

REVIEW

doi: 10.5455/medarch.2018.72.141-144

MED ARCH. 2018 APR; 72(2): 141-144

RECEIVED: FEB 10, 2018 | ACCEPTED: MAR 25, 2018

¹Acute Trauma and Surgical Unit. North West of Anglia Foundation Trust, Huntingdon, United Kingdom

²Health Centre of Salamis, Salamis, Greece

³Hellenic Center for Diseases Control and Prevention, Athens, Greece

⁴Nursing Technological Institute of Central Greece, Athens, Greece

Corresponding author: Charalambos Charalambous. Acute Trauma and Surgical Unit. North West of Anglia Foundation Trust, United Kingdom. ORCID ID: <http://www.orcid.org/0000-0001-7753-0163>. E-mail: CharalambosRN@gmail.com

Evaluation of the Validity and Reliability of the Waterlow Pressure Ulcer Risk Assessment Scale

Charalambos Charalambous¹, Agoritsa Koulori², , Aristidis Vasilopoulos³, Zoe Roupa⁴

ABSTRACT

Introduction: Prevention is the ideal strategy to tackle the problem of pressure ulcers. Pressure ulcer risk assessment scales are one of the most pivotal measures applied to tackle the problem, much criticisms has been developed regarding the validity and reliability of these scales. **Objective:** To investigate the validity and reliability of the Waterlow pressure ulcer risk assessment scale. **Method:** The methodology used is a narrative literature review, the bibliography was reviewed through Cinahl, Pubmed, EBSCO, Medline and Google scholar, 26 scientific articles where identified. The articles where chosen due to their direct correlation with the objective under study and their scientific relevance. **Results:** The construct and face validity of the Waterlow appears adequate, but with regards to content validity changes in the category age and gender can be beneficial. The concurrent validity cannot be assessed. The predictive validity of the Waterlow is characterized by high specificity and low sensitivity. The inter-rater reliability has been demonstrated to be inadequate, this may be due to lack of clear definitions within the categories and differentiating level of knowledge between the users. **Conclusion:** Due to the limitations presented regarding the validity and reliability of the Waterlow pressure ulcer risk assessment scale, the scale should be used in conjunction with clinical assessment to provide optimum results.

Keywords: Waterlow, Pressure ulcer, validity, reliability, risk assessment scale.

1. INTRODUCTION

A pressure ulcer (PU) is defined as: a) Localised injury to the skin, and/or underlying tissue, usually over a bony prominence, as a result of pressure (including pressure associate with shear); b) A number of contributing factors are also associated with PU the primary of which is impaired mobility (1).

PU are identified as one of seven medical conditions that affect a large number of individuals and require expensive treatment (2). Apart from the financial problem, the development of a PU has negative psychological and physiological impact upon the individual (3). Downie et al. (4) proposed that 95% of all PU are preventable. Jalali and Rezaie (5) reported that prevention is the best strategy to tackle the PU problem. One of the most pivotal preventative measure is the risk assessment of an individual for PU development (6).

This can be achieved using a valid, reliable and clinically relevant risk assessment scale (RAS) (7). Pancorbo-Hidalgo et al. (8) proposed that the use of a RAS can easily identify

an individual at risk and apply preventative measures however, much criticism has been applied concerning the clinimetrics of those scales (9).

2. OBJECTIVE

The aim of the review was to evaluate the validity and reliability of the Waterlow pressure ulcer risk assessment scale and provide an understanding of the current bibliography.

3. METHOD

The method of the review was a narrative review. The scientific bibliography was reviewed between March 2017 and April 2017 through the databases Cinahl, Pubmed, EBSCO, Medline and Google scholar.

The key words used were Waterlow AND pressure ulcer AND validity AND reliability AND risk assessment scale, 26 scientific articles where identified. The inclusion criteria for the articles where the existence of full text the direct correlation with the objective under study and their scientific credibility (peer review articles).

© 2018 Charalambos Charalambous, Agoritsa Koulori, Aristidis Vasilopoulos, Zoe Roupa

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

4. RESULTS

4.1. VALIDITY OF THE WATERLOW SCALE

Validity refers to the degree that a scale measures what it claims to measure (Haesler 2014). In the case of PU RAS, it would be how well it predicts the danger of an individual to develop a PU (10). Validity is not an one time process on the contrary, to establish that a tool is valid some basic aspects must be fulfilled in depth of time (11).

