

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

Accuracy and clinical effectiveness of risk prediction tools for pressure injury occurrence: An umbrella review

Bethany Hillier^{1,2}

Katie Scandrett^{1,2}

April Coombe^{1,2}

Tina Hernandez-Boussard³

Ewout Steyerberg⁴

Yemisi Takwoingi^{1,2}

Vladica Velickovic^{5,6}

Jacqueline Dinnes^{1,2*}

Affiliations

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

* Corresponding author:

E-mail: j.dinnes@bham.ac.uk (JD)

Keywords

Sensitivity, specificity, AUC, AUROC, prognostic model, clinical scale, pressure injury, pressure ulcer, incidence, umbrella review, overview

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

34 ABSTRACT

35 **Background**

36 Pressure injuries (PIs) pose a substantial healthcare burden and incur significant costs worldwide.
37 Several risk prediction models to allow timely implementation of preventive measures and
38 potentially reduce healthcare system burden are available and in use. The ability of risk prediction
39 tools to correctly identify those at high risk of PI (prognostic accuracy) and to have a clinically
40 significant impact on patient management and outcomes (effectiveness) is not clear.

41 We aimed to evaluate the prognostic accuracy and clinical effectiveness of risk prediction tools for PI,
42 and to identify gaps in the literature.

43 **Methods and Findings**

44 The umbrella review was conducted according to Cochrane guidance. MEDLINE, Embase, CINAHL,
45 EPISTEMONIKOS, Google Scholar and reference lists were searched to identify relevant systematic
46 reviews. Risk of bias was assessed using adapted AMSTAR-2 criteria. Results were described
47 narratively.

48 We identified 16 reviews that assessed prognostic accuracy and 10 that assessed clinical
49 effectiveness of risk prediction tools for PI. The 16 reviews of prognostic accuracy evaluated 63 tools
50 (39 scales and 24 machine learning models), with the Braden, Norton, Waterlow, Cubbin-Jackson
51 scales (and modifications thereof) the most evaluated tools. Meta-analyses from a focused set of
52 included reviews showed that the scales had sensitivities and specificities ranging from 53%-97% and
53 46%-84%, respectively. Only 2/16 reviews performed appropriate statistical synthesis and quality
54 assessment. One review assessing machine learning based algorithms reported high prognostic
55 accuracy estimates, but some of which were sourced from the same data within which the models
56 were developed, leading to potentially overoptimistic results.

57 Two randomised trials assessing the effect of PI risk assessment tools on incidence of PIs were
58 identified from the 10 systematic reviews of clinical effectiveness; both were included in a Cochrane
59 review and assessed as high risk of bias. Both trials found no evidence of an effect on PI incidence.

60 **Conclusions**

61 Our findings underscore the lack of high-quality evidence for the accuracy of risk prediction tools for
62 PI. There is no reliable evidence to suggest that using existing risk prediction tools effectively reduces
63 the incidence of PIs. Further research is needed on their clinical effectiveness, but only once
64 promising prediction tools have been developed and appropriately validated.

65 INTRODUCTION

66 Pressure injuries (PI), also known as pressure ulcers or decubitus ulcers, have an estimated global
67 prevalence of 12.8% among hospitalised adults,¹ and place a significant burden on healthcare
68 systems (estimated at \$26.8 billion per year in the US alone²). PIs are most common in individuals
69 with reduced mobility, limited sensation, poor circulation, or compromised skin integrity, and can
70 affect those in community settings and long-term care as well as hospital settings. Effective
71 prevention of PI requires multicomponent preventive strategies such as mattresses, overlays, and
72 other support systems, nutritional supplementation, repositioning, dressings, creams, lotions, and
73 cleansers.³ It is therefore important to correctly identify those most at risk of PI to allow timely and
74 targeted implementation of preventive measures, to reduce harm and consequently burden to
75 healthcare systems.⁴

76 Numerous clinical assessment scales (e.g. Braden^{5,6}, Norton⁷ and Waterlow⁸) and statistical risk
77 prediction models for assessing the risk of PI are available however, many are limited by reliance on
78 subjective clinical judgment and do not appear to meet basic standards for the development or
79 reporting of risk prediction models.⁹ Nevertheless, many such tools are in routine clinical usage. For
80 example, in certain hospitals and long-term care settings in the US, healthcare professionals must
81 conduct mandatory risk assessments for PI for all patients for the purposes of risk stratification and
82 clinical triage.

83 Despite the apparent lack of sound methods for development and validation (including external
84 validation) of available risk prediction tools, there is a considerable body of evidence evaluating their
85 clinical utility, much of which has been synthesised in systematic reviews and meta-analyses.⁹ Clinical
86 utility includes both prognostic accuracy and clinical effectiveness. Prognostic accuracy is estimated
87 by applying a numeric threshold above (or below) which there is a greater risk of PI, with study
88 results presented using accuracy metrics such as sensitivity, specificity or the area under the receiver
89 operating characteristic (ROC) curve.¹⁰ Resulting accuracy is driven not only by the nominated
90 threshold for defining participants as at low or high risk for PI but by other study factors including
91 population and setting.¹¹ Clinical effectiveness, or the ability of a tool to impact on health outcomes
92 such as the incidence or severity of PI, is related both to the accuracy of the tool (or its ability to
93 correctly identify those most likely to develop PI) and to the uptake and implementation of the tool
94 in practice. Demonstrating a change in health outcomes as a result of use of a risk prediction tool is
95 vital to encourage implementation.¹²

96 Using an umbrella review approach, we aimed to provide a comprehensive overview of available
97 systematic reviews that consider the prognostic accuracy and clinical effectiveness of PI risk
98 prediction tools.

99 METHODS

100 Protocol registration and reporting of findings

101 We followed Cochrane guidance for conducting umbrella reviews¹³, and ‘Preferred Reporting Items
102 for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies’ (PRISMA-DTA)
103 reporting guidelines¹⁴ (see Appendix 1). The protocol was registered on Open Science Framework
104 (<https://osf.io/tepyk>).

105 Literature search

106 Electronic searches of MEDLINE, Embase via Ovid and CINAHL Plus EBSCO from inception to January
107 2023 were developed and conducted by an experienced information specialist (AC), employing well-

108 established systematic review and prognostic search filters,¹⁵⁻¹⁷ combined with appropriate keywords
109 related to PIs. Simplified supplementary searches in EPISTEMONIKOS and Google Scholar were also
110 undertaken (see Appendix 2 for further details). Screening of search results and full texts were
111 conducted independently and in duplicate by two reviewers (BH, JD, YT, KS), with disagreements
112 resolved by a third reviewer.

113 [Eligibility criteria for this umbrella review](#)

114 Published English-language systematic reviews of risk prediction tools developed for adult patients at
115 risk of PI in any setting were included. Clinical risk assessment scales and models developed using
116 statistical or machine learning (ML) methods were eligible (models exclusively using pressure sensor
117 data were not considered). Risk prediction tools could be applied by any healthcare professional
118 using any threshold for classifying patients as high or low risk and using any PI classification system¹⁸⁻
119 ²¹ as a reference standard. For prognostic accuracy, we required accuracy metrics, such as sensitivity
120 and specificity, to be presented but did not require full 2x2 classification tables to be reported.
121 Reviews on diagnosing or staging suspected or existing PIs were excluded.

122 To be considered ‘systematic’, reviews were required to report a thorough search of at least two
123 electronic databases and at least one other indication of systematic methods (e.g. explicit eligibility
124 criteria, formal quality assessment of included studies, adequate data presentation for
125 reproducibility of results, or review stages (e.g. search screening) conducted independently in
126 duplicate).

127 [Data extraction and quality assessment](#)

128 Data extraction forms (Appendix 3) were informed by the CHARMS checklist (CHecklist for critical
129 Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) and Cochrane
130 Prognosis group template.^{22,23} Data extraction items included review characteristics, number of
131 studies and participants, study quality and results.

132 The methodological quality of included systematic reviews was assessed using AMSTAR-2 (A
133 Measurement Tool to Assess Systematic Reviews)²⁴, adapted for systematic reviews of risk prediction
134 models (Appendix 4). Our adapted AMSTAR-2 contains six critical items, and limitations in any of
135 these items reduces the overall validity of a review.²⁴ Quality assessment and data extraction were
136 conducted by one reviewer and checked by a second (BH, JD, KS), with disagreements resolved by
137 consensus.

138 [Synthesis methods](#)

139 Reviews about prognostic accuracy and clinical effectiveness of risk prediction tools were considered
140 separately. Review methods and results were tabulated and a narrative synthesis provided.
141 Prognostic accuracy results from reviews including a statistical synthesis were tabulated according to
142 risk prediction tool.

143 Considerable overlap in risk prediction tools and included primary studies was noted between
144 reviews. For risk prediction tools that were included in multiple meta-analyses, we focused our
145 synthesis on the review(s) with the most recent search date or most comprehensive (based on
146 number of included studies) and most robust estimate of prognostic accuracy (judged according to
147 the appropriateness of the meta-analytic method used, e.g. use of recommended hierarchical
148 approaches for test accuracy data²⁵). The prognostic accuracy of risk prediction tools that were
149 included in three or fewer reviews, was reported only if an appropriate method of statistical
150 synthesis¹³ was used.

151 For clinical effectiveness results, reviews with the most recent search date or most comprehensive
152 overview of available studies and that at least partially met more of the AMSTAR-2 criteria²⁴ were
153 prioritised for narrative synthesis.

