DEVELOPMENT AND VALIDATION OF RISK PREDICTION TOOLS FOR PRESSURE INJURY OCCURRENCE: AN UMBRELLA REVIEW

Bethany Hillier[1,2], Katie Scandrett[1,2], April Coombe[1], Tina Hernandez-Boussard[3], Ewout Steyerberg[4], Yemisi Takwoingi[1,2], Vladica Velickovic[5,6], Jacqueline Dinnes[1,2]

[1] Biostatistics, Evidence Synthesis, Test Evaluation And Prediction Modelling (BESTEAM), Institute of Applied Health Research, 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

1. Background

- Epidemiology and Costs: Pressure injuries (PIs) have a global prevalence of 13%, with two-thirds being hospital-acquired (HAPI). They impose significant financial burdens, costing $27 billion annually in the U.S. due to 2.5 million cases, and increase hospital stays by an average of 10 days for patients over 75 years old.
- Risk Factors and Prevention: PIs typically develop on bony areas such as heels and coccyx, predominantly in patients with limited mobility. Effective prevention involves multifaceted strategies including support surfaces, nutritional guidance, and regular repositioning.
- Challenges in Risk Assessment: Traditional risk assessment relies on clinical scales like Braden and Norton, which are subjective. Recent advancements in machine learning and statistical models promise better accuracy but require robust validation.
- Research Gap: Systematic reviews on PI risk prediction tools are often limited in scope and vary in quality, complicating comparisons and understanding.
- Study Objective: This study conducts an umbrella review to systematically evaluate and summarize evidence on the development, validation, and utility of risk prediction tools for PIs.

2. Methods

- **Protocol and Reporting Standards:**
  - Followed Cochrane Handbook for umbrella reviews.
  - Reported findings adhering to PRISMA guidelines.
  - Protocol registered on the Open Science Framework.
- **Literature Search:**
  - Conducted by an experienced information specialist in January 2023.
  - Searches executed in MEDLINE, Embase via Ovid, CINAHL Plus EBSCO, EPISTEMONIKOS, and Google Scholar.
  - Combined systematic review and prognostic search filters with PI-related terms.
  - Independent dual-review of titles, abstracts, and full texts.
- **Eligibility Criteria:**
  - English-language systematic reviews focusing on risk prediction models for adult patients at risk of PI.
  - Included reviews of both clinical tools and models developed using statistical or ML methods, with or without validation.
- **Data Extraction and Quality Assessment:**
  - Utilized CHARMS checklist and Cochrane Prognosis group template for data extraction.
  - AMSTAR-2 adapted for risk prediction models used for assessing methodological quality.
  - Dual-review process for extraction and quality assessment.
- **Synthesis Methods:**
  - Grouped reviews based on reporting of model development and validation.
  - Narrative synthesis and tabulation of methods and performance measures.
  - No statistical synthesis conducted; findings on prognostic accuracy and clinical effectiveness reported separately.

3. Results

- **Characteristics of Included Reviews**
- **Search and Selection Process:**
  - After de-duplication, 6,301 unique records were identified.
  - 110 records assessed in full-text, with 103 obtained and one additional from references.
  - 45 reviews listed available risk prediction models; 28 met eligibility for inclusion.
  - 5 reviews detailed model development and internal validation, one also included external validation.16 reviews focused on reporting accuracy data, 10 reviews evaluated clinical effectiveness, with three covering both accuracy and effectiveness.