Face Validity

Face validity refers to if a scale appears to measure what is intended to measure (12). Waterlow gives the impression that fulfils face validity, initially due to the items included that relate to PU development (9). Secondly, it utilises a scoring system to produce a sum, which is enabling the user to allocate the individual in one of the risk categories (13). Face validity due to its subjectivity is not extensively reported still, it remains a useful indicator as the users are more likely to complete a tool with high face validity (12).

Content Validity

Content Validity is related to the degree that the items included in a scale are suitable to measure the outcome under investigation on the target population, without sacrificing the ease of use (14). It is estimated that over 200 PU risk factors have been identified, although not all of them are suitable to be included in a RAS (15).

Factors are most commonly picked through extensive bibliographical reviews or/and expert panels consensus (12). Waterlow (13) based her factor selection on personal knowledge, literature review which was very limited at the time and at the pre-existence Norton RAS (16). The Waterlow scale was revised in 2005 in the light of new evidence and a more comprehensive explanation were provided concerning the risk factor selection. Nevertheless, even on the reviewed version the risk factors selected are not supported by sufficient evidence, leaving a sense of doubt upon the factors included (17).

The Waterlow (17) revised scale included the following factors: built/weight for height BMI, assessment of the skin, gender, age, continence, mobility, nutrition, medication, tissue malnutrition and neurological deficits and major surgery or trauma. The inclusion of such a wide variety of risk factors is enabling the scale to cover a wide variety of patients although it encloses the danger to over predict the risk of an individual and making its use complex (18). Coleman et al. (19) tried to identify the risk factors that are suitable to predict independently PU development in a systematic review. The result indicated that the factors emerging most frequently as independent predictors were mobility, perfusion and skin status. Other factors such as skin moisture, age, haematological measures, nutrition, and sensory perception were found to be regularly reported but not as frequent. Medication and gender were reported as not significant factors. Medication as an independent factor of PU development is generally not supported and is unlikely that any medication will cause a PU, rather is more likely to be a surrogate indicator (19). Gender was reported as not significant by Coleman et al. (19), on the contrary, Waterlow (17) supports that physiological

differences between genders are affecting the tolerability towards PU damage. Maklebust and Magnan (20) identified no statistical relationship between gender and PU development, the authors concluded that if factors that are known to decline by age are included in a RAS then the use of gender as an independent factor is unnecessary. Overall the risk factors included in Waterlow RAS appear to relate with PU development and the content validity of the tool seem to be at an acceptable level. Adjustments in the categories gender and medication can have a beneficial effect on the content validity of the tool.

Construct Validity

Construct validity involves the demonstration of the relationship between the concepts under study and the relating theory (21). Construct validity is related with content validity, as the items included in a scale are describing the outcome under measurement (22). The theory behind Waterlow RAS is that by including factors related to PU development and scoring them, a sum will be provided indicating the risk of an individual to develop PU (Waterlow 2005). To establish construct validity, Kottner and Dassen (23) measured the convergent between Waterlow, Braden and a Visual Analogue Scale. Correlation coefficient showed that Waterlow was measuring something similar with the other scales ($r=0.51$ VAS, $r=-0.71$). By showing a degree of convergence between the scales, it indicates that the Waterlow RAS is identifying the domains of risk effectively supporting the construct validity of the scale (12). Tannen et al. (24) tested the correlation of Waterlow RAS with tools measuring different outcomes to establish construct validity, it found statistically significant correlation between Waterlow RAS and the care dependency scale ($r=-0.061$, $p<0.001$). The authors support that the construct validity of Waterlow as a PURAS is justified, because PU risk increases as the care dependency increases. Although, the association between the two scales can lead to the assumption that Waterlow is measuring general care needs and not PU risk development (24). Overall the construct validity of the Waterlow RAS needs further investigation and assessment, as there has not been a large enough volume of studies produced (23).

Criterion Validity

Criterion validity is assessing the validity of a tool by comparing and observing its performance against the existing gold standard (25). In reference to Waterlow, no tool has been identified as the gold standard, thus the measurement of criterion validity is undergone by comparing the scale with other existent relevant scales (26). The degree of agreement between the scales is referred as concurrent validity (27). Gould et al. (26) examined the concurrent validity of the Norton, Waterlow and Braden RAS versus the consensus views of an expert's panel. Using patient scenarios 236 nurses produced in total 941 assessments, Waterlow RAS had the highest degree of agreement (20%), although according to the authors it is still in low levels. The lack of a clear gold standard PU RAS, complicates the proof of criterion/ concurrent validity, as the validity of other tools is questionable (14).

