Validation and clinical impact of paediatric pressure ulcer risk assessment scales: A systematic review.

Kottner J[1], Hauss A, Schlüer AB, Dassen T.

International Journal of Nursing Studies. 2013 Jun;50(6):807-18. doi: 10.1016/j.ijnurstu.2011.04.014. Epub 2011 Jun 8.

報告者 楊素月

2014/10/07

緣由

- 本院是否應用本篇文獻，做為本院 Skin Bundle Care 的規範及查檢表?
- Overall:

☑ Yes
☐ No

– 本篇許多建議係源自於NPUAP EPUAP 2009年 guideline，下次Journal Club 將評讀另一篇2013年的文獻: Pressure Ulcer Treatment Strategies Comparative Effectiveness，代收集更完善的實證資料後，再進一步整合

續曲

BACKGROUND

Pressure ulcer risk assessment using an age-appropriate, valid and reliable tool is recommended for clinical paediatric practice.

Introduction

- A pressure ulcer (PU) is localized injury to the skin and/or underlying tissue as a result of pressure, or pressure in combination with shear (National Pressure Ulcer Advisory Panel (NPUAP) and European Pressure Ulcer Advisory Panel (EPUAP), 2009).
- While this phenomenon has gained a great deal of **attention in adults**, **far less is known about PUs in the paediatric population** including children from birth to adolescence (Baharestani and Pope, 2007).
- Recent investigations indicate that PUs are common **in infants and children**: reported **PU prevalence rates** including all PU categories range from approximately **3%** (Dixon and Ratliff, 2005 and Noonan et al., 2006) **to 23%** (Suddaby et al., 2005) **to 27%** (Schlüer et al., 2009).

步驟 1：系統性文獻回顧探討的問題為何？

P	0~18yrs children
I	Pressure ulcer risk assessment Instrument
C	/
O	Pressure ulcer incidence (Sensitivity and Specificity)

步驟 2：系統性文獻回顧的品質如何？(FAITH)

F - 研究是否找到 (Find) 所有的相關證據？

最好的狀況是？
良好的文獻搜尋至少應包括二個主要的資料庫(如：Medline, Cochrane考科藍實證醫學資料庫, EMBASE 等)，並且加上文獻引用檢索(參考文獻中相關研究、Web of Science, Scopus或 Google Scholar)、試驗登錄資料等。文獻搜尋應不只限於英文，並且應同時使用 MeSH字串及一般檢索詞彙(text words)。

Methods Search →P. 809

A concurrent search was run in the **databases MEDLINE** (1950 to December 03, 2010) and **EMBASE** (1989 to December 03, 2010) via Ovid. The database **CINAHL** (1982 to December 2010) was searched **using EBSCOHOST.** To identify as many relevant sources as possible a broad search strategy was applied containing terms to identify the population and the condition. Due to incomplete reporting, inconsistent indexing, and a lack of availability of subject headings, no other search filters were used to maximize sensitivity (De Vet et al., 2008, Leeflang et al., 2008 and Kottner et al., 2011).

步驟 2：系統性文獻回顧的品質如何？(FAITH)

F - 研究是否找到 (Find) 所有的相關證據？

最好的狀況是？
良好的文獻搜尋至少應包括二個主要的資料庫(如：Medline, Cochrane考科藍實證醫學資料庫, EMBASE 等)，並且加上文獻引用檢索(參考文獻中相關研究、Web of Science, Scopus或 Google Scholar)、試驗登錄資料等。文獻搜尋應不只限於英文，並且應同時使用 MeSH字串及一般檢索詞彙(text words)。

Methods Search →P. 809

For both searches subject headings and natural language terms were used in all fields:

#1 child* OR infant* OR pediatric*

#2 pressure ulcer OR pressure ulcers OR bedsore OR bedsores OR bed sore OR bed sores OR decubitus