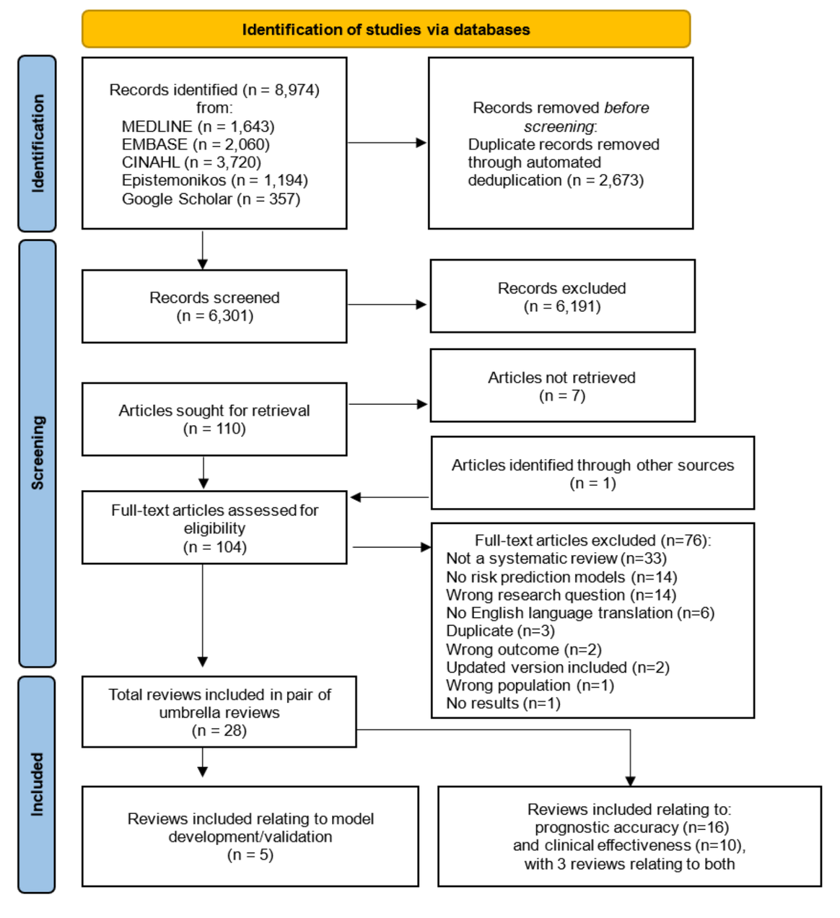
Identification of studies via databases
Identification
Records identified (n = 8,974) from:
MEDLINE (n = 1,643)
EMBASE (n = 2,060)
CINAHL (n = 3,720)
Epistemonikos (n = 1,194)
Google Scholar (n = 357)
Records removed before screening:
Duplicate records removed through automated deduplication (n = 2,673)
Screening
Records screened (n = 6,301)
Records excluded (n = 6,191)
Articles sought for retrieval (n = 110)
Articles not retrieved (n = 7)
Articles identified through other sources (n = 1)
Full-text articles assessed for eligibility (n = 104)
Full-text articles excluded (n=76):
Not a systematic review (n=33)
No risk prediction models (n=14)
Wrong research question (n=14)
No English language translation (n=6)
Duplicate (n=3)
Wrong outcome (n=2)
Updated version included (n=2)
Wrong population (n=1)
No results (n=1)
Included
Total reviews included in pair of umbrella reviews (n = 28)
Reviews included relating to model development/validation (n = 5)
Reviews included relating to: prognostic accuracy (n=16) and clinical effectiveness (n=10), with 3 reviews relating to both

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

3. Results

- **Review Characteristics:**
  - Publication years ranged from 2006 to 2023.
  - Half of the reviews conducted meta-analyses.
  - Reviews included between 1 and 70 primary studies, with participant numbers ranging from 528 to over 1.27 million.
  - Maximum number of prediction tools reviewed was 35.
- **Inclusion Specifics:**
  - 15 reviews had specific inclusion criteria based on setting or patient population (e.g., hospital inpatients, acute care, surgical patients, long-term care, elderly, bedridden).
  - 8 reviews exclusively analyzed the Braden or Waterlow scales.