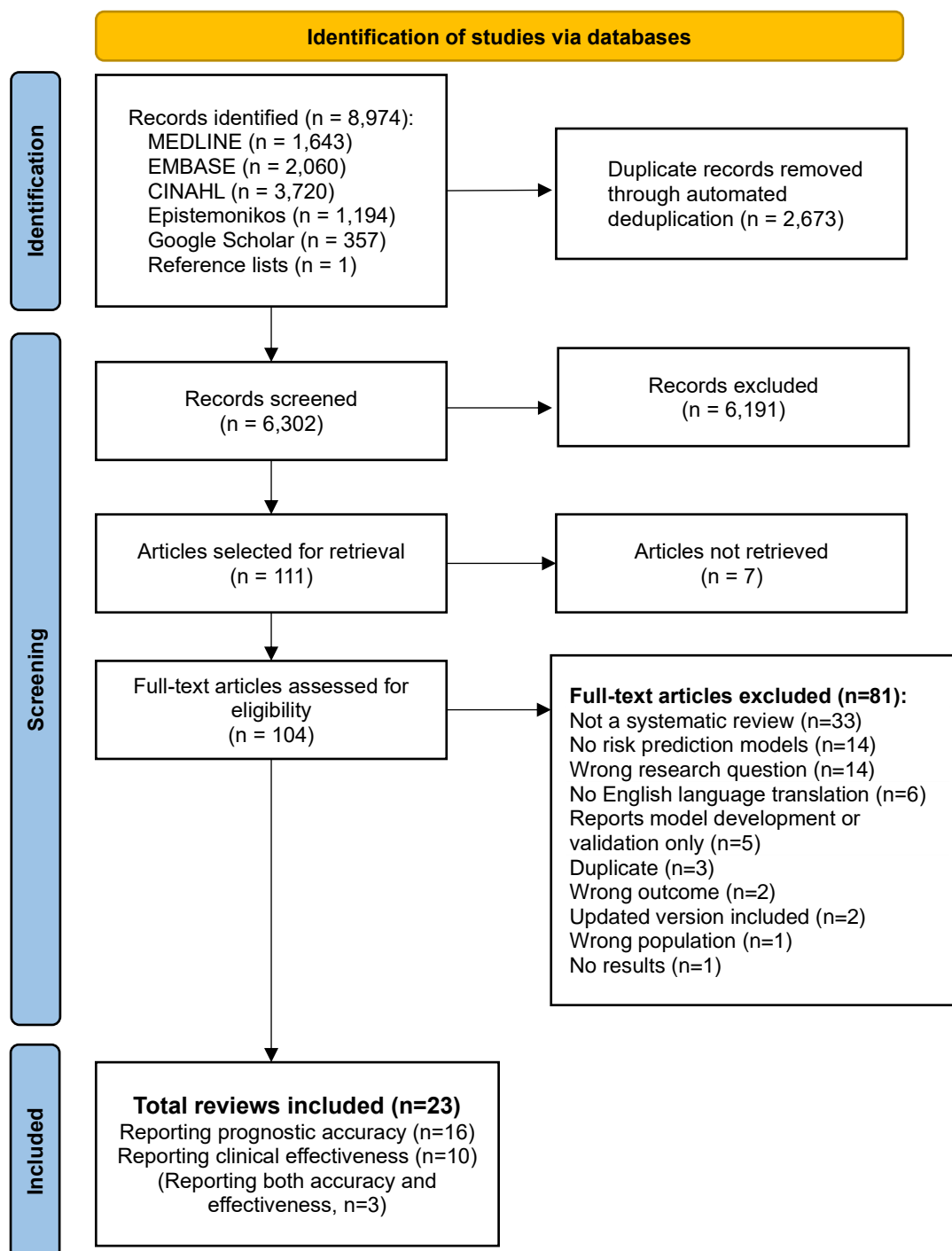
154 RESULTS

155 Characteristics of included reviews

156 A total of 110 records were selected for full-text assessment from 6302 unique records. We could
157 obtain the full text of 104 publications, of which 23 reviews met all eligibility criteria (Figure 1), 16
158 reported accuracy data²⁴⁻³⁹ and 10 reported clinical effectiveness data^{25 29 34 40-46} (three reported both
159 accuracy and effectiveness data^{25 29 34}). Table 1 and Figure 2 provide an overview of the
160 characteristics, methods and methodological quality of all 23 reviews (see Appendix 5 for full
161 details).

162

163 Figure 1. PRISMA²⁶ flowchart: identification, screening and selection process



164

165 List of full-text articles excluded, with reasons, is given in Appendix 5.

166

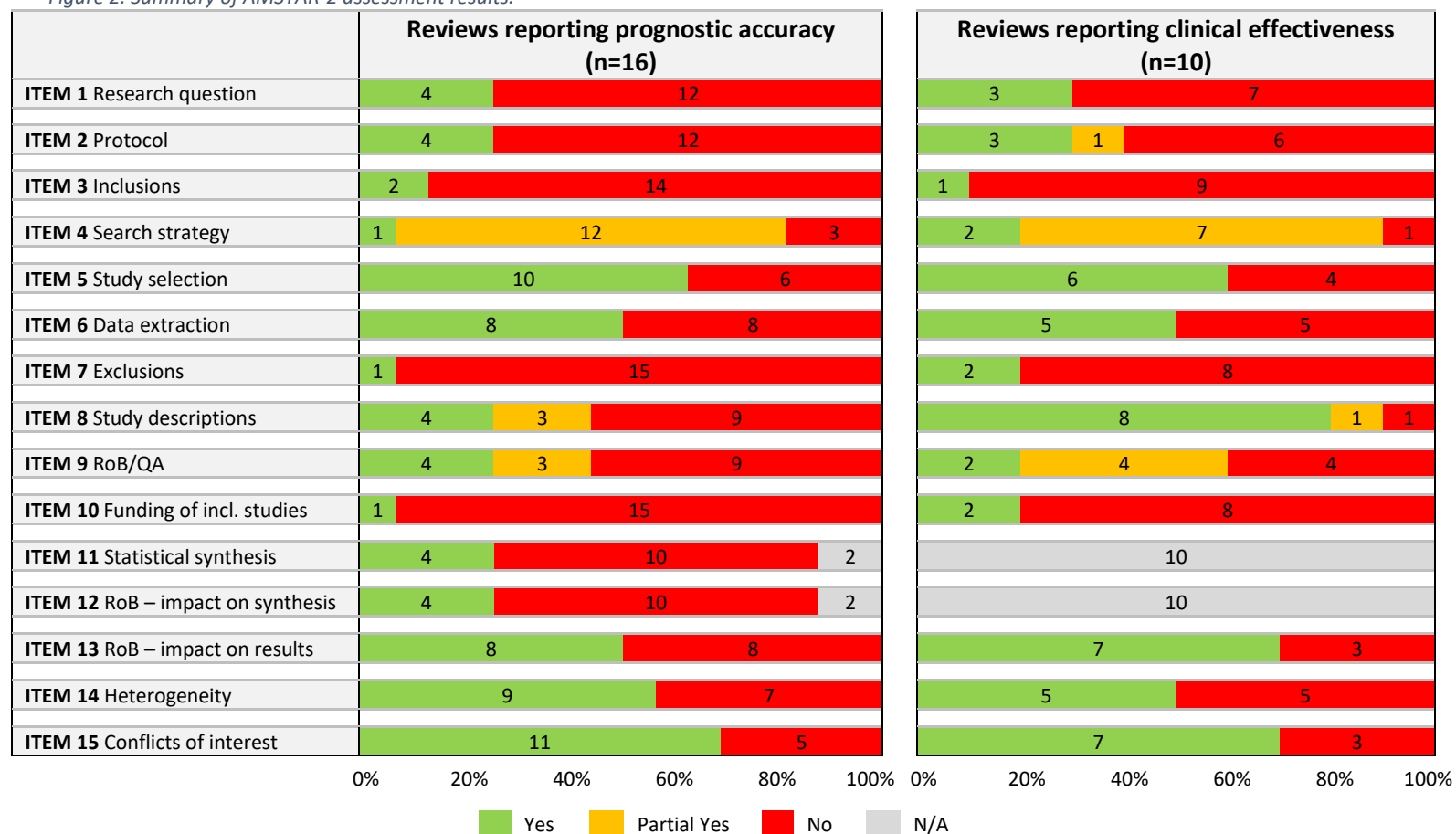
167 Table 1. Summary of included systematic review characteristics

Review characteristic	Reviews on prognostic accuracy of risk prediction tools (N=16)	Reviews on clinical effectiveness of risk prediction tools (N=10)	All included reviews (N=23)
Median (range) year of publication	2016 (2006 – 2022)	2014 (2006 – 2021)	2016 (2006 – 2022)
Eligibility criteria			
Participants			
Adults only	9 (56) ^A	5 (50)	13 (57) ^A
Any age	2 (13)	1 (10)	3 (13)
No age restriction reported	5 (31)	4 (40)	7 (30)
Presence of PI at baseline			
Excluded (no PI at baseline)	5 (31)	2 (20)	6 (26)
NS	11 (69)	7 (70)	17 (74)
Setting			
Any healthcare setting	6 (38)	5 (50)	10 (43)
Hospital	2 (13)	2 (20)	4 (22)
Long-term care	2 (13)	0 (0)	2 (9)
Acute care (incl. surgical and ICU)	5 (31)	2 (20)	6 (26)
Long-term, acute or community settings	1 (6)	1 (10)	1 (4)
Risk assessment tools			
Any prediction tool or scale	5 (31)	6 (60)	9 (39)
Specified clinical scale(s)	8 (50)	2 (20)	10 (43)
ML-based prediction models	1 (6)	0 (0)	1 (4)
PI prevention strategies	0 (0)	1 (10)	1 (4)
NS	2 (13)	1 (10)	2 (9)
PI classification system			
Any	1 (6)	0 (0)	1 (4)
Accepted standard classifications	1 (6)	1 (10)	2 (9)
Several specified classification systems (NPUAP, EPUAP, AHCP or TDCPS)	3 (19)	1 (10)	3 (13)
PI stage predefined ²⁷ /defined by study authors	1 (6)	0 (0)	1 (4)
NS	10 (63)	8 (80)	16 (70)
Source of data			
Prospective only	4 (25)	2 (20)	4.5 (20) ^B
Prospective or retrospective	1 (6)	2 (20)	2.5 (7) ^B
NS	11 (69)	6 (60)	16 (60)
Study design restrictions			
Yes	7 (44)	6 (60)	12 (52)
No	2 (13)	2 (20)	3 (13)
NS	7 (44)	2 (20)	8 (39)
Phase of development/evaluation of tools			
External evaluations only	1 (6)	N/A	N/A
Validation studies (internal or external NS)	1 (6)	N/A	N/A
NS	14 (88)	N/A	N/A
Review methods			
Median (range) no. sources^C searched	6 (2 – 14)	5 (3 – 14)	6 (2 – 14)
Publication restrictions:			
End date (year)			
2000-2009	1 (6)	3 (30)	3 (13)
2010-2019	12 (75)	6 (60)	16 (70)
2020-2023	3 (19)	1 (10)	4 (17)
Language			

English only	6 (38)	6 (60)	9 (39)
2 languages	0 (0)	2 (20)	2 (9)
>2 languages	2 (13)	2 (20)	3 (13)
No restrictions	3 (19)	1 (10)	4 (17)
NS	5 (31)	0 (0)	5 (22)
Quality assessment tool^D			
PROBAST	0 (0) ^E	N/A	0 (0) ^E
QUADAS	2 (13)	N/A	2 (9)
QUADAS-2	7 (44)	N/A	7 (30)
JBİ tools	1 (6)	2 (20)	3 (13)
CASP	2 (13)	1 (10)	2 (9)
Cochrane RoB tool	0 (0)	1 (10)	1 (4)
Other	1 (6)	4 (40)	5 (22)
None	3 (19)	2 (20)	4 (17)
Meta-analysis included	13 (81)	0 (0)	13 (57)
Method of meta-analysis (% of reviews incl. meta-analysis)			
Univariate RE/FE model (depending on heterogeneity assessment)	2 (15) ^F	N/A	N/A
Univariate RE model	5 (38) ^F	N/A	N/A
Hierarchical model (for DTA studies)	2 (15)	N/A	N/A
Unclear/NS	4 (31) ^F	N/A	N/A
Volume of evidence			
Median (range) no. studies	19 (2 – 70)	4 (1 – 20)	13 (1 – 70)
Median (range) no. participants	11,729 (609 – 221,541)	1,910 (528 – 2,273)	6,106 (528 – 221,541)
Median (range) no. tools	4 (1 – 28)	3 (2 – 9)	3 (1 – 28)

168 Figures are number (%) of reviews, unless otherwise specified. ^A one review²⁸ restricted to aged >60 years; ^B one review²⁹
169 states either prospective or retrospective data eligible for Research Question 1, but prospective only for Research Question
170 2, hence 0.5 added to each category; ^C including databases, bibliographies or registries; ^D reviews may fall into multiple
171 categories, therefore total number within domain not necessarily equal to N (100%); ^E one review³⁰ reported use of
172 PROBAST in methods, but did not present any PROBAST results; ^F one review conducts univariate meta-analysis for single
173 estimate, e.g. AUC³¹, RR³² or OR³³; AHCPR – Agency for Health Care Policy and Research; AUC – area under the curve; CASP
174 – Critical Appraisal Skills Programme; DTA – diagnostic test accuracy; EPUAP – European Pressure Ulcer Advisory Panel; FE –
175 fixed effects; ICU – intensive care unit; JBİ – Joanna Briggs Institute; ML – machine learning; NPUAP – National Pressure
176 Ulcer Advisory Panel; NS – not stated; OR – odds ratio; PI – pressure injury; PROBAST – Prediction model Risk of Bias
177 Assessment; QUADAS (2) – Quality Assessment of Diagnostic Accuracy Studies (Version 2); RE – random effects; RR – risk
178 ratio; TDCPS – Torrance Developmental Classification of Pressure Sore.
179

Figure 2. Summary of AMSTAR-2 assessment results.