#3 #1 AND #2

評讀結果：v 是 □否 □不清楚

Search results p.810

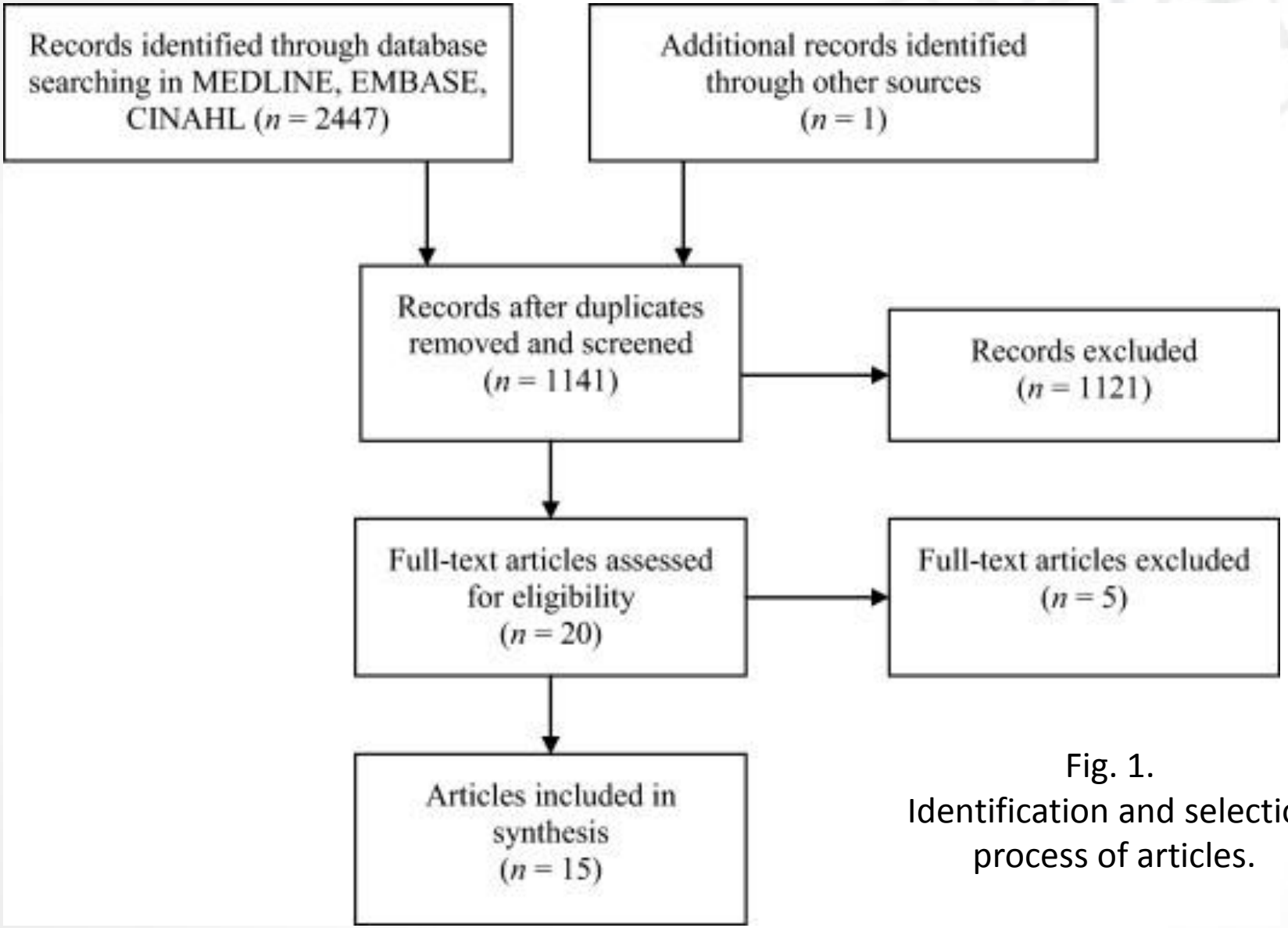


Fig. 1.
Identification and selection process of articles.

步驟 2：系統性文獻回顧的品質如何(FAITH)

A - 文獻是否經過嚴格評讀 (Appraisal)？

最好的狀況是？應根據不同臨床問題的文章類型，選擇適合的評讀工具，並說明每篇研究的品質(如針對治療型的臨床問題，選用隨機分配、盲法、及完整追蹤的研究類型)。

Data extraction and assessment of methodological quality →P. 809

Methodological quality of included papers **was assessed by using the QUADAS tool** (Whiting et al., 2003). This instrument was developed based on systematic reviews of methodological literature and three Delphi rounds with nine experts in diagnostic accuracy studies. This tool was designed to be used in systematic reviews and its use is recommended (Buntinx et al., 2009**). The QUADAS checklist is also recommended for the quality assessment of studies included Cochrane Systematic Reviews of Diagnostic Test Accuracy** (Reitsma et al., 2009).