Review characteristics	Reviews on model development and validation (N=5)	Reviews on accuracy or clinical effectiveness (N=23)
Median (range) year of publication	2021 (2019 – 2023)	2016 (2006 – 2022)
Eligibility criteria		
Population A		
Any population	2 (40)	3 (13)
Bedridden	1 (20)	0 (0)
Inpatients	1 (20)	5 (22)
Adults	0 (0)	12 (52)
No PIs at baseline	0 (0)	6 (26)
NS	1 (20)	5 (22)
Setting A		
Any healthcare setting	0 (0)	3 (13)
Hospital	1 (20)	5 (22)
Long-term care	0 (0)	2 (9)
Acute care (incl. surgical and ICU)	0 (0)	8 (35)
Long-term or acute	0 (0)	1 (4)
NS	4 (80)	8 (35)
Prediction models		
ML models	4 (80)	1 (4)
ML or statistical models	1 (20)	0 (0)
Any tools/scales	0 (0)	9 (39)
Specified clinical scales	0 (0)	10 (43)
Other	0 (0)	1 (4)
NS	0 (0)	2 (9)
PI classification system		
Any	0 (0)	1 (4)
Accepted standard classifications	0 (0)	2 (9)
Several specified classification systems (NPUAP, EPUAP, AHCPR or TDCPS)	0 (0)	3 (13)
Other	0 (0)	1 (4)
NS	5 (100)	16 (70)
Source of data		
Prospective only	0 (0)	4.5 (20) †
Prospective or retrospective	1 (20)	2.5 (11) †
NS	4 (80)	16 (70)
Study design restrictions		
Yes	1 (20)	13 (57)
No	0 (0)	4 (17)
NS	4 (80)	6 (26)
Review methods		
Median (range) no. sourcesB searched	5 (2 – 8)	6 (2 – 14)
Publication restrictions:		
End date (year)		
2000-2009	0 (0)	3 (13)
2010-2019	1 (20)	16 (70)
2020-2023	4 (80)	4 (17)
Language		
English only	4 (80)	9 (39)
2 languages	0 (0)	2 (9)
>2 languages	0 (0)	3 (13)
No restrictions	0 (0)	4 (17)
NS	1 (20)	5 (22)
Quality assessment tool A		
PROBAST	3 (60)	1 (4)
QUADAS	0 (0)	2 (9)
QUADAS-2	0 (0)	7 (30)
JBI tools	0 (0)	3 (13)
CASP	0 (0)	2 (9)
Other	0 (0)	6 (26)
None	2 (40)	4 (17)
Meta-analysis included	1 (20)	13 (57)
Method of meta-analysis (% of reviews incl. meta-analysis)		
Univariate RE/FE model (depending on heterogeneity assessment)	1 (100) ‡	2 (15) ‡
Univariate RE model	0 (0)	5 (38) ‡
Hierarchical model (for DTA studies)	0 (0)	2 (15)
Unclear/NS	0 (0)	4 (31) ‡
Volume of evidence		
Median (range) no. studies	22 (3 – 35)	13 (1 – 70)
Median (range) no. participants	234,105 (6,674 – 1,278,148)	6,106 (528 – 221,541)
Median (range) no. tools	21 (3 – 35)	3 (1 – 28)

Table 1. Summary of included systematic review characteristics

Figures are number (%) of reviews, unless otherwise specified. A reviews may fall into multiple categories, therefore total number within domain not necessarily equal to N (100%); B including databases, bibliographies or registries; † one review 54 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; ‡ one review conducts univariate meta-analysis for single estimate, e.g. c-statistic 52, AUC 61 RR 70 or OR71; NS – not stated; PI – pressure injury; ICU – intensive care unit; ML – machine learning; NPUAP – National Pressure Ulcer Advisory Panel; EPUAP – European Pressure Ulcer Advisory Panel; AHCPR – Agency for Health Care Policy and Research; TDCPS – Torrance Developmental Classification of Pressure Sore; QUADAS (2) – Quality Assessment of Diagnostic Accuracy Studies (Version 2); PROBAST – Prediction model Risk of Bias Assessment; JBI – Joanna Briggs Institute; CASP – Critical Appraisal Skills Programme; FE – fixed effects; RE – random effects; DTA – diagnostic test accuracy.

- **Quality Assessment Tools Used:**
  - PROBAST tool used in four reviews for quality assessment.
  - QUADAS-2 used in seven reviews, and original QUADAS in two.
  - Six reviews did not conduct a quality assessment of primary studies.


Reviews reporting model development / validation (n=5)
Reviews reporting prognostic accuracy (n=16)
Reviews reporting clinical effectiveness (n=10)
ITEM 1 Research question
ITEM 2 Protocol
ITEM 3 Inclusions
ITEM 4 Search strategy
ITEM 5 Study selection
ITEM 6 Data extraction
ITEM 7 Exclusions
ITEM 8 Study descriptions
ITEM 9 RoB/QA
ITEM 10 Funding
ITEM 11 Synthesis
ITEM 12 RoB - synthesis
ITEM 13 RoB - results
ITEM 14 Heterogeneity
ITEM 15 CoI
0% 20% 40% 60% 80% 100%
Yes
Partial Yes
No
N/A

Figure 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?; CoI – Conflicts of interest; N/A – Not Applicable; RoB – Risk of Bias; QA – quality assessment.