Predictive validity

Predictive validity can be described as the consistency of a scale in predicting how an individual will perform in the future, based on measurements of today (12). Therefore, if a patient is assessed with a PU RAS today and found to be at high risk then it is expected that they will develop a PU in the future (28). Predictive validity is measured by specificity or true negative, this refers to the proportions of the patients that were identified as low risk and did not develop PU and sensitivity or true positive, which refers to the percentage of the patients that were identified as high risk and developed PU (9).

In a systematic review by Pancorbo-Hidalgo et al. (8), seven studies were analyzed concerning the Waterlow predictive validity. It was observed good predictive validity (OR=2.05, CI 95%=1.11-3.76) with high sensitivity (82, 4%) but low specificity (27.4%). The results were consistent with other studies supporting the high specificity, low sensitivity status of Waterlow RAS (29, 30). Transferring the results to clinical practice, patients assessed with the Waterlow RAS have increased possibility to be identified as high risk of developing PU and receiving unnecessary preventative equipment, resulting in a higher financial burden on healthcare settings (8).

Predictive validity has been criticized for its applicability in PU RAS (31). This is due to the fact that it is unethical to hold available preventative measures from patients, so that the predictive validity measurement would be possible (14). In reality, when an individual is identified as high risk they receive preventative measures, which in turn will show Waterlow RAS to poorly perform concerning specificity (32). The inadequate predictive validity of the Waterlow RAS proposes that the scale should not be used on its own for the prediction of PU development, but in conjunction with clinical judgement. Together, they can provide valuable guidance in the allocation of preventative measures (6, 14).

4.2. RELIABILITY OF THE WATERLOW RISK ASSESSMENT SCALE

To be practical in everyday clinical practice, Waterlow must have the ability to produce consistent results (33). There are two major reliability aspects: the inter-rater reliability, which is the degree to which two raters operating independently assign the same ratings (agreement) for an attribute being measured (34), and the intra-rater reliability, which refers to the ability of a scale to produce the same ratings in different points of time when used by the same rater. This allows to the assumption of stability in the individual's condition, which can be difficult to achieve as the patients at risk for PU development condition can change quite quickly (35, 14).

Inter ratter reliability

According to a systematic review by Kottner et al. (23) examining the intra and inter-rater reliability of Waterlow, no studies examining the intra-ratter reliability of the scale were identified. The lack of intra-ratter reliability studies might be explained by the fact that in clinical practise it is highly unlikely the same patient to be assessed by the same health care professional each time their condition needs to be evaluated, thus much of the focus has been given on the inter-rater reliability (14).

Edwards (36) in a cross sectional observational study assessed the intra-rater reliability of the Waterlow in a sample of 40 elderly patients in the community, with or without PU. A second assessor with the same knowledge level as the researcher concerning PU risk assessment knowledge was chosen, which reduced the possibility of bias due to knowledge level difference between the raters (9). Low levels of agreement between the raters were observed (25%) (33). Categories where disagreement was observed the most were skin type, BMI and mobility, the authors concluded that the subjectivity due to the need of the assessor's opinion in the scoring might lead to high disagreement (9). Low levels of inter-rater reliability were found also by Watkinson (37) in his comparative study between Waterlow, Norton and Braden. The agreement between the raters concerning Waterlow reached 55.6%. The raters chosen for the study were nine registered nurses, one enrolled nurse and two student nurses. The sample consistent of nine patient admitted in an acute hospital (37). Watkinson (37) states that the low agreement percentage is a result of differing knowledge level concerning PU risk assessment. Cook et al. (38) assessed the inter-rater reliability of an adapted Waterlow, minor changes were made to the categories continence, neurological deficit and medication; 15 patients were assessed by two different nurses daily for a period of seven days. Inter-rater reliability was assessed by percentage and correlation. The results proposed week to moderate level of reliability between the raters (55, 5%, $r=0.36$). Similarly to the previous study by Watkinson et al. (37), differing level of knowledge was reported as having an impact on the results. Inter-rater reliability of the Waterlow seems to be inadequate (14). Although the difference in the pathophysiology of the patients makes it difficult to identify if low reliability is caused by different perception of the patients state by the assessor, or differentiation in the interpretation of the tool (33). Additionally, the lack of clear definition (BMI, skin status) within some categories might lead to misunderstanding and wrong scoring (39, 36) and to the differentiating knowledge level amongst the health care professional using the tool (37, 38). Waterlow (17) states that high reliability is achievable through education and continuous use of the scale.