評讀結果：v 是 □否 □不清楚

Table 1
Suggested items for paediatric pressure ulcer risk assessment scales.

Items	Setting/population (Author, year)												
	PICU (Bedi, 1993)	Paediatric (Quigley and Curley, 1996)	PICU (Garvin, 1997)	Neonates (Huffines and Logsdon, 1997)	Paediatric (Pickersgill, 1997)	PICU (Cockett, 1998)	PICU (Olding and Patterson, 1998)	Paediatric (Waterlow, 1998)	Paediatric (Barnes, 2004)	Paediatric (Suddaby et al., 2005)	Paediatric (Willock et al., 2007)	Paediatric burn patients (Gordon, 2009)	Total
Mobility	X	X	X	X	X	X	X	X	X	X	X	X	12
Nutrition/food intake		X	X	X		X	X	X		X	X	X	9
Continence	X				X	X	X		X		X	X	7
Weight according to age	X				X	X	X		X		X		6
Skin condition/ skin damage	X				X	X	X	X	X				6
Devices (casts, splints, tubes, lines)						X		X	X	X	X	X	6
Moisture		X	X	X						X		X	5
Tissue perfusion/ oxygenation/ cyanosis	X	X								X	X		4
Hemodynamic status/inotropic support	X					X	X					X	4
Neurological/mental state/Glasgow Coma Scale	X			X		X			X				4
Major surgery/trauma	X					X		X	X				4
Sensory perception		X	X						X	X			4
Activity		X		X						X			3
Friction & shear		X							X	X			3
Sedation						X			X		X		3
Respiratory status/ ventilation						X	X	X					3
Body temperature						X	X				X		3
Appetite	X				X								2
Medication	X				X								2
Gestational age				X									1
Age	X												1
Cachexia	X												1
Circulatory/vascular disease	X												1
Infection	X												1
Physical disabilities								X					1
Head injury								X					1
Severe illness								X					1
ICU								X					1
Anaemia											X		1
Albumin											X		1
Percent total body surface area burned												X	1
Prior/current PU												X	1
Increased prominences of bones												X	1

QUADAS 診斷準確性研究的質量評估 (Quality assessment of Diagnostic Accuracy Studies)

Table 2
Quality assessment according QUADAS.

Items	NSRAS	Huffines and Logsdon (1997)	Braden Q	Curley et al. (2003)	Burn pressure ulcer skin risk assessment scale	Gordon (2009)
1. Was the spectrum of patients representative of the patients who will receive the test in practice?		Y		Y		Y
2. Were selection criteria clearly described?		Y		Y		Y
3. Is the reference standard likely to correctly classify the target condition?		N		N		N
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?		Y		Y		Y
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?		Y		Y		Y
6. Did patients receive the same reference standard regardless of the index test result?		Y		Y		Y
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?		Y		Y		Y
8. Was the execution of the index test described in sufficient detail to permit replication of the test?		Y		Y		Y
9. Was the execution of the reference standard described in sufficient detail to permit its replication?		U		Y		Y
10. Were the index test results interpreted without knowledge of the results of the reference standard?		U		N		U
11. Were the reference standard results interpreted without knowledge of the results of the index test?		U		N		U
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?		Y		Y		Y
13. Were uninterpretable/intermediate test results reported?		U		U		N
14. Were withdrawals from the study explained?		U		U		N

Y=yes; N=no; U=unclear

QUADAS 診斷準確性研究的質量評估 (Quality assessment of Diagnostic Accuracy Studies)

Table 3
Results of prospective diagnostic accuracy studies.