3. Results (continued)

Review author (publication year)	DEV/ VAL (no. studies)	Setting of included studies; data sources	Model development algorithms	Internal validation method	Brief description of study quality	Summary of model performance results
Dweekat (2023) (1)	DEV; unclear	HAPI/CAPI n=32; SRPI n=2; detection of PI (effect on length of stay) n=1; nursing home residents n=2	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=7	No RoB assessment	N/A
Jiang (2021) (2)	DEV	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; SVM n=2; NN n=2; RF n=1; MTS n=1; BN n=1; gradient boosting 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.	F-score ranged between 0.377 (ML Su MTS) and 0.670 (ML Su LR); g-means ranged between 0.628 (ML Kaewprag BN) and 0.822 (ML Su MTS); sensitivity ranged between 0.478 (ML Kawprag) and 0.848 (ML Yang); specificity ranged between 0.703 (ML Deng) and 0.988 (ML Su LR)
Ribeiro (2021) (3)	DEV	SRPI cardiovascular n=2; SRPI critical care n=1 EHRs used in n=2 models	ANN n=1; XGBoost n=1; RF n=1	Split sample n=2; NS n=1	No RoB assessment	N/A
Shi (2019) (4)	DEV; VAL	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 Prospective n=10; retrospective n=11 VAL Long-term care n=3; specific acute care (e.g. ICU) n=2; general (acute care) hospital n=2 Prospective n=3; retrospective n=4	LR n=16; cox regression n=5; ANN n=1; DT n=1; discriminant analysis n=1; C4.5 machine learning (DT induction algorithm) 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-statisticsd ranged between 0.61 (interRAI PURS) and 0.90 (TNH-PUPP); O/E ratiosd ranged between 0.91 (Berlowitz MDS) and 1.0 (prePURSE study tool) Pooled C-statisticsd TNH-PUPP: 0.86 (95% CI 0.81–0.90), n=2 Fragmgent scale: 0.79 (95% CI 0.77–0.82), n=1e Berlowitz 11-item model: 0.75 (95% CI 0.74–0.76), n=2 Berlowitz MDS model: 0.73 (95% CI 0.72–0.74), n=2 interRAI PURS: 0.65 (95% CI 0.60–0.69), n=3 Compton: 0.81 (95% CI 0.78–0.84), n=2 Pooled O/E ratiosd 0.99 (95% CI 0.95–1.04), n=2 Berlowitz MDS 0.94 (95% CI 0.88–1.01), n=2
Zhou (2022) (5)	DEV	SRPI n=3; ICU n=11; hospitalised n=6; rehabilitation centre n=1; hospice n=1 EHR n=18; Medical Information Mart for Intensive care 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.	F1 score ranged between 0.02 (ML Nakagami) and 0.99 (ML Song [2]); AUC ranged between 0.78 (ML Delparte) and 0.99 (ML Song [2]); sensitivity tanged between 0.08 (ML Cai) and 0.99 (ML Song [2]); specificity ranged between 0.63 (ML Delparte) and 1 (ML Cai)

Table 2. Results of reviews reporting model development and validation

AUC – area under curve; ANN – artificial neural network; BN – Bayesian network; CAPI – community-acquired pressure injury; CANTRIP - reCurrent Additive Network for Temporal RIsk Prediction; CV – cross-validation; DEV – development; DT – decision tree; EHRs – electronic health records; EN – elastic net; GB – gradient boosting; HAPI – hospital-acquired pressure injury; ICU – intensive care unit; KNN – k-nearest neighbors; LDA – linear discriminant analysis; LSTM – long short-term memory; LR – logistic regression; ML – machine learning; MLP – multilayer perception; MTS – Mahalanobis-Taguchi system; NB – naïve Bayes; NN – neural network; O/E – observed vs expected; PI – pressure injury; 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.
aModel only considered in this review.
bExternal validation of model included in review.
cAppears to be a model validation study but the review only included model development studies.
dValues from fixed-effects meta-analyses, pooling development and external validation study estimates together.
eOne data source but included two C-statistic values (one for model development and one for internal validation) that were subsequently pooled

23 reviews provided limited details on model development or validation and mainly involved clinically-derived tools such as Braden (n=20) and Waterlow (n=14) scales.