4.3. EASE OF USE

A RAS can be reliable and valid in the highest degree, although without high levels of completeness and usability will probably never be accepted by the users and thus applied in the clinical settings (14). Waterlow (17) ensure that the tool is user friendly by including explanations and cards for different factors and aspects (18). Although there is no direct study investigating the ease of use, some authors report that the scale is user-friendly and unambiguous (37). On the contrary, Banks and Bale (40) reported that community nurses found the scale time consuming. Community settings have differences from hospitals and difficulties completing direct skin observation and limited access on patient information might affect its use (41). A clear conclusion cannot be extracted for the ease of use as the volume of evidence is

limited and further investigation is needed focusing only on this aspect.

5. CONCLUSION

Based on the evidence the construct and face validity of Waterlow is acceptable, but with regards to content validity changes in the category age and gender can be beneficial. Concurrent validity of the RAS is not clear as the scale seems to measure something similar with other PU RAS, although the relevance with a scale that measures general life needs, raises doubts and requires further investigation. The predictive validity of the Waterlow is characterized by high specificity and low sensitivity. The limitation in the measurement of predictive validity caused by the application of preventative measures is not allowing a clear conclusion to be extracted and is proposing that Waterlow should be used with clinical judgement. Additionally the inter-rater reliability of Waterlow has been found to be inadequate. This may be due to lack of clear definitions within the categories and differentiating level of knowledge between the users. Waterlow is relatively easy to use, although in community settings there is an indication that the tool has limitations due to the limited information access. Waterlow is not the perfect RAS concerning the validity and reliability levels, but it can be helpful if it's used in conjunction with clinical judgement, serving as an indicator of PU risk rather than a diagnostic tool.

- **Authors contribution:** All the authors have contributed in the design, revision and final version to be published equally and they are qualified to be reported as authors.
- **Conflict of interest:** none declared.