181 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
 182 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?;
 183 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
 184 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
 185 Applicable; RoB – Risk of Bias; QA – quality assessment. Further details on AMSTAR items are given in Appendix 4, and results per review are given in Appendix 5.

186 Reviews were published between 2006 and 2022. Approximately half (12/23, 52%) restricted
187 inclusion to adult populations (Table 1), three included any age group, and nine (39%) did not report
188 any age restrictions. Six reviews (6/23, 26%) specified only populations without PIs at baseline for
189 inclusion. Acute care was the most common setting across both review questions, 5/16 (31%) and
190 4/10 (40%) for accuracy and effectiveness reviews, respectively. Quality assessment tools varied, with
191 QUADAS-2 (n=7) or QUADAS (n=2) being most common for reviews of accuracy (9/16, 56%). One
192 accuracy review³⁰ reported use of both QUADAS-2 and PROBAST tools in their methods, but only
193 reported QUADAS-2 results.

194 Reviews of accuracy predominantly focused on studies using any (5/16, 31%) or pre-specified (8/16,
195 50%) risk assessment tools or scales, one included only ML-based prediction models.³⁰ A total of 63
196 risk prediction tools were reported across the reviews, including 24 ML models. The number of
197 included risk prediction tools in a single review ranged from one³⁴⁻³⁹ to 28³². Only two reviews
198 reported eligibility criteria related to the development or validation of the risk prediction tools. One³³
199 (6%) excluded evaluation studies that used the same data that was used to develop the tool and the
200 other²⁹ included only “validated risk assessment instruments”, however this was not further defined
201 and the review included studies reporting the original development of risk prediction tools.

202 The majority (13/16, 81%) of accuracy reviews conducted a statistical synthesis of data, however only
203 two utilised currently recommended hierarchical approaches for the meta-analysis of test accuracy
204 data,^{36 40} seven conducted univariate meta-analysis of individual accuracy measures (e.g. sensitivity
205 and specificity separately, or AUC³¹, RR³² or odds ratio³³) and four did not clearly report the type of
206 analysis approach used.

207 Of the 10 systematic reviews evaluating the clinical effectiveness of risk prediction tools, two only
208 considered the reliability of risk assessment scales^{41 42} and eight considered effects on patient
209 outcomes (one of which also considered tool reliability⁴³). More than half of reviews (6, 60%)
210 compared use of PI risk assessment scales to clinical judgement alone or ‘standard care’. The number
211 of included studies ranged from one⁴⁴ to 20⁴⁵ and the sample sizes of primary studies ranged from
212 one (one subject and 110 raters, in an inter-rater reliability study⁴⁶) to 3,027 patients. Reported
213 outcomes included the incidence of PIs (7/10), preventative interventions prescribed (5/10) and
214 interrater reliability (3/10) (reported in Appendix 5). One (Cochrane) review used the Cochrane RoB
215 tool for quality assessment of included studies and three used JBI (n=2) or CASP (n=1) tools. Due to
216 heterogeneity in study design, risk prediction tools and outcomes evaluated, none of the included
217 reviews provided any form of statistical synthesis of study results.

218 [Methodological quality of included reviews](#)

219 The quality of included reviews was generally poor (Figure 2; Appendix 5). The AMSTAR-2 items that
220 were most consistently met (yes or partial yes) were: comprehensiveness of the search (19/23, 83%),
221 study selection independently in duplicate (15/23, 65%), and conflicts of interest reported (17/23,
222 74%).

223 Of the 16 accuracy reviews, four (25%)^{30 35 36 40} used an appropriate method of quality assessment of
224 included studies (i.e. QUADAS or QUADAS-2 dependent on publication year) and presented
225 judgements per study. Of the 10 effectiveness reviews, two (20%)^{29 47} used an appropriate method of
226 quality assessment (the Cochrane tool for assessing risk of bias⁴⁸ and a criteria consistent with AHRQ
227 Methods Guide for Effectiveness and Comparative Effectiveness Reviews⁴⁹, respectively) and
228 provided judgements per study. Four reviews either reported quality assessment results per study
229 (n=3^{41 45 50}) or were considered to use an appropriate quality assessment tool (n=1³³) (AMSTAR-2
230 criterion partially met).

231 Of the accuracy reviews that included a statistical synthesis, 31% (4/13)^{31 32 36 40} used an appropriate
232 method of meta-analysis and investigated sources of heterogeneity. Two reviews^{36 40} used
233 recommended hierarchical approaches to meta-analysis of test accuracy data (the bivariate model³⁶
234 and hierarchical summary ROC (HSROC) model⁴⁰) and one³¹ calculated summary AUC using random
235 effects meta-analysis.³²

236 Compared to the reviews of accuracy, reviews of effectiveness more commonly provided adequate
237 descriptions of primary studies (8/10, 80% vs 4/16, 25%) (Figure 2). No other major differences
238 across review questions were noted.

239 Results from reviews evaluating the prognostic accuracy of risk prediction models

240 Five of 16 accuracy reviews were prioritised for narrative synthesis (Tables 2-3) and are reported
241 below according to risk prediction tool. Four of the five reviews did not include development study
242 estimates within their meta-analyses, but this information could not be ascertained for the review³⁰
243 of ML-based models. None of these reviews assessed the quality of development methods for the
244 prediction tools considered in their statistical syntheses.

245 Braden, and modified Braden scales

246 The most recent and largest review³⁶ of the Braden scale (60 studies, including 49,326 patients),
247 which used hierarchical bivariate meta-analysis, reported an overall summary sensitivity of 0.78 (95%
248 CI 0.74, 0.82; 15,241 patients) and specificity of 0.72 (95% CI 0.66, 0.78; 34,085 patients) across all
249 reported thresholds (range ≤ 10 to ≤ 20). Summary sensitivities and specificities ranged from 0.79
250 (95% CI 0.76, 0.82) and 0.66 (95% CI 0.55, 0.75) at the lowest cut-offs for identification of high-risk
251 patients (≤ 15 in 15 studies) to 0.82 (95% CI 0.73, 0.89) and 0.70 (95% CI 0.62, 0.77) using a cut-off of
252 18 (15 studies), respectively. Heterogeneity investigations suggested higher accuracy for predicting
253 pressure injury risk in patients with a mean age of 60 years or less, in hospitalised patients
254 (compared to long-term care facility residents) and in Caucasian populations (compared to Asian
255 populations).³⁶ The review noted a high risk of bias for the 'index test' section of the QUADAS-2
256 assessment in approximately a third of included studies, but failed to provide details or reasons for
257 this assessment.

258 Two modified versions of the Braden scale^{51 52} were included by Park and colleagues.⁵³ Summary
259 sensitivities were 0.97 (95% CI 0.92, 0.99; 125 patients from four studies)⁵¹ and 0.89 (95% CI 0.71,
260 0.98; 27 patients from two studies)⁵², and summary specificities were 0.70 (95% CI 0.66, 0.73; 563
261 patients)⁵¹ and 0.71 (95% CI 0.67, 0.75; 599 patients).⁵² The review was rated critically low on the
262 AMSTAR-2 assessment, with only 3/15 (20%) criteria fulfilled. Despite reporting the use of QUADAS-2
263 for their risk of bias assessment, QUADAS-2 results were not reported, except that none of the
264 included studies were estimated to be at high risk.

265 Cubbin & Jackson scale

266 Zhang and colleagues⁴⁰ included six studies evaluating the original Cubbin & Jackson scale⁵⁴ (800
267 patients). Summary sensitivity and specificity were both reported as 0.84 (95% CIs 0.59, 0.95 and
268 0.66, 0.93, respectively)⁴⁰ suggesting that this represents the point on the HSROC curve where
269 sensitivity equals specificity, particularly as reported thresholds ranged from 24 to 34. The review
270 authors concluded that although the accuracy of the Cubbin & Jackson scale was higher than the
271 EVARUCI scale and the Braden scale, low quality of evidence and significant heterogeneity limit the
272 strength of conclusions that can be drawn.

Table 2. Findings related to prognostic accuracy, by model: Characteristics and quality of studies included within reviews

Review author (publication year)	Tool(s) evaluated n studies; N participants	Brief description of included studies	Brief description of included study quality	Method of meta-analysis
Huang ³⁶ (2021)	Braden ^{5 6} n = 60; N = 49,326	Setting: hospital (n=45; includes 22 in ICU or other acute units), LTCF (n=15) Sample size: 25 to 10,098 Mean age: range 31.7±10.9 to 84.6±7.9 Design: 47 prospective, 13 retrospective Braden cut-off (out of 23): range ≤10 to ≤20 Development study not included.	QUADAS-2: Patient selection: low RoB in 11/60 (18%), unclear RoB (>50%); low concern about applicability in 44/60 (73%), high concern for applicability in 16; Index test: low RoB in 39/60 (65%), high RoB (approx. 33%); low concern about applicability in 51/60 (85%); Reference standard: low RoB in 58/60 (97%), unclear RoB in 2 (3%); low concern about applicability in 51/60 (85%), unclear concern in 9 (15%); Flow and timing: low RoB in 50/60 (83%); unclear RoB in 10 (17%).	Bivariate meta-analysis; SROC constructed; and subgroup/stratified analyses to explore heterogeneity
Zhang ⁴⁰ (2021)	4 tools evaluated in meta-analyses	Studies not described according to prediction tool. All prospective	Studies not described according to prediction tool QUADAS-2: Overall judgement was "not so satisfactory".	HSROC model for >3 studies, or univariate fixed- or random-effects models if ≤3 studies; meta-regression heterogeneity investigation
	Braden ^{5 6} n = 18; N = 11,167	Cut-offs used range from 10.5 to 20. Development study not included.		
	Cubbin & Jackson ⁵⁴ n = 6; N = 800	Cut-offs used range from 24 to 34. Development study not included.		
	EVARUCI ⁵⁵ n = 3; N = 3,063	Cut-offs: >11, >11.5, NS Development study not included.		'Inconsistency' (I ² statistic) of studies was found to be 0%, therefore univariate fixed-effects models used
	Waterlow ⁸ n = 4; N = 1,000	Cut-offs: 12, 16, <25, 20.5 Development study not included.		
Park ²⁸ (2016b)	3 tools evaluated	Studies not described according to prediction tool Cut-offs selected "by following the one which the study researcher(s) indicated to be the most effective".		DTA meta-analysis (random effects) using MetaDiSc; no further details