- **Detailed Model Analysis:**
  - Five reviews reported extensive details on model development and validation, covering 62 prediction models, with 40 unique to these reviews.
  - Models predominantly used electronic health records for development, with logistic regression being the most common approach followed by machine learning techniques like random forests and support vector machines.
- **Validation and Quality Assessment:**
  - Internal validation methods were inconsistently reported; some reviews noted a lack of validation information for a significant number of studies.
  - Two reviews utilized PROBAST for quality assessment, finding a high risk of bias in the majority of studies.
- **Model Performance and External Validation:**
  - Performance measures varied widely across reviews, with metrics like C-statistics, F1 scores, and G-means demonstrating considerable variation.
  - External validation was limited but included some models validated in long-term and acute care settings, with mixed risk of bias and performance metrics reported.
- **Overview of Tools:**
  - Identified 116 risk prediction tools across all included reviews.
  - Tools were divided based on the use of machine learning (ML) methods: 45% utilized ML (52/116) while 55% did not (64/116).
- **Data Collection and Model Development:**
  - Data used for model development varied: 19% used prospectively collected data (22/116), 40% used retrospectively collected data (46/116, including 18 ML-based models using electronic health records), and data collection methods were not reported for 41% (48/116).
  - Common settings for model development included hospital inpatients (34 tools) and long-term care settings such as rehabilitation units (20 tools).

3. Results (continued)

Tool characteristics	ML-based models (N=52, 45%)	Non-ML tools (N=64, 55%)	Total (N=116)
No. of included reviewsA considered in			
0	0 (0)	13 (20)	13 (11)
1	30 (58)	23 (36)	53 (46)
2	8 (15)	9 (14)	17 (15)
>2	14 (27)	19 (30)	33 (28)
Development study details			
Median (range) year of publication	2020 (2000 – 2022)	1998 (1962 – 2015)	2006 (1962 – 2022)
Source of data			
Prospective	4 (8)	18 (28)	22 (19)
Retrospective	36 (69)	10 (16)	46 (40)
NS	12 (23)	36 (56)	48 (41)
Setting			
Hospital	13 (25)	11 (17)	24 (21)
Long-term care (incl. end-of-life and rehab)	6 (12)	14 (22)	20 (17)
Acute care (incl. surgical and ICU)	28 (54)	24 (38)	52 (45)
Mixed settings	1 (2)	1 (2)	2 (2)
Other	2 (4)	2 (3)	4 (3)
NS	2 (4)	12 (19)	14 (12)
Study population age			
Adults	31 (60)	34 (53)	65 (56)
Any	5 (10)	3 (5)	8 (7)
NS	16 (31)	27 (42)	43 (37)
Baseline condition			
PIs at baseline	1 (2)	0 (0)	1 (1)
No PIs at baseline	11 (21)	19 (30)	30 (26)
NS	40 (77)	45 (70)	85 (73)
Development methods			
Development method/algorithmB			
ML algorithms	52 (100)	0 (0)	52 (45)
Logistic regression	35 (67)	15 (23) †	50 (43)
Cox regression	0 (0)	5 (8)	5 (4)
Clinical expertise	0 (0)	2 (3)	2 (2)
NS	0 (0)	44 (69) ‡	44 (38)
Internal validation method			
Cross-validation	21 (40)	3 (5) **	24 (21)
Data splitting	11 (21)	0 (0)	11 (9)
Not done / NS	20 (38) *	61 (95)	81 (70)
Median (range) no. of final predictorsC	6 (3 – 23)	8 (2 – 15)	7 (2 – 23)
Study cohort			
Median (range) total sample size	3,000 (27 – 1,252,313)	320 (15 – 31,150)	686 (15 – 1,252,313)
Median (range) number of events	206 (8 – 86,410)	51 (9 – 1,350)	94 (8 – 86,410)
Median (range) proportion of events (% of sample size)	10.65% (0.42% – 80.00%)	14.84% (1.18% – 46.67%)	14.40% (0.42% – 80.00%)

Table 3. Summary of tool characteristics, from review-level data

A the 28 included studies fully eligible for the umbrella reviews; B tools use multiple methods, therefore total number not equal to N (100%); C counting of final predictors may vary between models: some authors may count individual factors, while others consider domains or subscales; † one study also used discriminant analysis for model development; ‡ many seemed to use clinical expertise, but development methods were not clearly reported; * one review35 implies 5 models did not implement internal validation; ** '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.

