul Hartmann AG (Heidenheim, Germany), part of HARTMANN GROUP. The contract with the University of Birmingham was agreed on the legal understanding that the authors had the freedom to publish results regardless of the findings.

YT, JD, BH and AC are funded by the National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre (BRC). This paper presents independent research supported by the NIHR Birmingham BRC at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Author Contributions

Conceptualisation: Bethany Hillier, Katie Scandrett, April Coombe, Tina Hernandez-Boussard, Ewout Steyerberg, Yemisi Takwoingi, Vladica Velickovic, Jacqueline Dinnes

Data curation: Bethany Hillier, Katie Scandrett, April Coombe, Jacqueline Dinnes

Formal analysis: Bethany Hillier, Katie Scandrett, Jacqueline Dinnes

Funding acquisition: Yemisi Takwoingi, Vladica Velickovic, Jacqueline Dinnes

Investigation: Bethany Hillier, Katie Scandrett, April Coombe, Yemisi Takwoingi, Jacqueline Dinnes

Methodology: Bethany Hillier, Katie Scandrett, April Coombe, Tina Hernandez-Boussard, Ewout Steyerberg, Yemisi Takwoingi, Vladica Velickovic, Jacqueline Dinnes

Project administration: Bethany Hillier, Yemisi Takwoingi, Jacqueline Dinnes

Resources: Bethany Hillier, Katie Scandrett

Supervision: Yemisi Takwoingi, Jacqueline Dinnes

Writing – original draft: Bethany Hillier, Katie Scandrett, April Coombe, Jacqueline Dinnes

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Writing – review & editing: Bethany Hillier, Katie Scandrett, April Coombe, Tina Hernandez-Boussard, Ewout Steyerberg, Yemisi Takwoingi, Vladica Velickovic, Jacqueline Dinnes

Acknowledgements

We would like to thank Mrs. Rosie Boodell (University of Birmingham, UK) for her help in acquiring the publications necessary to complete this piece of work.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

References

1. Li Z, Lin F, Thalib L, et al. Global prevalence and incidence of pressure injuries in hospitalised adult patients: A systematic review and meta-analysis. *International Journal of Nursing Studies* 2020;105:103-546. doi: 10.1016/j.ijnurstu.2020.103546
2. Padula WV, Delarmente BA. The national cost of hospital-acquired pressure injuries in the United States. *Int Wound J* 2019;16(3):634-40. doi: 10.1111/iwj.13071 [published Online First: 2019/01/28]
3. Theisen S, Drabik A, Stock S. Pressure ulcers in older hospitalised patients and its impact on length of stay: a retrospective observational study. *J Clin Nurs* 2012;21(3-4):380-7. doi: 10.1111/j.1365-2702.2011.03915.x [published Online First: 2011/12/09]