	Braden ⁵⁶ n = 25; N = 10,547	cut-off: 13 (n=2); 16 (n=8); 17 (n=2); 18 (n=9); 19 (n=3); 20 (n=1) Development study not included.		
	Norton ⁷ n = 5; N = 2,408	cut-offs: 14 (n=2); 16 (n=3) Development study not included.		
	Waterlow ⁸ n = 5; N = 1,406	15 (n=1); 16 (n=2); 17 (n=1); NS (n=1) Development study not included.		
Park ⁵³ (2016a)	5 tools evaluated	Described below according to prediction tool	QUADAS-2: Studies not described according to prediction tool "None had 'high risk'"	DTA meta-analysis (random-effects) using MetaDiSc; Cochrane Handbook (2010) ⁵⁶ and Walter 2002 ⁵⁷ cited
	Braden – modified by Song & Choi ⁵¹ n = 4; N = 688	Prospective (4/4), recruiting patients with no PI at baseline (hospital ward (n=2) or ICU (n=3); mean age in the 50s (n=2), 60s (n=2). Classification used: AHCPR (n=3), Bergstrom (n=1). Braden scale cut-off used: <21 (n=1), <23 (n=1), <24 (n=2) Development study not included.		
	Braden – modified by Pang & Wong ⁵² n = 2; N = 626	Prospective (2/2), recruiting patients with no PI at baseline (OS ward (n=1) or NS (n=1); mean age 79.4 and 54.1. Classification used: NPUAP (n=2) Braden scale cut-off used: <19 (n=1), <14 (n=1) Development study not included.		
	Cubbin & Jackson ⁵⁴ n = 4; N = 662	Prospective (4/4); ICU patients for all studies (1 in surgical ICU), with no PI at baseline (n=3); mean age in the 50s (n=2), 60s (n=2). Classification used: AHCPR (n=2), NPUAP (n=1), Lowthian (n=1). C&J scale cut-off used: <24 (n=2), <26 (n=1), <28 (n=1) Development study not included.		
	Norton ⁷ n = 7; N = 2,899	Prospective (6/7); inpatients with no PI at baseline (1 LTC, 2 'hospital', 1 ICU, 1 ICU & wards); mean age in the 50s (n=1), 60s (n=3), or 80s (n=1), or NS (n=2). Classification used: AHCPR (n=3), NPUAP (n=2), EPUAP (n=1), TDCPS (n=1). Norton scale cut-off used: <14 (n=2, but reported as 3 in paper), <15 (n=2), <16 (n=3) Development study not included.		
	Waterlow ⁸ n = 6; N = 1,268	Prospective (6/6); all male* inpatients aged over 60 on average with no PI at baseline (3 included ICU patients).		

		Classification used: AHCPR (n=2), NPUAP (n=2), EPUAP (n=1), TDCPS (n=1). Waterlow scale cut-off used: <9 (n=1), <15 (n=1), <16 (n=2), <17 (n=1), NS (n=1) Development study not included.		
Qu ³⁰ (2022)	Models by ML algorithm type	Characteristics only reported overall, not by algorithm type.	QUADAS-2 details reported overall, not by algorithm type.	RevMan (Moses & Littenberg method ^{58,59}) for analysis of quantitative data (SROC plot presented) and Bayesian DTA-NMA
	Decision Tree n = 14; N = 118,292	Conducted in: hospital patients (n= 13); surgical patients (n=3), ICU (n=5), CVD patients (n=2), cancer patients (n=1), LTC (n=1)	2/14 high RoB; 10/14 low RoB; 2/14 unclear RoB	
	Logistic Regression n = 14; N = 195,927	Unclear whether development, internal validation or external validation studies included.	4/14 high RoB; 9/14 low RoB; 1/14 unclear RoB	
	Neural Network n = 9; N = 97,815		1/9 high RoB; 7/9 low RoB; 1/9 unclear RoB	
	Random Forest n = 7; N = 161,334		1/7 high RoB; 6/7 low RoB.	
	Support Vector Machine n = 9; N = 152,068		1/9 high RoB; 8/9 low RoB.	

274 * as reported in review's text. However, the table reports a mixture of female and male participants for all studies, with a mean female proportion of 50.73%.
275 AHCPR – Agency for Health Care Policy and Research; CI – confidence interval; CVD – cardiovascular disease; DTA – diagnostic test accuracy; EPUAP – European Pressure Ulcer Advisory Panel;
276 (H)SROC – (hierarchical) summary receiver operating characteristic curve; ICU – intensive care unit; LTC(F) – long-term care (facility); ML – machine learning; N – number of participants; n –
277 number of studies; NMA – network meta-analysis; NS – not stated; NPUAP – National Pressure Ulcer Advisory Panel; PI – pressure injury; PPPU – Panel for the Prediction and Prevention of
278 Pressure Ulcers; QUADAS – Quality Assessment of Diagnostic Accuracy Studies; RCT – randomised controlled trial; RoB – risk of bias; TDCPS – Torrance Developmental Classification of Pressure
279 Sore.

Table 3. Summary estimates of accuracy parameters (main results from statistical syntheses), by prediction tool

Review author (publication year)	n studies; N participants	Sensitivity (95% CI) (N = no. participants with PI)	Specificity (95% CI) (N = no. participants without PI)	Likelihood ratios (95% CI)	DOR (95% CI)	AUROC (95% CI)
TOOL: Braden⁵⁶ (1987)						
Huang ³⁶ (2021)	n = 60; N = 49,326	0.78 (0.74, 0.82)^A N=15,241 By cut-off: ≤15 (n=15): 0.79 (0.76, 0.82) 16 (n=19): 0.75 (0.67, 0.82) 17 (n=4): 0.69 (0.61, 0.76) 18 (n=15): 0.82 (0.73, 0.89) ≥19 (n=7): 0.78 (0.65, 0.87)	0.72 (0.66, 0.78)^A N=34,085 By cut-off: ≤15: 0.66 (0.55, 0.75) 16: 0.85 (0.70, 0.93) 17: 0.86 (0.50, 0.97) 18: 0.70 (0.62, 0.77) ≥19: 0.54 (0.44, 0.63)	PLR 2.80 (2.30, 3.50)^A NLR 0.30 (0.26, 0.35)^A	9.00 (7.00, 13.00)^A By cut-off: ≤15: 7.00 (4.00, 12.00) 16: 17.00 (8.00, 36.00) 17: 14.00 (2.00, 103.00) 18: 11.00 (6.00, 20.00) ≥19: 4.00 (2.00, 7.00)	0.82 (0.79, 0.85)^A By cut-off: ≤15: 0.80 (0.76, 0.83) 16: 0.84 (0.80, 0.87) 17: 0.73 (0.69, 0.77) 18: 0.83 (0.79, 0.86) ≥19: 0.67 (0.63, 0.71)
Zhang ⁴⁰ (2021)	n = 18; N = 11,167	0.78 (0.68, 0.85)^B	0.61 (0.40, 0.79)^B	PLR 2.00 (1.24, 3.24) NLR 0.36 (0.25, 0.52)	5.52 (2.61, 11.67)	0.78
Park ²⁸ (2016b)	n = 25; N = 10,547	0.72 (0.69, 0.74)^A	0.63 (0.62, 0.64)^A	PLR 2.31 (1.98, 2.69)^A NLR 0.43 (0.36, 0.51)^A	6.50 (4.64, 9.11)^A	0.79^A (SE = 0.02)
TOOL: Modified Braden scales: Braden – modified by Song & Choi⁵¹ (1991)						
Park ⁵³ (2016a)	n = 4; N = 688	0.97 (0.92, 0.99)^A N=125	0.70 (0.66, 0.73)^A N=563	PLR 3.47 (1.33, 9.06)^A NLR 0.08 (0.04, 0.19)^A	56.56 (21.88, 146.21)^A	0.95^A (SE 0.02)
TOOL: Braden – modified by Pang & Wong⁵² (1998)						
Park ⁵³ (2016a)	n = 2; N = 626	0.89 (0.71, 0.98)^A N=27	0.71 (0.67, 0.75)^A N=599	PLR 2.87 (1.88, 4.38)^A NLR 0.17 (0.06, 0.49)^A	16.06 (4.75, 54.35)^A	Not calculated
TOOL: Cubbin & Jackson⁵⁴ (1991)						
Zhang ⁴⁰ (2021)	n = 6; N = 800	0.84 (0.59, 0.95)^B	0.84 (0.66, 0.93)^B	PLR 5.12 (2.70, 9.70) NLR 0.19 (0.08, 0.49)	26.45 (13.51, 51.78)	0.90
Park ⁵³ (2016a)	n = 4; N = 662	0.67 (0.60, 0.74)^A N=194	0.75 (0.71, 0.79)^A N=468	PLR 2.80 (1.66, 4.72)^A NLR 0.34 (0.15, 0.76)^A	9.46 (2.41, 37.22)^A	0.82^A (SE 0.06)
TOOL: EVARUCI⁵⁵ (2001)						
Zhang ⁴⁰ (2021)	n = 3; N = 3,063	0.84 (0.79, 0.89)^A	0.68 (0.66, 0.70)^A	PLR 2.32 (2.14, 2.51)^A NLR 0.25 (0.19, 0.35)^A	9.79 (6.81, 14.07)^A	0.82^A
TOOL: Norton⁷ (1962)						
Park ⁵³ (2016a)	n = 7; N = 2,899	0.75 (0.70, 0.79)^A N=383	0.57 (0.55, 0.59)^A N=2,516	PLR 1.77 (1.26, 2.50)^A NLR 0.49 (0.32-0.76)^A	7.57 (2.53, 22.64)^A	0.82^A (SE 0.05)
Park ²⁸ (2016b)	n = 5; N = 2,408	0.76 (0.71, 0.80)^A	0.55 (0.53, 0.57)^A	PLR 1.58 (1.07, 2.34)^A NLR 0.47 (0.29, 0.76)^A	6.41 (1.72, 23.88)^A	0.84^A (SE 0.07)