Author	Instrument (score range)	Setting/sample	Exclusion criteria	Methods	Reference standard	Results	Comments (Quality Score according QUADAS)
Huffines and Logsdon (1997) **NSRAS**	Items "General physical condition", "Activity", "Nutrition" of the Neonatal Skin Risk Assessment Scale for Predicting Skin Break down (NSRAS) (3–12)	Neonatology, $n=32$, convenience, intact skin, mean age 33 weeks' gestation	Pre-existing PUs or lacerations on skin, genetic dermatologic conditions	Assessment with NSRAS within 24 h post delivery; skin observation period two month, until discharge or skin breakdown	Skin breakdown	Incidence 6/32 (19%); at cut-off 5 sensitivity = 0.83, specificity = 0.81, PPV = 0.50, NPV = 0.95, LR+ = 4.3, LR− = 0.2	Reference standard not clearly defined, LRs recalculated (7/14)
Curley et al. (2003) **Braden Q**	Braden Q (7–28)	PICU, $n=322$, convenience, intact skin, mean age 3 years (21 days to 8 years)	Pre-existing PUs, patients with intracardiac shunting and/or unrepaired congenital heart disease	Assessment with Braden Q and skin assessments by two trained nurses independently until discharge	Pressure ulcers stage I to IV (NPUAP, 1989)	Incidence 86/322 (26.7%); for stage II and higher PUs AUC = 0.83, at cut-off 16 sensitivity = 0.88, specificity = 0.58, PPV = 0.15, NPV = 0.98, LR+ = 2.1, LR− = 0.2 Only subscales "Sensory perception", "Mobility", "Tissue perfusion" contributed greater than 0.7 to AUC	LR− recalculated (8/14)
Gordon (2009) **Burn pressure ulcer skin risk assessment scale**	Burn Pressure Ulcer Skin Risk Assessment Scale (?)	3 Paediatric burn intensive care units, $n=163$, convenience, mean age 7.2 years (2 month to 18 years)	Not speaking English or Spanish, length of stay <3 days, primary diagnosis not acute burn injury, total body surface area burned >85%	Risk assessment on post-op day one, two or three; daily skin inspection by trained nurses until PU occurrence or discharge	Pressure ulcers stage I to IV, unstageable, DTI (NPUAP, 2007)	Incidence 44/163 (27%); based on a logistic regression model comparing predicted vs. observed PUs sensitivity = 0.54, specificity = 0.95; PPV = 0.80, NPV = 0.85, LR+ = 11.1, LR− = 0.48 Only items "Prior/current PU", "Percent total body surface area burned", "Number of splints" were significant predictors	Mean age and results recalculated; sample size determination for logistic regression was performed; no cut-off point established (9/14)

J. Kottner et al. / International Journal of Nursing Studies 50 (2013) 807–818

步驟 2：系統性文獻回顧的品質如何(FAITH)

A - 文獻是否經過嚴格評讀 (Appraisal) ?

最好的狀況是？
應根據不同臨床問題的文章類型，選擇適合的評讀工具，並說明每篇研究的品質(如針對治療型的臨床問題，選用隨機分配、盲法、及完整追蹤的研究類型)。

Data extraction and assessment of methodological quality→P. 809

Data from interrater **reliability and agreement studies** were gathered using a third data extraction sheet. It contained: authors, years, instrument, setting, sample of raters, sample of subjects, methods and results. The recently developed **QAREL checklist** (Lucas et al., 2010**) was used to evaluate the methodological quality of included interrater reliability and agreement studies**. In an iterative process a group of researchers with expertise in diagnostic research developed this tool for use in systematic reviews. The QAREL checklist was chosen because this seems to be the only available quality appraisal tool for reliability studies at the moment. Data extraction and methodological evaluation was conducted independently by two reviewers. Disagreements were resolved by consensus.

評讀結果：v 是 □否 □不清楚

QAREL研究可靠性質量評讀工具
(A quality appraisal tool for studies of reliability)

NSRAS (6-24) | Starkid skin scale (6-24) | Glamorgan scale (0-42) | Burn pressure ulcer skin risk assessment scale

Table 4
Quality assessment according QAREL.