4. Sullivan N, Schoelles K. Preventing In-Facility Pressure Ulcers as a Patient Safety Strategy. *Annals of Internal Medicine* 2013;158(5.2):410-16. doi: 10.7326/0003-4819-158-5-201303051-00008
5. Institute for Quality and Efficiency in Health Care (IQWiG). Preventing pressure ulcers. Cologne, Germany 2006 [updated 2018 Nov 15. Available from: https://www.ncbi.nlm.nih.gov/books/NBK326430/?report=classic accessed Feb 2023].
6. Haesler E. European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel and Pan Pacific Pressure Injury Alliance. Prevention and Treatment of Pressure Ulcers/Injuries: Clinical Practice Guideline. 2019 [Available from: https://internationalguideline.com/2019 accessed Feb 2023].
7. Walker RM, Gillespie BM, McInnes E, et al. Prevention and treatment of pressure injuries: A meta-synthesis of Cochrane Reviews. *Journal of Tissue Viability* 2020;29(4):227-43. doi: 10.1016/j.jtv.2020.05.004
8. Shi C, Dumville JC, Cullum N, et al. Beds, overlays and mattresses for preventing and treating pressure ulcers: an overview of Cochrane Reviews and network meta-analysis. *Cochrane Database Syst Rev* 2021;8(8):Cd013761. doi: 10.1002/14651858.CD013761.pub2 [published Online First: 2021/08/16]
9. Russo CA, Steiner C, Spector W. Hospitalizations Related to Pressure Ulcers, 2006. HCUP Statistical Brief: Agency for Healthcare Research and Quality, Rockville, MD. 2008.
10. Braden B, Bergstrom N. A Conceptual Schema for the Study of the Etiology of Pressure Sores. *Rehabilitation Nursing* 1987;12(1):8-16. doi: 10.1002/j.2048-7940.1987.tb00541.x
11. Bergstrom N, Braden BJ, Laguzza A, et al. The Braden Scale for Predicting Pressure Sore Risk. *Nurs Res* 1987;36(4):205-10.
12. Norton D. Geriatric nursing problems. *Int Nurs Rev* 1962;9:39-41.
13. Waterlow J. Pressure sores: a risk assessment card. *Nursing Times* 1985;81:49-55.
14. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016;69:245-7. doi: 10.1016/j.jclinepi.2015.04.005 [published Online First: 2015/04/18]
15. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLOS Medicine* 2013;10(2):e1001381. doi: 10.1371/journal.pmed.1001381
16. Siontis GCM, Tzoulaki I, Castaldi PJ, et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *Journal of Clinical Epidemiology* 2015;68(1):25-34. doi: 10.1016/j.jclinepi.2014.09.007
17. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and Methods in Clinical Prediction Research: A Systematic Review. *PLOS Medicine* 2012;9(5):e1001221. doi: 10.1371/journal.pmed.1001221
18. Van Calster B, Steyerberg EW, Wynants L, et al. There is no such thing as a validated prediction model. *BMC Medicine* 2023;21(1):70. doi: 10.1186/s12916-023-02779-w