Review author (publication year)	n studies; N participants	Sensitivity (95% CI) (N = no. participants with PI)	Specificity (95% CI) (N = no. participants without PI)	Likelihood ratios (95% CI)	DOR (95% CI)	AUROC (95% CI)
TOOL: Waterlow⁸ (1985)						
Zhang ⁴⁰ (2021)	n = 4; N = 1,000	0.63 (0.48, 0.76)^B	0.46 (0.22, 0.71)^B	PLR 1.16 (0.66, 2.01) NLR 0.82 (0.40, 1.67)	1.42 (0.40, 5.07)	0.56
Park ⁵³ (2016a)	n = 6; N = 1,268	0.55 (0.49, 0.62)^A N=246	0.82 (0.80, 0.85)^A N=1,222	PLR 2.89 (1.74, 4.79)^A NLR 0.46 (0.31, 0.70)^A	9.22 (6.43, 13.23)^A	0.82^A (SE 0.03)
Park ²⁸ (2016b)	n = 5; N = 1,406	0.53 (0.47, 0.60)^A	0.84 (0.81, 0.86)^A	PLR 3.09 (1.63, 5.83)^A NLR 0.49 (0.34, 0.72)^A	9.06 (6.30, 13.04)^A	0.81^A (SE 0.03)
ML models:^C						
Qu ³⁰ (2022) DT models	n = 14; N = 118,292	0.66 (0.42, 0.84) N=7,557	0.90 (0.78, 0.96) N=110,735	PLR 6.9 (3.2, 14.7) NLR 0.37 (0.20, 0.69)	18 (7, 49)	0.88 (0.85, 0.91)
Qu ³⁰ (2022) LR models	n = 14; N = 195,927	0.71 (0.60, 0.80) N=9046	0.83 (0.75, 0.89) N=186,881	PLR 4.3 (3.1, 5.9) NLR 0.35 (0.26, 0.46)	12 (9, 17)	0.84 (0.81, 0.87)
Qu ³⁰ (2022) NN models	n = 9; N = 97,815	0.73 (0.55, 0.86) N=9488	0.78 (0.65, 0.87) N=88,327	PLR 3.3 (2.1, 5.0) NLR 0.35 (0.21, 0.59)	9 (5, 19)	0.82 (0.79, 0.85)
Qu ³⁰ (2022) RF models	n = 7; N = 161,334	0.72 (0.26, 0.95) N=5486	0.96 (0.80, 0.99) N=155,848	PLR 16.3 (2.4, 108.9) NLR 0.29 (0.07, 1.29)	56 (3, 1258)	0.95 (0.93, 0.97)
Qu ³⁰ (2022) SVM models	n = 9; N = 152,068	0.81 (0.69, 0.90) N=6562	0.81 (0.59, 0.93) N=145,506	PLR 4.3 (1.8, 9.9) NLR 0.23 (0.13, 0.39)	19 (6, 54)	0.88 (0.85, 0.90)

281 ^A summary statistic pooled across multiple thresholds; ^B estimate derived from HSROC, but method for choosing summary point unclear; ^C it is not reported, at review level, whether the
282 results presented below are pooled across development, internal validation or external validation/evaluation studies.
283 AUROC – area under the receiver operating characteristic curve; CI – confidence interval; DT – decision tree; DOR – diagnostic odds ratio; HSROC – hierarchical summary receiver operating
284 characteristic curve; LR – logistic regression; ML – machine learning; NLR – negative likelihood ratio; NN – neural network; NS – not stated; PLR – positive likelihood ratio; RF – random forest;
285 SE – standard error; SVM – support vector machine.

286 Norton scale

287 Park and colleagues⁵³ pooled data from seven studies (2,899 participants) evaluating the Norton
288 scale, across thresholds ranging from <14 to <16. They reported summary sensitivity of 0.75 (95% CI
289 0.70, 0.79) and specificity 0.57 (95% CI 0.55, 0.59). A further four reviews presented statistically
290 synthesised results for the Norton scale (Appendix 5), including one review by Chou and colleagues²⁹
291 which included nine studies (5,444 participants) but only reported median values for accuracy
292 parameters.

293 Waterlow scale

294 Although Zhang and colleagues⁴⁰ included the fewest participants (4 studies; 1,000 participants) of all
295 six reviews that conducted a statistical synthesis of the accuracy of the Waterlow scale⁸, they
296 provided the most recent review. It was rated highest on AMSTAR-2 criteria and appropriately used
297 the HSROC model for meta-analysis across thresholds ranging from 12 to 25. Summary sensitivity
298 was 0.63 (95% CI 0.48, 0.76) and summary specificity 0.46 (95% CI 0.22, 0.71) (Table 3). A second
299 review⁵³ reported summary sensitivity of 0.55 (95% CI 0.49, 0.62) and specificity 0.82 (95% CI 0.80,
300 0.85) (6 studies; 1268 participants).

301 Machine learning algorithms

302 Qu and colleagues³⁰ conducted separate meta-analyses of 25 studies according to ML algorithm type
303 (Table 2). The review rated critically low on AMSTAR-2 items, with only 6/15 (40%) criteria fulfilled,
304 and reported using Bayesian DTA meta-analysis. The review did not restrict inclusion to external
305 evaluations of the models, and the authors did not report which estimates were sourced from
306 development data or external data. The summary AUC for the five algorithms ranged from 0.82 (95%
307 CI 0.79, 0.85; 9 studies with 97,815 participants) for neural network-based models to 0.95 (95% CI
308 0.93, 0.97; 7 studies with 161,334 participants) for random forest models (Table 3). The latter
309 approach also had the highest summary specificity 0.96 (95% CI 0.80, 0.99), with sensitivity 0.72
310 (95% CI 0.26, 0.95). The highest summary sensitivity was observed for support vector machine
311 models (0.81, 95% CI 0.69, 0.90) with summary specificity 0.81 (95% CI 0.59, 0.93) (9 studies,
312 152,068 participants). The remaining algorithms had summary sensitivities ranging from 0.66
313 (decision tree models) to 0.73 (neural network models) (Table 3). Two additional ML algorithms
314 evaluated in the included studies (Bayesian networks and LOS (abbreviation not explained)) had too
315 few studies to allow meta-analysis (Appendix 5).

316 Other scales

317 In addition to the risk prediction tools reported above, Zhang and colleagues⁴⁰ reported on the
318 EVARUCI scale⁵⁵, presenting summary sensitivity and specificity of 0.84 (95% CI 0.79, 0.89) and 0.68
319 (95% CI 0.66, 0.70), respectively (3 studies; 3,063 participants). These results were pooled across
320 thresholds, 11 and 11.5 (one not reported).

321 Beyond the results covered by our five prioritised reviews, three further modifications of the Braden
322 scale were evaluated in statistical syntheses: Braden modified by Kwong⁶⁰, the 4-factor model⁶¹ and
323 'extended Braden'⁶¹, revealing variable performance with high uncertainty.^{32 29 42} Another two
324 modified versions of the Norton scale (by Ek⁶², and by Bienstein⁶³) were also included in one review's
325 meta-analyses³², but only risk ratios were reported. Three additional scales (revised "Jackson &
326 Cubbin"⁶⁴, EMINA⁶⁵ and PSPS⁶⁶) were evaluated in one statistical synthesis each.^{32 29} Full details can
327 be found in Appendix 5 Table S4.

328 Appendix 5 Table S5 reports data for another 17 risk prediction tools, each associated with a single
329 primary study (therefore not covered in detail in the text above), and another two tools,
330 Sunderland⁶⁷ and RAPS⁶⁸, which are assessed in two primary studies each.

331 Results from reviews evaluating the clinical effectiveness of risk prediction models

332 Table 4 provides an overview of results from four^{29 45 47 50 69} of the 10 reviews reporting clinical
333 effectiveness, including one Cochrane review⁴⁷ which identified two randomised controlled trials
334 (RCTs) of risk prediction tools and assessed risk of bias using the Cochrane RoB tool⁴⁸. The remaining
335 reviews used broader eligibility criteria for study inclusion and a range of different quality
336 assessment tools, with some reviews reaching varying conclusions about the methodological quality
337 of the same studies. Given the overlap in study inclusion between reviews, a summary of the
338 included comparative studies is provided below.

339 One individually randomised trial (Webster and colleagues⁷⁰) and one cluster randomised trial (Saleh
340 and colleagues⁷¹) were considered to be at high risk of bias by the Cochrane review authors. The
341 individually randomised trial⁷⁰ was included in three additional reviews^{29 44 47 50}, each of which
342 considered the trial to be 'good quality'²⁹, 'valid'⁴⁴, or 'high quality'⁵⁰. The trial was conducted in
343 1,231 hospital inpatients and found no evidence of a difference in PI incidence between patients
344 assessed with either the Waterlow scale or Ramstadius tool compared with clinical judgment alone
345 (RR 1.10, 95% CI 0.68, 1.81 for Waterlow and RR 0.79, 95% CI 0.46, 1.35 for Ramstadius). The trial
346 further showed no evidence of a difference in patient management or in PI severity when using a risk
347 assessment tool compared to clinical judgement.

348 The cluster randomised trial⁷¹ was considered to be of poor methodological quality in two reviews.²⁹
349 ⁴⁷ The trial included 521 patients at a military hospital and compared nurse training with mandatory
350 use of the Braden scale, to nurse training and optional use of the Braden scale, to no training. No
351 evidence of a difference in PI incidence was observed between groups: incidence rates were 22%,
352 22% and 15% (p=0.38), for the three groups respectively.

353 In both reviews by Lovegrove and colleagues,^{45 50} an uncontrolled comparison study⁷² was included.
354 The study assessed the clinical effectiveness of the Maelor scale,⁷³ and was rated as high quality
355 within the most recent review.⁵⁰ Preventive strategies and PI prevalence were compared across two
356 sites, an Irish hospital that used the Maelor scale (121 patients) and a Norwegian hospital that used
357 nurses' clinical judgement (59 patients). A higher rate of preventive strategies, as well as a lower PI
358 prevalence (12% vs. 54%), was reported for the Irish hospital. However, these results are likely to be
359 highly confounded by inherent differences in population and setting.

360 A non-randomised study by Gunningberg and colleagues⁷⁴ was included in two reviews, one of which
361 is reported in Table 4^{33 69} and was considered to be of relatively high quality. The study was
362 conducted in 124 patients in emergency and orthopaedic units and compared the use of a pressure
363 ulcer risk alarm sticker for patients with a modified Norton Score of <21 (indicating high-risk patients)
364 to standard care. No significant difference in the incidence of pressure ulcers between the Norton
365 scale and standard care groups was observed.

366 A non-randomised study⁷⁵ conducted in 233 hospice inpatients was included in three reviews,^{29 33 69}
367 one of which is reported in Table 4.⁶⁹ The study met six of eight quality criteria used by Health
368 Quality Ontario.⁶⁹ Use of a modified version of the Norton scale (Norton modified by Bale), in
369 conjunction with standardised use of preventive interventions based on risk score, was found to be
370 associated with lower risk of pressure ulcers when compared with nurses' clinical judgment alone
371 (RR 0.11, 95% CI 0.03, 0.46). The lack of randomisation limits the reliability of this result, and review
372 authors report that the modified Norton scale had not been validated.