Items	Huffines and Logsdon (1997)	Suddaby et al. (2005)	Willock et al. (2008)	Gordon (2008)	Gordon (2009)
1. Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied?	Y	Y	Y	U	Y
2. Was the test performed by raters who were representative of those to whom the authors intended the results to be applied?	Y	Y	Y	Y	U
3. Were raters blinded to the findings of other raters during the study?	Y	Y	Y	U	U
4. Were raters blinded to their own prior findings of the test under evaluation?	N/A	N/A	N/A	N/A	N/A
5. Were raters blinded to the subjects' disease status or the results of the accepted reference standard for the target disorder (or variable) being evaluated?	U	U	U	U	U
6. Were raters blinded to clinical information that was not intended to form part of the study design or testing procedure?	N/A	N/A	N/A	N/A	N/A
7. Were raters blinded to additional cues that are not part of the test?	U	N	U	U	U
8. Was the order of examination varied?	U	U	U	U	U
9. Was the stability (or theoretical stability) of the variable being measured taken into account when determining the suitability of the time interval among repeated measures?	Y	Y	Y	U	Y
10. Was the test applied correctly and interpreted appropriately?	U	U	U	U	U
11. Were appropriate statistical measures of agreement used?	N	N	Y	Y	Y

Y = yes; N = no; U = unclear; N/A = not applicable.

QAREL 研究可靠性質量評讀工具
(A quality appraisal tool for studies of reliability)

Table 5
Results of interrater reliability and agreement studies.

Author	Instrument (score range)	Setting	Raters (*k*)	Subjects (*n*)	Methods	Results	Comments (Quality Score according QAREL)
Huffines and Logsdon (1997)	Neonatal Skin Risk Assessment Scale for Predicting Skin Breakdown (NSRAS) (6–24)	Neonatology	Nurses providing care (*k* not stated)	Neonates, mean age 33 weeks' gestation (*n* = 32)	Independent assessment of neonates by researcher and the nurse providing care within 9 subsequent occasions (256 paired ratings)	*r* = –.27 to 1.00 across all items and rater pairs; *r* = 0.97 for "physical condition", "activity", "nutrition" across two raters	Calculation and kind of interrater reliability coefficients not clearly described (QS 4/9)
Suddaby et al. (2005)	Starkid Skin Scale (6–24)	Paediatric hospital	Staff nurses (*k* = 4), clinical nurse specialists (*k* = 3)	(*n* = 30)	Independent assessment of staff nurse and clinical nurse specialists on same day	*r* = 0.85 for total score; most differences (10/30) for "nutrition"	No interrater reliability study in a strict sense (QS 4/9)
Willock et al. (2008)	Glamorgan scale (0–42)	Paediatric wards of tertiary hospital	Randomly selected nurses (*k* = 15) and one researcher	(*n* = 15)	Independent assessment of children by researcher and one nurse within 10 min (15 paired ratings)	100% agreement between researcher and nurses rating all items; one disagreement for rating "inadequate nutrition"	No interrater reliability study in a strict sense (QS 5/9)
Gordon (2008)	Burn Pressure Ulcer Skin Risk Assessment Scale (?)	?	Burn nurses (*k* = 5)	(*n* = 21)	?	ICC = 0.99 for total score	QS (2/9)
Gordon (2009)	Burn Pressure Ulcer Skin Risk Assessment Scale (?)	3 paediatric burn intensive care units	Burn nurses (*k* = 2)	(*n* = 20)	Nurses rated patients at the same point in time	ICC = 0.99 for total score; ICC for items ranging from 0.86 to 1.00	(QS 3/9)

步驟 2：系統性文獻回顧的品質如何(FAITH)

I - 是否只納入 (included) 具良好效度的文章？

最好的狀況是？
僅進行文獻判讀是不足夠，系統性文獻回顧只納入至少要有一項研究結果是極小偏誤的試驗。

2.2. Study selection→P. 809

The results of the database and reference list searches were **screened by two investigators independently.** The **target population** consists of infants from 0 to 23 month (including premature newborns), children from 2 to 12 years, and adolescents from 13 to 18 years. No setting was excluded. **More specific inclusion criteria were set according to the four study questions:**

評讀結果：v 是 □否 □不清楚

Study selection

More specific inclusion criteria were set according to the four study questions:

(1) What PU risk scales for children currently exist?

(2) What is the diagnostic accuracy of their scores?

(3) Are the scores reliable and what is the degree of agreement?

(4) What is the clinical impact of risk scale scores in paediatric practice?