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

19. Wynants L, Calster BV, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020;369:m1328. doi: 10.1136/bmj.m1328
20. Ma J, Dhiman P, Qi C, et al. Poor handling of continuous predictors in clinical prediction models using logistic regression: a systematic review. *J Clin Epidemiol* 2023;161:140-51. doi: 10.1016/j.jclinepi.2023.07.017 [published Online First: 2023/08/02]
21. Dhiman P, Ma J, Qi C, et al. Sample size requirements are not being considered in studies developing prediction models for binary outcomes: a systematic review. *BMC Medical Research Methodology* 2023;23(1):188. doi: 10.1186/s12874-023-02008-1
22. Moriarty AS, Meader N, Snell KIE, et al. Predicting relapse or recurrence of depression: systematic review of prognostic models. *Br J Psychiatry* 2022;221(2):448-58. doi: 10.1192/bjp.2021.218
23. Andaur Navarro CL, Damen JAA, Takada T, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ* 2021;375:n2281. doi: 10.1136/bmj.n2281
24. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12-22. doi: 10.1016/j.jclinepi.2019.02.004 [published Online First: 20190211]
25. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;356:i6460. doi: 10.1136/bmj.i6460
26. Riley RD, van der Windt D, Croft P, et al. Prognosis research in healthcare: concepts, methods, and impact: Oxford University Press 2019.
27. Snell KIE, Levis B, Damen JAA, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). *BMJ* 2023;381:e073538. doi: 10.1136/bmj-2022-073538
28. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine* 2019;170(1):51-58. doi: 10.7326/M18-1376
29. Chen HL, Shen WQ, Liu P. A Meta-analysis to Evaluate the Predictive Validity of the Braden Scale for Pressure Ulcer Risk Assessment in Long-term Care. *Ostomy/wound management* 2016;62(9):20-8.
30. Baris N, Karabacak BG, Alpar SE. The Use of the Braden Scale in Assessing Pressure Ulcers in Turkey: A Systematic Review. *Advances in skin & wound care* 2015;28:349-57. doi: 10.1097/01.ASW.0000465299.99194.e6
31. He W, Liu P, Chen HL. The Braden Scale cannot be used alone for assessing pressure ulcer risk in surgical patients: a meta-analysis. *Ostomy/wound management* 2012;58:34-40.
32. Huang C, Ma Y, Wang C, et al. Predictive validity of the braden scale for pressure injury risk assessment in adults: A systematic review and meta-analysis. *Nursing open* 2021;8:2194-207. doi: 10.1002/nop2.792
33. Park SH, Choi YK, Kang CB. Predictive validity of the Braden Scale for pressure ulcer risk in hospitalized patients. *Journal of Tissue Viability* 2015;24:102-13. doi: 10.1016/j.jtv.2015.05.001
34. Wei M, Wu L, Chen Y, et al. Predictive Validity of the Braden Scale for Pressure Ulcer Risk in Critical Care: A Meta-Analysis. *Nursing in critical care* 2020;25:165-70. doi: 10.1111/nicc.12500
35. Wilchesky M, Lungu O. Predictive and concurrent validity of the Braden scale in long-term care: A meta-analysis. *Wound Repair and Regeneration* 2015;23:44-56. doi: 10.1111/wrr.12261
36. Dweekat OY, Lam SS, McGrath L. Machine Learning Techniques, Applications, and Potential Future Opportunities in Pressure Injuries (Bedsores) Management: A Systematic Review. *International journal of environmental research and public health* 2023;20(1) doi: 10.3390/ijerph20010796
37. Jiang M, Ma Y, Guo S, et al. Using Machine Learning Technologies in Pressure Injury Management: Systematic Review. *JMIR Medical Informatics* 2021;9(3):e25704. doi: 10.2196/25704