373 Finally, a 'before-and-after' study⁷⁶ of 181 patients in various hospital settings was included in two
374 reviews.^{33 69} The Health Quality Ontario considered the study to meet all quality criteria.⁶⁹ Use of the

Table 4. Systematic reviews evaluating clinical effectiveness

Review author (publication year)	Models included	Setting of included studies; study design; sample size	Included outcomes	Brief description of study quality	Relevant results from included studies
Lovegrove ⁵⁰ (2021)	Braden; Maelor score; Norton; Ramstadius; Waterlow	Acute care hospital n=1, inpatient units n=1, ICU n=1, internal medicine and oncology wards n=1; Design: cross-sectional survey n=2, RCT n=1, observational inter-rater reliability n=1; Sample size 45 to 1231	PI risk scores; PI incidence; PI preventative interventions; interrater reliability (reliability results covered in Appendix 5)	RoB assessed using JBI tools or analytical cross-sectional study appraisal checklist. The RCT was judged as high quality. Of the remaining studies, two were judged as high quality and one as moderate quality; inclusion criteria not clearly stated and no strategies to deal with confounding.	<ul style="list-style-type: none"> • There were no differences in patient management ('pressure care plan' and use of a special mattress) based on PI risk assessment method (clinical judgement, Ramstadius tool or Waterlow score). PI incidence difference between groups not significant ($p = 0.44$) (Webster 2011⁷⁰). • A hospital that used the Maelor scale reported a higher rate of PI preventive strategies, and a lower PI prevalence (12% vs. 54%), than a site that used nurses' clinical judgement (Moore 2015⁷²).
Moore ⁴⁷ (2019)	Braden; Waterlow; Ramstadius	Military hospital n=1, internal medicine and oncology wards n=1; Design: RCT n=1, cluster randomised trial n=1; Sample sizes 286 and 1231	PI incidence; severity of PIs	RoB assessed using Cochrane tool (Higgins 2011 ⁴⁸). Both studies at high RoB due to blinding issues. One study at RoB also due to baseline imbalance and incorrect analyses.	<ul style="list-style-type: none"> • No differences in PI incidence when using Braden scale or clinical judgement (Braden vs. clinical judgement+training, RR 0.97, 95% CI 0.53-1.77; Braden vs clinical judgement RR 1.43, 95% CI 0.77-2.68) (Saleh 2009⁷¹). • No difference in PI incidence when using a risk assessment tool compared to clinical judgement (RR 1.10, 95% CI 0.68-1.81 and RR 0.79, 95% CI 0.46-1.35, for Waterlow and Ramstadius respectively) (Webster 2011⁷⁰). • No difference in PI severity based on risk assessment tools vs. clinical judgement (Webster 2011⁷⁰).
Chou ²⁹ (2013)	Norton modified by Bale; Braden; Waterlow; Ramstadius	Hospital n=2, hospice n=1; Design: non-randomised n=1, cluster randomised trial n=1, RCT n=1; Sample size 240 to 1231	PI incidence, severity of PIs; PI preventative interventions	RoB assessed with criteria consistent with AHRQ Methods Guide for Effectiveness and Comparative Effectiveness Reviews. One RCT was rated as good quality and the other as poor due to randomisation and blinding issues. The cohort study was rated as poor; there were blinding issues and confounding was not investigated.	<ul style="list-style-type: none"> • No difference in PI incidence when using a risk assessment tool compared to clinical judgement (RR 1.10, 95% CI 0.68-1.81 and RR 0.79, 95% CI 0.46-1.35, for Waterlow and Ramstadius respectively) (Webster 2011⁷⁰). • The modified version of the Norton scale with use of preventive interventions is associated with lower risk of PIs compared with clinical judgment (RR 0.11, 95% CI 0.03-0.46) (Bale 1995⁷⁵). • No difference in risk of PIs when one of three interventions was used (22% vs. 22% vs. 15%, $p=0.38$ for nurse training+mandatory Braden scale, nurse training+optional Braden scale and no training respectively) (Saleh 2009⁷¹).

Review author (publication year)	Models included	Setting of included studies; study design; sample size	Included outcomes	Brief description of study quality	Relevant results from included studies
Health Quality Ontario ⁶⁹ (2009)	Norton; Norton modified by Bale; Norton modified by Ek 97	Hip fracture inpatients n=1, palliative care/hospice n=1, neurosurgery, general medicine, orthopaedic, and oncology units n=1; Design: prospective controlled (contemporaneous controls) n=1, before-and-after n=1; Sample size 124 to 223	PI incidence; PI preventative interventions	RoB assessment criteria name not given. Two studies met 6/8 and one study met all quality assessment requirements. In the studies that didn't meet all requirements, there were blinding and loss to follow-up issues. One study used a version of the Norton scale that was not validated.	<ul style="list-style-type: none"> • Compared a strategy that gave high-risk patients (based on modified Norton score) a risk alarm sticker to standard care. No significant difference between the groups in the incidence of PIs (Gunningberg 1999⁷⁴). • Compared a strategy where patients received a pressure support system allocated according to the modified Norton scale to one where the nurse chose whether to give a special mattress. Using the scale significantly reduced the incidence of PIs (22.4% vs. 2.5%, p<0.0001) (Bale 1995⁷⁵). • Compared the Norton scale with training to standard care. There was a significant difference in the number of preventative interventions (18.96 vs. 10.75, for Norton and usual care respectively). Interventions were used earlier for Norton vs. usual care (on day 1, 61% vs. 50%, p<0.002). No significant difference in the incidence of PIs between the groups (Hodge 1990⁷⁶).

376

377

AHRQ – Agency for Healthcare Research; CASP – Critical Appraisal Skills Checklist; CI – confidence interval; ICU – intensive care unit; JBI – Joanna Briggs Institute; PI – pressure injury; RCT – randomised controlled trial; RoB – risk of bias; RR – Risk Ratio; S.S. – Suriadi Sanada Scale.

378 Norton scale with additional training for staff was associated with significant differences in the
379 number of preventative interventions prescribed compared to standard care (18.96 vs. 10.75,
380 respectively). Preventative interventions were also introduced earlier in the intervention group (on
381 day 1, 61% vs. 50%, $P < 0.002$ for Norton and usual care, respectively). However, no significant
382 difference in the incidence of PIs was detected between the groups.

383 DISCUSSION

384 This umbrella review summarises data from a total of 23 systematic reviews of studies evaluating the
385 clinical utility of a total of 63 PI risk prediction tools. Despite the large number of available reviews,
386 quality assessment using an adaptation of AMSTAR-2 suggested that the majority were conducted to
387 a relatively poor standard or did not meet reporting standards for systematic reviews.^{14 26} Of the 15
388 items included in AMSTAR-2, only two (for accuracy reviews) and four (for effectiveness reviews)
389 criteria were more consistently met (more than 60% of reviews scoring ‘Yes’). All other criteria were
390 fully met by less than half of reviews. The primary studies included in the reviews were particularly
391 poorly described in the accuracy reviews, making it difficult to determine exactly what was evaluated
392 and in whom. In particular, the source of the data was poorly reported. Only one review³³ explicitly
393 restricted to accuracy estimates from external validations, and only one review²⁹ described whether
394 estimates were sourced from training or external data. The extent to which we could reliably
395 describe and comment on the content of the reviews is limited and high-quality evidence for the
396 accuracy and clinical effectiveness of PI risk prediction models may be lacking.

397 Prognostic accuracy of risk prediction models

398 Of the 16 reviews focused on the predictive accuracy of included models, only two used appropriate
399 methods for both quality assessment and statistical synthesis of accuracy data^{36 40}, one of which³⁶
400 evaluated only the Braden scale. Only one review³³ pre-specified the exclusion of studies reporting
401 tool development only, one review restricted to “validated risk assessment instruments” only²⁹, and
402 none of the reviews discussed the importance of appropriate validation of prediction models. Only
403 two reviews conducted meta-analyses at different cut-offs for determination of high risk^{29 36}; the
404 remaining reviews combined data regardless of the threshold used. Combining data across different
405 thresholds to estimate summary sensitivity and specificity is discouraged as it yields clinically
406 uninterpretable and non-generalisable estimates, because the estimates do not relate to a particular
407 threshold.²⁵

408 Results of meta-analyses suggested that risk prediction scales have moderate sensitivities and
409 somewhat lower specificities, typically in the range of around 70% to 85% for sensitivity and as low
410 as 30% to 40% for specificity for some tools. Without a detailed review of the primary study
411 publications for these models, it is not possible to assess which, if any, of these risk assessment
412 scales might outperform the others. It seems that limited comparative studies comparing the
413 accuracy of different tools are available.

414 For the ML-based models, one review³⁰ meta-analysed accuracy data by algorithm type. The results
415 of the meta-analyses are not informative for clinical practice but may be a useful way of identifying
416 which ML algorithms may be more suited to the data. Results suggested that specificities for random
417 forest or decision tree models could reach 90% or above with associated sensitivities in the range of
418 66% to 72%, however relatively wide confidence intervals around these summary estimates reflect
419 considerable variation in model performance. Moreover, some of these estimates came from internal
420 validations within model development studies, and may not be transferable to other settings.⁷⁷
421 Authors should make it clear where accuracy estimates are derived from to avoid overinterpretation
422 of results.

423 Clinical effectiveness of risk prediction scales

424 Prediction models, like any test used for diagnostic or prognostic purposes, require evaluation in the
425 care pathway to identify the extent to which their use can impact on health outcomes.⁷⁸ Of the 10
426 reviews assessing clinical effectiveness of PI risk prediction tools, the only primary studies suggesting
427 potential patient benefits from the use of risk prediction tools,^{72 75 76} were non-randomised and are
428 likely to be at high risk of bias. In contrast, two randomised trials,^{70 71} (both considered at high risk of
429 bias by the Cochrane review⁴⁷) suggest that use of structured risk assessment tools does not reduce
430 the incidence of PIs. We should recognise that effectiveness outcomes largely depend on the
431 availability and efficacy of preventative measures, and conclusions regarding the clinical effectiveness
432 of these risk assessment tools cannot be confidently drawn from the limited evidence available. All
433 reviews included studies that assessed the use of risk assessment scales developed by experts, and
434 no evidence is available evaluating the clinical effectiveness of empirically derived prediction models
435 or ML algorithms.

436 Other existing evidence

437 Moore and colleagues⁴⁷ recently updated their review (published after our search was conducted⁷⁹)
438 and reported no new randomised trials that assessed the effect of risk assessment tools on PI
439 incidence.

440 We have separately reviewed⁹ available evidence for the development and validation of risk
441 prediction tools for PI occurrence. Almost half (52/116, 45%) of available tools were developed using
442 ML methods (as defined by review authors), 40% (46/116) were based on clinical expertise or
443 unclear methods, and only 18 (16%) were identified as having used statistical modelling methods.
444 The reviews varied in methodological quality and reporting; however, the reporting of prediction
445 model development in the original primary studies appears to be poor. For example, across all
446 prediction tools identified, the internal validation approach was unclear and unidentifiable for 70%
447 (81/116) of tools, and only one review identified and included external validation studies (n=7
448 studies).