步驟 2：系統性文獻回顧的品質如何(FAITH)

I - 是否只納入 (included) 具良好效度的文章？

最好的狀況是？
僅進行文獻判讀是不足夠，系統性文獻回顧只納入至少要有一項研究結果是極小偏誤的試驗。

2.2. Study selection→P. 809

(1)What PU risk scales for children currently exist?

We included every source introducing or describing **a standardized PU risk scale.**

評讀結果：v 是 □否 □不清楚

步驟 2：系統性文獻回顧的品質如何(FAITH)

I - 是否只納入 (included) 具良好效度的文章？

最好的狀況是？
僅進行文獻判讀是不足夠，系統性文獻回顧只納入至少要有一項研究結果是極小偏誤的試驗。

2.2. Study selection→P. 809

(2)What is the diagnostic accuracy(診斷準確性) of their scores?

Studies were included when they used a prospective design and PU development (incidence) of any PU category as reference standard. Prospective designs are considered crucial for PU risk validation (Pancorbo-Hidalgo et al., 2006), because the degree of PU risk (quantified by risk scales) was regarded as the predictor for possible subsequent PU development. Consequently**, studies using cross sectional designs were excluded** due to their inability to establish relations between possible predictors and the outcome PU development. **Temporality can only investigated in longitudinal studies** (Twisk, 2003). Studies using other validation approaches (e.g., known groups) were also excluded.

評讀結果：v 是 □否 □不清楚

步驟 2：系統性文獻回顧的品質如何(FAITH)

I - 是否只納入 (included) 具良好效度的文章？

最好的狀況是？
僅進行文獻判讀是不足夠，系統性文獻回顧只納入至少要有一項研究結果是極小偏誤的試驗。

2.2. Study selection→P. 809

(3)Are the scores reliable and what is the degree of agreement?

All studies comparing PU risk scale scores of different raters using the same scale (interrater) or of the same raters using the same scale at different times (intrarater) were included (Kottner et al., 2011). No raters or rater pairings were excluded.

評讀結果：v 是 □否 □不清楚

步驟 2：系統性文獻回顧的品質如何(FAITH)

I - 是否只納入 (included) 具良好效度的文章？

最好的狀況是？
僅進行文獻判讀是不足夠，系統性文獻回顧只納入至少要有一項研究結果是極小偏誤的試驗。

2.2. Study selection→P. 809

(4)What is the clinical impact of risk scale scores in the paediatric practice?

To evaluate the clinical impact of PU risk scale scores in the paediatric practice we included RCTs, CCTs and before-after studies. The application of a pressure ulcer risk scale was regarded as intervention compared to the use of another or no risk scale. PU incidence (as defined by the authors) was regarded as outcome.

評讀結果：v 是 □否 □不清楚

步驟 2：系統性文獻回顧的品質如何(FAITH)

T - 作者是否以表格和圖表「總結」(total up) 試驗結果？

最好的狀況是？應該用至少 1 個摘要表格呈現所納入的試驗結果。若結果相近，可針對結果進行統合分析(meta-analysis)，並以「森林圖」(forest plot)呈現研究結果，最好再加上異質性分析。

文章內未使用統合分析(meta-analysis)及運用「森林圖」(forest plot)呈現研究結果，故亦無進行異質性分析。

評讀結果：□ 是 v否 □不清楚

步驟 2：系統性文獻回顧的品質如何(FAITH)

H - 試驗的結果是否相近 - 異質性 (Heterogeneity) ?

最好的狀況是？
在理想情況下，各個試驗的結果應相近或具同質性，若具有異質性，作者應評估差異是否顯著(卡方檢定)。根據每篇個別研究中不同的PICO及研究方法，探討造成異質性的原因。

文章內容未提到

評讀結果：□ 是 v否 □不清楚

RESULTS

- The search yielded 1141 hints, 15 publications describing or applying 12 paediatric pressure ulcer risk scales were included.
- Three of these scales (Neonatal Skin Risk Assessment Scale for Predicting Skin Breakdown, Braden Q Scale, Burn Pressure Skin Risk Assessment Scale) were investigated in prospective validation studies.
- Empirical evidence about interrater reliability and agreement is available for four instruments
  - Neonatal Skin Risk Assessment Scale for Predicting Skin Breakdown
  - Starkid Skin Scale
  - Glamorgan Scale
  - Burn Pressure Ulcer Risk Assessment Scale
- No studies were identified investigating the clinical impact.