- **Model Characteristics and Internal Validation:**
  - Sample sizes, where reported (n=92), ranged dramatically from 1,577 to over 1.25 million.
  - Internal validation approaches were poorly documented, with specific methods unidentified for 70% of models (81/116).
- **Predictors Used in Models:**
  - Information on included predictors was detailed for 53 of the 116 tools.
  - Most common predictors included mobility (51%), predisposing diseases/conditions (49%), medical treatment/care (42%), and continence (42%).
  - Other frequently mentioned predictors were age, nutrition, mental status, activity, skin conditions, and lab values.
  - Some tools also integrated established risk prediction scales such as Braden scores (six tools) and the Norton score (one tool).

Predictor category	No. of tools predictor appears in
Mobility	27 (51)
Pre-disposing conditions	26 (49)
Continence	22 (42)
Receiving medical treatment/care	22 (42)
Age	21 (40)
Mental Status	20 (38)
Nutrition	20 (38)
Activity	19 (36)
Laboratory values	18 (34)
Skin	18 (34)
General Health	16 (30)
Body	13 (25)
Gender	11 (21)
Surgery duration	7 (13)
Ability to ambulate	6 (11)
Braden score	6 (11)
Pressure injury	6 (11)
Friction, shear, pressure	5 (9)
Medical unit, ward, visit	5 (9)
Length of stay	4 (8)
Hygiene	3 (6)
Ethnicity	2 (4)
Pain	2 (4)
Smoking	2 (4)
'Special' (not explained)	2 (4)
Isolation	1 (2)
Norton score	1 (2)

Table 4. Summary of tool characteristics, from review-level data

Figures are given as count (% out of 53 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.

- **Limitations**
- **Review Standards and Search Strategy:**
  - Conducted following Cochrane guidelines with a highly sensitive search designed by an information specialist.
  - Excluded non-English publications due to time and resources but used them to identify additional models where possible.
- **Quality Assessment Challenges:**
  - Used AMSTAR-2 for quality assessment, which is not specifically designed for diagnostic or prognostic studies.
  - Adaptations were made to AMSTAR-2, but these focused more on reporting quality rather than methodological quality.
  - Potential need for establishing specific criteria for assessing systematic reviews of prediction models.
- **Limitations in Review Findings:**
  - Significant gaps in detail about risk prediction models and their performance as reported by the systematic reviews.
  - Reporting quality likely affected by the lack of contemporary guidelines during the development of many traditional (non-ML) risk prediction tools, most of which predate 2000. Contrastingly, ML-based models, published mostly post-2000 (median year 2020), benefitted from more recent reporting guidelines.

4. CONCLUSIONS

- **Evidence Scope and Review Quality:**
  - Extensive evidence on risk prediction scales, tools, and models for pressure injuries (PIs) summarized across multiple systematic reviews of varied methodological quality.
  - Only five systematic reviews comprehensively reported on the development and validation of models specifically designed to predict the risk of PIs.
- **Current Standards and Model Validation:**
  - Many available models, including those utilizing machine learning (ML), fail to meet current standards for development and reporting of risk prediction models.
  - Most models have not undergone validation outside the original populations in which they were developed, limiting their generalizability.
- **Optimal Model Identification and Research Needs:**
  - Identifying the optimal risk prediction model for PI requires a high-quality systematic review focused on primary literature adhering to stringent methodological standards.
  - Current lack of consensus on the best risk prediction model for PI underscores the need for more standardized and rigorous research methodologies in future studies.

This overview emphasizes the critical need for enhanced methodological rigor and standardized reporting in the development and evaluation of risk prediction models to improve their reliability and applicability in clinical settings.

References

1. 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
2. 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
3. Ribeiro F, Fidalgo F, Silva A, et al. Literature review of machine-learning algorithms for pressure ulcer prevention: Challenges and opportunities: MDPI 2021.
4. 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: doi:https://dx.doi.org/10.1016/j.ijnurstu.2018.08.005
5. 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: doi:https://dx.doi.org/10.1002/nop2.1429

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.

Funding

This work was commissioned and supported by Paul Hartmann AG (Heidenheim, Germany). 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 KS 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.

Conflicting Interests

VV is an employee of Paul Hartmann AG; ES and THB received consultancy fees from Paul Hartmann AG. All other authors received no personal funding or personal compensation from Paul Hartmann AG and have declared that no competing interests exist.

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