449 ML-based models may have potential for identifying those at risk of PI, as suggested by one review³⁰
450 included in this umbrella review. However, it is important to consider the lack of transparency in
451 reporting of model development methods and model performance, and the concerning lack of
452 model validation in populations outside of the original model development sample.⁹

453 Strengths and limitations

454 We have conducted the first umbrella review that summarise the prognostic accuracy and clinical
455 effectiveness of prediction models for risk of PI. We followed Cochrane guidance¹³, with a highly
456 sensitive search strategy designed by an experienced information specialist. Although we excluded
457 non-English publications due to time and resource constraints, where possible these publications
458 were used to identify additional eligible risk prediction models. To some extent, our review is limited
459 by the use of AMSTAR-2 for quality assessment of included reviews. AMSTAR-2 was not designed for
460 assessing systematic reviews of diagnostic or prognostic studies. Although we made some
461 adaptations, many of the existing and amended criteria relate to the quality of reporting of the
462 reviews as opposed to methodological quality. There is scope for further work to establish criteria for
463 assessing systematic reviews of prediction models.

464 The primary limitation of our study lies in the limited detail available on risk prediction tools and
465 their performance within the included systematic reviews. To ensure comprehensive model
466 identification, we adopted a broad definition of 'systematic,' potentially influencing the depth of
467 information provided in the reviews, and the reporting quality in many primary studies contributing

468 to these reviews may be suboptimal. Notably, excluding ML-based models, over half of the existing
469 risk prediction tools were published prior to 2000, before the publication of original versions of
470 reporting guidelines for diagnostic accuracy studies⁸⁰ and risk prediction models.⁸¹

471 CONCLUSIONS

472 In conclusion, this umbrella review comprehensively summarises the prognostic accuracy and clinical
473 effectiveness of risk prediction tools for developing PIs. The included systematic reviews used poor
474 methodology and reporting, limiting our ability to reliably describe and evaluate their content. ML-
475 based models demonstrated potential, with high specificity reported for some models. Wide
476 confidence intervals highlight the variability in current evaluations, and external validation of ML
477 tools may be lacking. The prognostic accuracy of clinical scales and statistically derived prediction
478 models has a substantial range of specificities and sensitivities, motivating further model
479 development with high quality data and appropriate statistical methods.

480 Regarding clinical effectiveness, a reduction of PI incidence is unclear due the overall uncertainty and
481 potential biases in available studies. This underscores the need for further research in this critical
482 area, once promising prediction tools have been developed and appropriately validated. In particular,
483 the clinical impact of newer ML-based models currently remains largely unexplored. Despite these
484 limitations, our umbrella review provides valuable insights into the current state of PI risk prediction
485 tools, emphasising the need for robust research methods to be used in future evaluations.

486

487 Acknowledgements

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

490 Author Contributions

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

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

494 **Formal analysis:** Bethany Hillier, Katie Scandrett, Jacqueline Dinnes

495 **Funding acquisition:** Yemisi Takwoingi, Vladica Velickovic, Jacqueline Dinnes

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

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

499 **Project administration:** Bethany Hillier, Yemisi Takwoingi, Jacqueline Dinnes

500 **Resources:** Bethany Hillier, Katie Scandrett

501 **Supervision:** Yemisi Takwoingi, Jacqueline Dinnes

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

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

505 Funding

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

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

514 Conflicting Interests

515 I have read the journal's policy and the authors of this manuscript have the following competing
516 interests: VV is an employee of Paul Hartmann AG; ES and THB received consultancy fees from Paul
517 Hartmann AG. All other authors received no personal funding or personal compensation from Paul
518 Hartmann AG and have declared that no competing interests exist.

519

520 References

- 521 1. Li Z, Lin F, Thalib L, et al. Global prevalence and incidence of pressure injuries in hospitalised adult
522 patients: A systematic review and meta-analysis. *International Journal of Nursing Studies*
523 2020;105:103-546. doi: 10.1016/j.ijnurstu.2020.103546
- 524 2. Padula WV, Delarmente BA. The national cost of hospital-acquired pressure injuries in the United
525 States. *Int Wound J* 2019;16(3):634-40. doi: 10.1111/iwj.13071 [published Online First:
526 2019/01/28]
- 527 3. Sullivan N, Schoelles K. Preventing In-Facility Pressure Ulcers as a Patient Safety Strategy. *Annals of*
528 *Internal Medicine* 2013;158(5.2):410-16. doi: 10.7326/0003-4819-158-5-201303051-00008
- 529 4. Institute for Quality and Efficiency in Health Care (IQWiG). Preventing pressure ulcers. Cologne,
530 Germany 2006 [updated 2018 Nov 15. Available from:
531 <https://www.ncbi.nlm.nih.gov/books/NBK326430/?report=classic> accessed Feb 2023].
- 532 5. Braden B, Bergstrom N. A Conceptual Schema for the Study of the Etiology of Pressure Sores.
533 *Rehabilitation Nursing* 1987;12(1):8-16. doi: 10.1002/j.2048-7940.1987.tb00541.x
- 534 6. Bergstrom N, Braden BJ, Laguzza A, et al. The Braden Scale for Predicting Pressure Sore Risk. *Nurs*
535 *Res* 1987;36(4):205-10.
- 536 7. Norton D. Geriatric nursing problems. *Int Nurs Rev* 1962;9:39-41.
- 537 8. Waterlow J. Pressure sores: a risk assessment card. *Nursing Times* 1985;81:49-55.
- 538 9. Hillier B, Scandrett K, Coombe A, et al. Development and validation of risk prediction tools for
539 pressure injury occurrence: An umbrella review (pre-print). *MedRxiv* 2024 doi: TBC (UPDATE
540 ME)
- 541 10. Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC* 2009;19(4):203-11.
542 [published Online First: 2009/01/20]
- 543 11. Leeflang MM, Rutjes AW, Reitsma JB, et al. Variation of a test's sensitivity and specificity with
544 disease prevalence. *CMAJ* 2013;185(11):E537-44. doi: 10.1503/cmaj.121286 [published
545 Online First: 2013/06/24]
- 546 12. Maiga A, Farjah F, Blume J, et al. Risk Prediction in Clinical Practice: A Practical Guide for
547 Cardiothoracic Surgeons. *Ann Thorac Surg* 2019;108(5):1573-82. doi:
548 10.1016/j.athoracsur.2019.04.126 [published Online First: 2019/06/27]
- 549 13. Pollock M, Fernandes RM BL, Pieper D, Hartling L,. Chapter V: Overviews of Reviews. In: Higgins
550 JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA ed. *Cochrane Handbook for*
551 *Systematic Reviews of Interventions* version 63 (updated February 2022). Available from
552 www.training.cochrane.org/handbook: Cochrane 2022.
- 553 14. McInnes MDF, Moher D, Thoms BD, et al. Preferred Reporting Items for a Systematic Review and
554 Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA*
555 2018;319(4):388-96. doi: 10.1001/jama.2017.19163
- 556 15. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc*
557 2001;8(4):391-7. doi: 10.1136/jamia.2001.0080391 [published Online First: 2001/06/22]
- 558 16. Wilczynski NL, Haynes RB. Optimal Search Strategies for Detecting Clinically Sound Prognostic
559 Studies in EMBASE: An Analytic Survey. *Journal of the American Medical Informatics*
560 *Association* 2005;12(4):481-85. doi: 10.1197/jamia.M1752
- 561 17. Geersing G-J, Bouwmeester W, Zuithoff P, et al. Search Filters for Finding Prognostic and
562 Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews. *PLOS ONE*
563 2012;7(2):e32844. doi: 10.1371/journal.pone.0032844
- 564 18. NHS. Pressure ulcers: revised definition and measurement. Summary and recommendations 2018
565 [Available from: [https://www.england.nhs.uk/wp-content/uploads/2021/09/NSTPP-](https://www.england.nhs.uk/wp-content/uploads/2021/09/NSTPP-summary-recommendations.pdf)
566 [summary-recommendations.pdf](https://www.england.nhs.uk/wp-content/uploads/2021/09/NSTPP-summary-recommendations.pdf) accessed Feb 2023].
- 567 19. Haesler E. European Pressure Ulcer Advisory Panel, National Pressure Injury Advisory Panel and
568 Pan Pacific Pressure Injury Alliance. Prevention and Treatment of Pressure Ulcers/Injuries:
569 Clinical Practice Guideline. 2019 [Available from: <https://internationalguideline.com/2019>
570 accessed Feb 2023].

- 571 20. AHCPR. Pressure ulcer treatment. : Agency for Health Care Policy and Research 1994:1-25.
572 21. Harker J. Pressure ulcer classification: the Torrance system. *Journal of Wound Care* 2000;9(6):275-
573 77. doi: 10.12968/jowc.2000.9.6.26233
- 574 22. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical Appraisal and Data Extraction for
575 Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLOS Medicine*
576 2014;11(10):e1001744. doi: 10.1371/journal.pmed.1001744
- 577 23. Cochrane. DE form example prognostic models - scoping review: The Cochrane Collaboration: The
578 Prognosis Methods Group; [Available from: <https://methods.cochrane.org/prognosis/tools>
579 accessed Feb 2023].
- 580 24. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that
581 include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*
582 2017;358:j4008. doi: 10.1136/bmj.j4008
- 583 25. P M, Y T, JJ D, et al. Understanding meta-analysis. Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi
584 Y (editors) Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version
585 2.0 ed: Cochrane, 2023 (updated July 2023).
- 586 26. Moher D, Liberati A, Tetzlaff J, et al. Preferred Reporting Items for Systematic Reviews and Meta-
587 Analyses: The PRISMA Statement. *PLOS Medicine* 2009;6(7):e1000097. doi:
588 10.1371/journal.pmed.1000097
- 589 27. Munoz N, Posthauer ME. Nutrition strategies for pressure injury management: Implementing the
590 2019 International Clinical Practice Guideline. *Nutrition in Clinical Practice* 2022;37(3):567-
591 82.
- 592 28. Park SH, Lee YS, Kwon YM. Predictive Validity of Pressure Ulcer Risk Assessment Tools for Elderly:
593 A Meta-Analysis. *Western journal of nursing research* 2016;38:459-83. doi:
594 10.1177/0193945915602259
- 595 29. Chou R, Dana T, Bougatsos C, et al. Pressure ulcer risk assessment and prevention: a systematic
596 comparative effectiveness review. *Annals of internal medicine* 2013;159(1):28-38.
- 597 30. Qu C, Luo W, Zeng Z, et al. The predictive effect of different machine learning algorithms for
598 pressure injuries in hospitalized patients: A network meta-analyses. *Heliyon*
599 2022;8(11):e11361. doi: 10.1016/j.heliyon.2022.e11361
- 600 31. Wang N, Lv L, Yan F, et al. Biomarkers for the early detection of pressure injury: A systematic
601 review and meta-analysis. *Journal of Tissue Viability* 2022;31:259-67. doi:
602 10.1016/j.jtv.2022.02.005
- 603 32. García-Fernández FP, Pancorbo-Hidalgo PL, Agreda JJS. Predictive Capacity of Risk Assessment
604 Scales and Clinical Judgment for Pressure Ulcers: A Meta-analysis. *Journal of Wound Ostomy*
605 *& Continence Nursing* 2014;41(1):24-34. doi: 10.1097/01.WON.0000438014.90734.a2
- 606 33. Pancorbo-Hidalgo PL, Garcia-Fernandez FP, Lopez-Medina IM, et al. Risk assessment scales for
607 pressure ulcer prevention: a systematic review. *J Adv Nurs* 2006;54(1):94-110. doi:
608 10.1111/j.1365-2648.2006.03794.x
- 609 34. Chen HL, Shen WQ, Liu P. A Meta-analysis to Evaluate the Predictive Validity of the Braden Scale
610 for Pressure Ulcer Risk Assessment in Long-term Care. *Ostomy/wound management*
611 2016;62(9):20-8.
- 612 35. He W, Liu P, Chen HL. The Braden Scale cannot be used alone for assessing pressure ulcer risk in
613 surgical patients: a meta-analysis. *Ostomy/wound management* 2012;58:34-40.
- 614 36. Huang C, Ma Y, Wang C, et al. Predictive validity of the braden scale for pressure injury risk
615 assessment in adults: A systematic review and meta-analysis. *Nursing open* 2021;8:2194-207.
616 doi: 10.1002/nop2.792
- 617 37. Park SH, Choi YK, Kang CB. Predictive validity of the Braden Scale for pressure ulcer risk in
618 hospitalized patients. *Journal of Tissue Viability* 2015;24:102-13. doi:
619 10.1016/j.jtv.2015.05.001