CONCLUSIONS

- Sound empirical evidence about the performance of paediatric pressure ulcer risk assessment scales is lacking.
- Based on the few results of this review no instrument can be regarded as superior to the others.
- Whether the application of pressure ulcer risk assessment scales reduces the pressure ulcer incidence in paediatric practice is unknown.
- Maybe clinical judgement is more efficient in evaluating pressure ulcer risk than the application of risk scale scores.

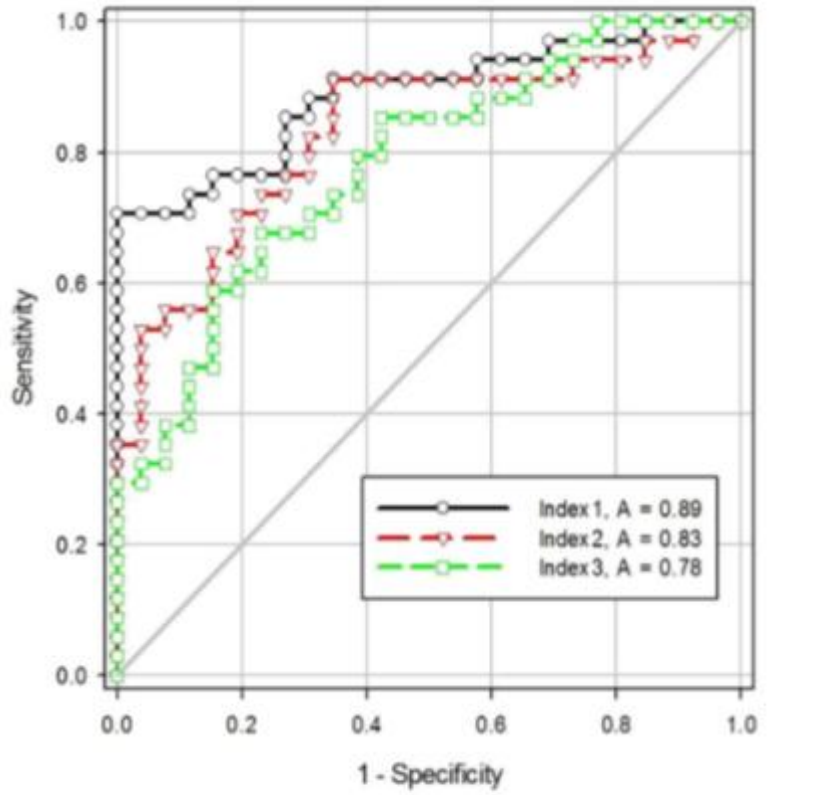
討　論

本院是否自行發展兒童壓瘡風險評估量表？

- 同意者：5人
- 保留者：10人
- 不同意：5人

補充資料

ROC曲線 (Receiver operating characteristic curve)



圖片來源：以斯帖統計顧問公司 (繪圖軟體: SigmaPlot)

- ROC曲線判別時，會以**對角線**為一個參考線，若是檢驗工具的ROC曲線剛好落在對角的參考線上，則表示檢驗工具對於此疾病的診斷沒有鑑別性。
- 若是**ROC曲線愈往左上方移動，表示檢驗工具對於疾病的敏感度愈高**，且偽陽性率愈低，亦即此工具的鑑別力較佳，而最靠近左上角的的點 (0,1) 是錯誤歸類最少的切點，其敏感度(sensitivity)是最大的且偽陰性率(1 – specificity)是最小的。

ROC Curve

一般在判別檢驗工具的好壞時，除了看曲線的圖形之外，也可以利用曲線下的面積(Area Under Curv; AUC)來判別ROC曲線的鑑別力，AUC數值的範圍從0到1，數值愈大愈好。以下為AUC數值一般的判別規則：

- AUC=0.5 (no discrimination 無鑑別力)
- 0.7≦AUC≦0.8 (acceptable discrimination 可接受的鑑別力)
- 0.8≦AUC≦0.9 (excellent discrimination 優良的鑑別力)
- 0.9≦AUC≦1.0 (outstanding discrimination 極佳的鑑別力)


Thank you!