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

38. Qu C, Luo W, Zeng Z, et al. The predictive effect of different machine learning algorithms for pressure injuries in hospitalized patients: A network meta-analyses. *Heliyon* 2022;8(11):e11361. doi: 10.1016/j.heliyon.2022.e11361
39. Hillier B, Scandrett K, Coombe A, et al. Accuracy and clinical effectiveness of risk prediction tools for pressure injury occurrence: An umbrella review (pre-print). *MedRxiv* 2024 doi: 10.1101/2024.05.07.24307001
40. Pollock M, Fernandes RM BL, Pieper D, Hartling L,. Chapter V: Overviews of Reviews. In: Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA ed. Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022). Available from www.training.cochrane.org/handbook: Cochrane 2022.
41. Moher D, Liberati A, Tetzlaff J, et al. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine* 2009;6(7):e1000097. doi: 10.1371/journal.pmed.1000097
42. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc* 2001;8(4):391-7. doi: 10.1136/jamia.2001.0080391 [published Online First: 2001/06/22]
43. Wilczynski NL, Haynes RB. Optimal Search Strategies for Detecting Clinically Sound Prognostic Studies in EMBASE: An Analytic Survey. *Journal of the American Medical Informatics Association* 2005;12(4):481-85. doi: 10.1197/jamia.M1752
44. Geersing G-J, Bouwmeester W, Zuithoff P, et al. Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews. *PLOS ONE* 2012;7(2):e32844. doi: 10.1371/journal.pone.0032844
45. NHS. Pressure ulcers: revised definition and measurement. Summary and recommendations 2018 [Available from: https://www.england.nhs.uk/wp-content/uploads/2021/09/NSTPP-summary-recommendations.pdf accessed Feb 2023].
46. AHCPR. Pressure ulcer treatment. : Agency for Health Care Policy and Research 1994:1-25.
47. Harker J. Pressure ulcer classification: the Torrance system. *Journal of Wound Care* 2000;9(6):275-77. doi: 10.12968/jowc.2000.9.6.26233
48. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLOS Medicine* 2014;11(10):e1001744. doi: 10.1371/journal.pmed.1001744
49. Cochrane. DE form example prognostic models - scoping review: The Cochrane Collaboration: The Prognosis Methods Group; [Available from: https://methods.cochrane.org/prognosis/tools accessed Feb 2023].
50. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008. doi: 10.1136/bmj.j4008
51. Ribeiro F, Fidalgo F, Silva A, et al. Literature review of machine-learning algorithms for pressure ulcer prevention: Challenges and opportunities: MDPI 2021.
52. Shi C, Dumville JC, Cullum N. Evaluating the development and validation of empirically-derived prognostic models for pressure ulcer risk assessment: A systematic review. *International journal of nursing studies* 2019;89:88-103. doi: 10.1016/j.ijnurstu.2018.08.005
53. Zhou Y, Yang X, Ma S, et al. A systematic review of predictive models for hospital-acquired pressure injury using machine learning. *Nursing open* 2022;30 doi: 10.1002/nop2.1429
54. Pei J, Guo X, Tao H, et al. Machine learning-based prediction models for pressure injury: A systematic review and meta-analysis. *Int Wound J* 2023 doi: 10.1111/iwj.14280 [published Online First: 2023/06/20]
55. Barghouthi EaD, Owda AY, Asia M, et al. Systematic Review for Risks of Pressure Injury and Prediction Models Using Machine Learning Algorithms. *Diagnostics (Basel, Switzerland)* 2023;13(17) doi: 10.3390/diagnostics13172739
56. Chou R, Dana T, Bougatsos C, et al. Pressure ulcer risk assessment and prevention: a systematic comparative effectiveness review. *Annals of internal medicine* 2013;159(1):28-38.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