- 620 38. Wei M, Wu L, Chen Y, et al. Predictive Validity of the Braden Scale for Pressure Ulcer Risk in
621 Critical Care: A Meta-Analysis. *Nursing in critical care* 2020;25:165-70. doi:
622 10.1111/nicc.12500
- 623 39. Wilchesky M, Lungu O. Predictive and concurrent validity of the Braden scale in long-term care: A
624 meta-analysis. *Wound Repair and Regeneration* 2015;23:44-56. doi: 10.1111/wrr.12261
- 625 40. Zhang Y, Zhuang Y, Shen J, et al. Value of pressure injury assessment scales for patients in the
626 intensive care unit: Systematic review and diagnostic test accuracy meta-analysis. *Intensive &*
627 *critical care nursing* 2021;64:103009. doi: 10.1016/j.iccn.2020.103009
- 628 41. Kottner J, Dassen T, Tannen A. Inter- and intrarater reliability of the Waterlow pressure sore risk
629 scale: A systematic review. *International Journal of Nursing Studies* 2009;46:369-79. doi:
630 10.1016/j.ijnurstu.2008.09.010
- 631 42. Tayyib NAH, Coyer F, Lewis P. Pressure ulcers in the adult intensive care unit: a literature review of
632 patient risk factors and risk assessment scales. *Journal of Nursing Education and Practice*
633 2013;3(11):28-42.
- 634 43. Baris N, Karabacak BG, Alpar SE. The Use of the Braden Scale in Assessing Pressure Ulcers in
635 Turkey: A Systematic Review. *Advances in skin & wound care* 2015;28:349-57. doi:
636 10.1097/01.ASW.0000465299.99194.e6
- 637 44. Gaspar S, Peralta M, Marques A, et al. Effectiveness on hospital-acquired pressure ulcers
638 prevention: a systematic review. *International Wound Journal* 2019;16(5):1087-102. doi:
639 10.1111/iwj.13147
- 640 45. Lovegrove J, Miles S, Fulbrook P. The relationship between pressure ulcer risk assessment and
641 preventative interventions: a systematic review. *Journal of wound care* 2018;27(12):862-75.
- 642 46. Kelly J. Inter-rater reliability and Waterlow's pressure ulcer risk assessment tool. *Nurs Stand*
643 2005;19(32):86-7, 90-2. doi: 10.7748/ns2005.04.19.32.86.c3851
- 644 47. Moore ZEH, Patton D. Risk assessment tools for the prevention of pressure ulcers. *Cochrane*
645 *Database of Systematic Reviews* 2019 doi: 10.1002/14651858.CD006471.pub4
- 646 48. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of
647 bias in randomised trials. *BMJ* 2011;343:d5928. doi: 10.1136/bmj.d5928
- 648 49. AHRQ Methods for Effective Health Care. Methods Guide for Effectiveness and Comparative
649 Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US)
650 2008.
- 651 50. Lovegrove J, Ven S, Miles SJ, et al. Comparison of pressure injury risk assessment outcomes using
652 a structured assessment tool versus clinical judgement: A systematic review. *Journal of*
653 *Clinical Nursing* 2021 doi: 10.1111/jocn.16154 [published Online First: 2021/12/01]
- 654 51. Song M, Choi KS. Factors predicting development of decubitus ulcers among patients admitted
655 for neurological problems. *The Journal of Nurses Academic Society* 1991;21(1):16-26.
- 656 52. Pang SM, Wong TK. Predicting pressure sore risk with the Norton, Braden, and Waterlow scales in
657 a Hong Kong rehabilitation hospital. *Nursing Research* 1998;47(3):147-53.
- 658 53. Park SH, Lee HS. Assessing Predictive Validity of Pressure Ulcer Risk Scales- A Systematic Review
659 and Meta-Analysis. *Iranian journal of public health* 2016;45(2):122-33.
- 660 54. Cubbin B, Jackson C. Trial of a pressure area risk calculator for intensive therapy patients.
661 *Intensive Care Nursing* 1991;7(1):40-44.
- 662 55. González-Ruiz J, Carrero AG, Blázquez MH, et al. Factores de riesgo de las úlceras por presión en
663 pacientes críticos. *Enfermería Clínica* 2001;11(5):184-90.
- 664 56. Macaskill P, Gatsonis C, Deeks J, et al. Cochrane handbook for systematic reviews of diagnostic
665 test accuracy: Version, 2010.
- 666 57. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for
667 diagnostic test data. *Stat Med* 2002;21(9):1237-56. doi: 10.1002/sim.1099
- 668 58. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a
669 summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med*
670 1993;12(14):1293-316. doi: 10.1002/sim.4780121403

- 671 59. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new
672 meta-analytic method. *Med Decis Making* 1993;13(4):313-21. doi:
673 10.1177/0272989x9301300408
- 674 60. Kwong E, Pang S, Wong T, et al. Predicting pressure ulcer risk with the modified Braden, Braden,
675 and Norton scales in acute care hospitals in Mainland China. *Appl Nurs Res* 2005;18(2):122-8.
676 doi: 10.1016/j.apnr.2005.01.001
- 677 61. Halfens R, Van Achterberg T, Bal R. Validity and reliability of the Braden scale and the influence of
678 other risk factors: a multi-centre prospective study. *International Journal of Nursing Studies*
679 2000;37(4):313-19.
- 680 62. Ek AC. Prediction of pressure sore development. *Scand J Caring Sci* 1987;1(2):77-84. doi:
681 10.1111/j.1471-6712.1987.tb00603.x
- 682 63. Bienstein C. Risikopatienten erkennen mit der erweiterten Nortonskala [Risk patients detected
683 with the extended Norton scale]. *Dekubitus - Prophylaxe und Therapie*. Frankfurt/Main:
684 Verlag Krankenpflege 1991.
- 685 64. Jackson C. The revised Jackson/Cubbin Pressure Area Risk Calculator. *Intensive Crit Care Nurs*
686 1999;15(3):169-75. doi: 10.1016/s0964-3397(99)80048-2
- 687 65. Fuentelsaz C. Validation of the EMINA scale: tool for the evaluation of risk of developing pressure
688 ulcers in hospitalized patients. *Enferm Clin [Internet]* 2001;11(3):97-103.
- 689 66. Lowthian P. The practical assessment of pressure sore risk. *Care-Science and Practice*
690 1987;5(4):3-7.
- 691 67. Lowery MT. A pressure sore risk calculator for intensive care patients: 'the Sunderland
692 experience'. *Intensive Crit Care Nurs* 1995;11(6):344-53. doi: 10.1016/s0964-3397(95)80452-
693 8
- 694 68. Lindgren M, Unosson M, Krantz AM, et al. A risk assessment scale for the prediction of pressure
695 sore development: reliability and validity. *Journal of advanced nursing* 2002;38(2):190-99.
- 696 69. Ontario HQ. Pressure ulcer prevention: an evidence-based analysis. *Ontario health technology
697 assessment series* 2009;9(2):1-104.
- 698 70. Webster J, Coleman K, Mudge A, et al. Pressure ulcers: effectiveness of risk-assessment tools. A
699 randomised controlled trial (the ULCER trial). *BMJ Quality & Safety*
700 2011;20(4):297. doi: 10.1136/bmjqs.2010.043109
- 701 71. Saleh M, Anthony D, Parboteeah S. The impact of pressure ulcer risk assessment on patient
702 outcomes among hospitalised patients. *J Clin Nurs* 2009;18(13):1923-9. doi: 10.1111/j.1365-
703 2702.2008.02717.x [published Online First: 2009/04/03]
- 704 72. Moore Z, Johansen E, Etten Mv, et al. Pressure ulcer prevalence and prevention practices: a cross-
705 sectional comparative survey in Norway and Ireland. *Journal of Wound Care* 2015;24(8):333-
706 39. doi: 10.12968/jowc.2015.24.8.333
- 707 73. Moore Z, Pitman S. Towards establishing a pressure sore prevention and management policy in
708 an acute hospital setting. *The All Ireland Journal of Nursing and Midwifery* 2000;1(1):7-11.
- 709 74. Gunningberg L, Lindholm C, Carlsson M, et al. Implementation of risk assessment and
710 classification of pressure ulcers as quality indicators for patients with hip fractures. *J Clin
711 Nurs* 1999;8(4):396-406. doi: 10.1046/j.1365-2702.1999.00287.x
- 712 75. Bale S, Finlay I, Harding KG. Pressure sore prevention in a hospice. *J Wound Care* 1995;4(10):465-
713 8. doi: 10.12968/jowc.1995.4.10.465
- 714 76. Hodge J, Mounter J, Gardner G, et al. Clinical trial of the Norton Scale in acute care settings. *Aust
715 J Adv Nurs* 1990;8(1):39-46.
- 716 77. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and
717 external validation. *J Clin Epidemiol* 2016;69:245-7. doi: 10.1016/j.jclinepi.2015.04.005
718 [published Online First: 2015/04/18]
- 719 78. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model
720 updating, and impact assessment. *Heart* 2012;98(9):691-8. doi: 10.1136/heartjnl-2011-
721 301247 [published Online First: 2012/03/07]

- 722 79. Moore Z, Avsar P, O'Connor T, et al. A systematic review of movement monitoring devices to aid
723 the prediction of pressure ulcers in at-risk adults. *International Wound Journal*
724 2023;20(2):579-608. doi: 10.1111/iwj.13902
- 725 80. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic
726 accuracy: explanation and elaboration. *Ann Intern Med* 2003;138(1):W1-12. doi:
727 10.7326/0003-4819-138-1-200301070-00012-w1
- 728 81. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction
729 model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann*
730 *Intern Med* 2015;162(1):W1-73. doi: 10.7326/m14-0698