57. García-Fernández FP, Pancorbo-Hidalgo PL, Agreda JJS. Predictive Capacity of Risk Assessment Scales and Clinical Judgment for Pressure Ulcers: A Meta-analysis. *Journal of Wound Ostomy & Continence Nursing* 2014;41(1):24-34. doi: 10.1097/01.WON.0000438014.90734.a2
58. Pancorbo-Hidalgo PL, Garcia-Fernandez FP, Lopez-Medina IM, et al. Risk assessment scales for pressure ulcer prevention: a systematic review. *J Adv Nurs* 2006;54(1):94-110. doi: 10.1111/j.1365-2648.2006.03794.x
59. Park SH, Lee HS. Assessing Predictive Validity of Pressure Ulcer Risk Scales- A Systematic Review and Meta-Analysis. *Iranian journal of public health* 2016;45(2):122-33.
60. Park SH, Lee YS, Kwon YM. Predictive Validity of Pressure Ulcer Risk Assessment Tools for Elderly: A Meta-Analysis. *Western journal of nursing research* 2016;38:459-83. doi: 10.1177/0193945915602259
61. Tayyib NAH, Coyer F, Lewis P. Pressure ulcers in the adult intensive care unit: a literature review of patient risk factors and risk assessment scales. *Journal of Nursing Education and Practice* 2013;3(11):28-42.
62. Wang N, Lv L, Yan F, et al. Biomarkers for the early detection of pressure injury: A systematic review and meta-analysis. *Journal of Tissue Viability* 2022;31:259-67. doi: 10.1016/j.jtv.2022.02.005
63. Zhang Y, Zhuang Y, Shen J, et al. Value of pressure injury assessment scales for patients in the intensive care unit: Systematic review and diagnostic test accuracy meta-analysis. *Intensive & critical care nursing* 2021;64:103009. doi: 10.1016/j.iccn.2020.103009
64. Zimmermann GS, Cremasco MF, Zanei SSV, et al. Pressure injury risk prediction in critical care patients: an integrative review. *Texto & Contexto-Enfermagem* 2018;27(3)
65. Chen X, Diao D, Ye L. Predictive validity of the Jackson−Cubbin scale for pressure ulcers in intensive care unit patients: A meta ‐ analysis. *Nursing in Critical Care* 2023;28(3):370-78. doi: 10.1111/nicc.12818
66. Mehicic A, Burston A, Fulbrook P. Psychometric properties of the Braden scale to assess pressure injury risk in intensive care: A systematic review. *Intensive & critical care nursing* 2024;83:103686. doi: 10.1016/j.iccn.2024.103686
67. Gaspar S, Peralta M, Marques A, et al. Effectiveness on hospital-acquired pressure ulcers prevention: a systematic review. *International Wound Journal* 2019;16(5):1087-102. doi: 10.1111/iwj.13147
68. Ontario HQ. Pressure ulcer prevention: an evidence-based analysis. *Ontario health technology assessment series* 2009;9(2):1-104.
69. Kottner J, Dassen T, Tannen A. Inter- and intrarater reliability of the Waterlow pressure sore risk scale: A systematic review. *International Journal of Nursing Studies* 2009;46:369-79. doi: 10.1016/j.ijnurstu.2008.09.010
70. Lovegrove J, Ven S, Miles SJ, et al. Comparison of pressure injury risk assessment outcomes using a structured assessment tool versus clinical judgement: A systematic review. *Journal of Clinical Nursing* 2021 doi: 10.1111/jocn.16154 [published Online First: 2021/12/01]
71. Lovegrove J, Miles S, Fulbrook P. The relationship between pressure ulcer risk assessment and preventative interventions: a systematic review. *Journal of wound care* 2018;27(12):862-75.
72. Moore ZEH, Patton D. Risk assessment tools for the prevention of pressure ulcers. *Cochrane Database of Systematic Reviews* 2019 doi: 10.1002/14651858.CD006471.pub4
73. Walther F, Heinrich L, Schmitt J, et al. Prediction of inpatient pressure ulcers based on routine healthcare data using machine learning methodology. *Scientific Reports* 2022;12(1):5044.
74. Anderson C, Bekele Z, Qiu Y, et al. Modeling and prediction of pressure injury in hospitalized patients using artificial intelligence. *BMC Med Inform Decis Mak* 2021;21(1):253. doi: 10.1186/s12911-021-01608-5 [published Online First: 20210830]
75. Hu YH, Lee YL, Kang MF, et al. Constructing Inpatient Pressure Injury Prediction Models Using Machine Learning Techniques. *Cin-Computers Informatics Nursing* 2020;38(8):415-23. doi: 10.1097/cin.0000000000000604

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

76. Hyun S, Moffatt-Bruce S, Cooper C, et al. Prediction Model for Hospital-Acquired Pressure Ulcer Development: Retrospective Cohort Study. *Jmir Medical Informatics* 2019;7(3) doi: 10.2196/13785
77. Nakagami G, Yokota S, Kitamura A, et al. Supervised machine learning-based prediction for in-hospital pressure injury development using electronic health records: A retrospective observational cohort study in a university hospital in Japan. *International Journal of Nursing Studies* 2021;119 doi: 10.1016/j.ijnurstu.2021.103932
78. Cai JY, Zha ML, Song YP, et al. Predicting the Development of Surgery-Related Pressure Injury Using a Machine Learning Algorithm Model. *Journal of Nursing Research* 2021;29(1) doi: 10.1097/jnr.0000000000000411
79. Aloweni F, Ang SY, Fook-Chong S, et al. A prediction tool for hospital-acquired pressure ulcers among surgical patients: Surgical pressure ulcer risk score. *Int Wound J* 2019;16(1):164-75. doi: 10.1111/iwj.13007 [published Online First: 2018/10/05]
80. Cramer EM, Seneviratne MG, Sharifi H, et al. Predicting the Incidence of Pressure Ulcers in the Intensive Care Unit Using Machine Learning. *EGEMS (Wash DC)* 2019;7(1):49. doi: 10.5334/egems.307 [published Online First: 20190905]
81. Ladios-Martin M, Fernández-de-Maya J, Ballesta-López FJ, et al. Predictive Modeling of Pressure Injury Risk in Patients Admitted to an Intensive Care Unit. *Am J Crit Care* 2020;29(4):e70-e80. doi: 10.4037/ajcc2020237
82. Su CT, Wang PC, Chen YC, et al. Data Mining Techniques for Assisting the Diagnosis of Pressure Ulcer Development in Surgical Patients. *Journal of Medical Systems* 2012;36(4):2387-99. doi: 10.1007/s10916-011-9706-1
83. Kaewprag P, Newton C, Vermillion B, et al. Predictive models for pressure ulcers from intensive care unit electronic health records using Bayesian networks. *Bmc Medical Informatics and Decision Making* 2017;17 doi: 10.1186/s12911-017-0471-z
84. Yang Q, Wang G, Jiang B, et al. Study on risk prediction model of unavoidable pressure ulcers in cancer patients based on decision tree. *Journal of Nursing Science* 2019;34(13):4-7.
85. Deng X, Wang Q, Li M, et al. Predicting the risk of hospital-acquired pressure ulcers in intensive care unit patients based on decision tree. *Chin J Prac Nurs* 2016;32:485-89.
86. Alderden J, Pepper GA, Wilson A, et al. Predicting Pressure Injury in Critical Care Patients: A Machine-Learning Model. *Am J Crit Care* 2018;27(6):461-68. doi: 10.4037/ajcc2018525
87. Chen HL, Yu SJ, Xu Y, et al. Artificial Neural Network: A Method for Prediction of Surgery-Related Pressure Injury in Cardiovascular Surgical Patients. *Journal of Wound Ostomy and Continence Nursing* 2018;45(1):26-30. doi: 10.1097/won.0000000000000388
88. Poss J, Murphy KM, Woodbury MG, et al. Development of the interRAI Pressure Ulcer Risk Scale (PURS) for use in long-term care and home care settings. *BMC geriatrics* 2010;10:67. doi: 10.1186/1471-2318-10-67
89. Page KN, Barker AL, Kamar J. Development and validation of a pressure ulcer risk assessment tool for acute hospital patients. *Wound Repair and Regeneration* 2011;19(1):31-37. doi: 10.1111/j.1524-475X.2010.00647.x
90. Berlowitz DR, Brandeis GH, Morris JN, et al. Deriving a risk-adjustment model for pressure ulcer development using the Minimum Data Set. *Journal of the American Geriatrics Society* 2001;49(7):866-71. doi: 10.1046/j.1532-5415.2001.49175.x
91. Schoonhoven L, Grobbee DE, Donders ART, et al. Prediction of pressure ulcer development in hospitalized patients: a tool for risk assessment. *Quality & Safety in Health Care* 2006;15(1):65-70. doi: 10.1136/qshc.2005.015362
92. Perneger TV, Raë AC, Gaspoz JM, et al. Screening for pressure ulcer risk in an acute care hospital: development of a brief bedside scale. *J Clin Epidemiol* 2002;55(5):498-504. doi: 10.1016/s0895-4356(01)00514-5

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

93. Berlowitz DR, Ash AS, Brandeis GH, et al. Rating long-term care facilities on pressure ulcer development: importance of case-mix adjustment. *Annals of Internal Medicine* 1996;124(6):557-63.
94. Compton F, Hoffmann F, Hortig T, et al. Pressure ulcer predictors in ICU patients: nursing skin assessment versus objective parameters. *J Wound Care* 2008;17(10):417-20, 22-4. doi: 10.12968/jowc.2008.17.10.31304
95. Song WY, Kang MJ, Zhang LY, et al. Predicting pressure injury using nursing assessment phenotypes and machine learning methods. *Journal of the American Medical Informatics Association* 2021;28(4):759-65. doi: 10.1093/jamia/ocaa336
96. Delparte JJ, Flett HM, Scovil CY, et al. Development of the spinal cord injury pressure sore onset risk screening (SCI-PreSORS) instrument: a pressure injury risk decision tree for spinal cord injury rehabilitation. *Spinal Cord* 2021;59(2):123-31. doi: 10.1038/s41393-020-0510-y
97. Cubbin B, Jackson C. Trial of a pressure area risk calculator for intensive therapy patients. *Intensive Care Nursing* 1991;7(1):40-44.
98. Jackson C. The revised Jackson/Cubbin Pressure Area Risk Calculator. *Intensive Crit Care Nurs* 1999;15(3):169-75. doi: 10.1016/s0964-3397(99)80048-2
99. Berlowitz DR, Ash AS, Brandeis GH, et al. Rating long-term care facilities on pressure ulcer development: Importance of case-mix adjustment. *Annals of Internal Medicine* 1996;124(6):557-63. doi: 10.7326/0003-4819-124-6-199603150-00003
100. Suriadi Sanada H, Sugama J, Thigpen B, et al. Development of a new risk assessment scale for predicting pressure ulcers in an intensive care unit. *Nursing in critical care* 2008;13(1):34-43.
101. Lowery MT. A pressure sore risk calculator for intensive care patients: 'the Sunderland experience'. *Intensive Crit Care Nurs* 1995;11(6):344-53. doi: 10.1016/s0964-3397(95)80452-8
102. Sprigle S, McNair D, Sonenblum S. Pressure Ulcer Risk Factors in Persons with Mobility-Related Disabilities. *Adv Skin Wound Care* 2020;33(3):146-54. doi: 10.1097/01.ASW.0000653152.36482.7d
103. Do Q, Lipatov K, Ramar K, et al. Pressure Injury Prediction Model Using Advanced Analytics for At-Risk Hospitalized Patients. *Journal of patient safety* 2022;18(7):e1083-e89.
104. McCreath HE, Bates-Jensen BM, Nakagami G, et al. Use of Munsell color charts to measure skin tone objectively in nursing home residents at risk for pressure ulcer development. *Journal of Advanced Nursing* 2016;72(9):2077-85. doi: 10.1111/jan.12974
105. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res* 2017;26(2):796-808. doi: 10.1177/0962280214558972 [published Online First: 2014/11/19]
106. Smith GC, Seaman SR, Wood AM, et al. Correcting for optimistic prediction in small data sets. *Am J Epidemiol* 2014;180(3):318-24. doi: 10.1093/aje/kwu140 [published Online First: 2014/06/24]
107. Riley RD, Collins GS. Stability of clinical prediction models developed using statistical or machine learning methods. *Biometrical Journal* 2023;65(8):2200302. doi: 10.1002/bimj.202200302
108. Salazar D, Vélez J, Salazar Uribe J. Comparison between SVM and Logistic Regression: Which One is Better to Discriminate? *Revista Colombiana de Estadística* 2012;35:223-37.
109. Ramspek CL, Jager KJ, Dekker FW, et al. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;14(1):49-58. doi: 10.1093/ckj/sfaa188 [published Online First: 2020/11/24]
110. de Hond AAH, Shah VB, Kant IMJ, et al. Perspectives on validation of clinical predictive algorithms. *npj Digital Medicine* 2023;6(1):86. doi: 10.1038/s41746-023-00832-9
111. Binuya MAE, Engelhardt EG, Schats W, et al. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol* 2022;22(1):316. doi: 10.1186/s12874-022-01801-8 [published Online First: 2022/12/12]

medRxiv preprint doi: https://doi.org/10.1101/2024.05.07.24306999; this version posted November 14, 2024. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

112. Riley RD, Archer L, Snell KIE, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *BMJ* 2024;384:e074820. doi: 10.1136/bmj-2023-074820

113. Hingorani AD, Windt DAvd, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ : British Medical Journal* 2013;346:e5793. doi: 10.1136/bmj.e5793

114. Qaseem A, Mir TP, Starkey M, et al. Risk Assessment and Prevention of Pressure Ulcers: A Clinical Practice Guideline From the American College of Physicians. *Annals of Internal Medicine* 2015;162(5):359-69. doi: 10.7326/m14-1567

115. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527. doi: 10.1136/bmj.h5527 [published Online First: 2015/10/28]

116. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162(1):W1-73. doi: 10.7326/m14-0698

117. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, et al. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *Journal of the American Medical Informatics Association* 2020;27(12):2011-15. doi: 10.1093/jamia/ocaa088

118. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378. doi: 10.1136/bmj-2023-078378