REFERENCES

1. NPUAP. Prevention and treatment of pressure ulcers: clinical practice guideline. 2014.
2. Vanderwee K, Clark M, Dealey C, Gunningberg L, Defloor T. Pressure ulcer prevalence in Europe: a pilot study. *Journal of evaluation in clinical practice*. 2007 Apr 1; 13(2): 227-35.
3. Gorecki C, Closs SJ, Nixon J, Briggs M. Patient-reported pressure ulcer pain: a mixed-methods systematic review. *Journal of pain and symptom management*. 2011 Sep 1; 42(3): 443-59.
4. Downie F, Guy H, Gilroy P, Royyal D, Davies S. Are 95% of hospital-acquired pressure ulcers avoidable? *Wounds UK*. 2013 Sep 1; 9(3).
5. Jalali R, Rezaie M. Predicting pressure ulcer risk: comparing the predictive validity of 4 scales. *Advances in skin & wound care*. 2005 Mar 1; 18(2): 92-7.
6. Smith LN, Booth N, Douglas D, Robertson WR, Walker A, Durie M, Eraser A, Hillan EH, Swaffield J. A critique of 'at risk' pressure sore assessment tools. *Journal of Clinical Nursing*. 1995 May 1; 4(3): 153-9.
7. Ayello EA, Braden B. How and why to do pressure ulcer risk assessment. *Advances in skin & wound care*. 2002 May 1; 15(3): 125-31.
8. Pancorbo-Hidalgo PL, Garcia-Fernandez FP, Lopez-Medina IM, Alvarez-Nieto C. Risk assessment scales for pressure ulcer prevention: a systematic review. *Journal of advanced nursing*. 2006 Apr 1; 54(1): 94-110.
9. Thompson D. An evaluation of the Waterlow pressure ulcer risk-assessment tool. *British Journal of Nursing*. 2005 Apr 28; 14(8): 455-9.
10. Haesler E. National Pressure Ulcer Advisory Panel, European Pressure Ulcer Advisory Panel and Pan Pacific Pressure Injury Alliance. Prevention and treatment of pressure ulcers: quick reference guide.[Internet]. 2014.
11. Edwards M. The levels of reliability and validity of the Waterlow pressure sore risk calculator. *Journal of wound care*. 1995 Sep 2; 4(8): 373-8.
12. Merkouris A. Methodology of Nursing Research. Athens: Ellin. 2008: 263.
13. Waterlow J. Pressure sores: a risk assessment card. *Nursing times*. 1985; 81(48): 49-55.
14. Walsh B, Dempsey L. Investigating the reliability and validity of the Waterlow risk assessment scale: a literature review. *Clinical nursing research*. 2011 May; 20(2): 197-208.
15. Byrne DW, Salzberg CA. Major risk factors for pressure ulcers in the spinal cord disabled: a literature review. *Spinal cord*. 1996 May; 34(5): 255.
16. Norton D, McLaren R, Exton-Smith AN. An investigation of geriatric nursing problems in hospital. Churchill Livingstone; 1962.
17. Waterlow J. Pressure ulcer prevention manual. 2005.
18. Bell J. Are pressure ulcer grading and risk assessment tools useful? *Wounds*. 2005; 1(2): 62.
19. Coleman S, Gorecki C, Nelson EA, Closs SJ, Defloor T, Halfens R, Farrin A, Brown J, Schoonhoven L, Nixon J. Patient risk factors for pressure ulcer development: systematic review. *International journal of nursing studies*. 2013 Jul 1; 50(7): 974-1003.
20. Maklebus J, Magnan MA. Risk factors associated with having a pressure ulcer: a secondary data analysis. *Advances in wound care: the journal for prevention and healing*. 1994; 7(6): 25-8; 31.
21. Roberts P, Priest H. Reliability and validity in research. *Nursing standard*. 2006; 20(44): 41-5.
22. Thomas DR. Existing tools: are they meeting the challenges of pressure ulcer healing? *Advances in wound care: the journal for prevention and healing*. 1997 Sep; 10(5): 86-90.
23. Kottner J, Dassen T. Pressure ulcer risk assessment in critical care: interrater reliability and validity studies of the Braden and Waterlow scales and subjective ratings in two intensive care units. *International journal of nursing studies*. 2010 Jun 1; 47(6): 671-7.
24. Tannen A, Balzer K, Kottner J, Dassen T, Halfens R, Mertens E. Diagnostic accuracy of two pressure ulcer risk scales and a generic nursing assessment tool. A psychometric comparison. *Journal of Clinical nursing*. 2010 Jun 1; 19(11-12): 1510-8.
25. Σαχίνη-Καρδάση Α. Μεθοδολογία Έρευνας: Εφαρμογές στο Χώρο της Υγείας. Γ' έκδ., Εκδόσεις ΒΗΤΑ, Αθήνα. 1997: 155-61 (in Greek).
26. Gould D, Goldstone L, Kelly D, Gammon J. Examining the validity of pressure ulcer risk assessment scales: a replication study. *International Journal of Nursing Studies*. 2004 Mar 1; 41(3): 331-9.
27. Ουζούνη Χ, Νακάκης Κ. Ή αξιοπιστία και η Εγκυρότητα των εργαλείων Μέτρησης σε Ποσοτικές μελέτες. *Νοσηλευτική*. 2011; 50(2): 231-9 (in Greek).
28. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*. 2008 Dec 1; 65(23): 2276-84.
29. Pang SM, Wong TK. Predicting pressure sore risk with the Norton, Braden, and Waterlow scales in a Hong Kong rehabilitation hospital. *Nursing Research*. 1998 May 1; 47(3): 147-53.
30. Balzer K, Pohl C, Dassen T, Halfens R. The Norton, Waterlow, Braden, and Care Dependency Scales. Comparing Their Validity When Identifying Patients' Pressure Sore Risk. *Journal of Wound Ostomy & Continence Nursing*. 2007 Jul 1; 34(4): 389-98.
31. Papanikolaou P, Clark M, Lyne PA. Improving the accuracy of pressure ulcer risk calculators: some preliminary evidence. *International journal of nursing studies*. 2002 Feb 1; 39(2): 187-94.
32. Anthony D, Parboteeah S, Saleh M, Papanikolaou P. Norton, Waterlow and Braden scores: a review of the literature and a comparison between the scores and clinical judgement. *Journal of clinical nursing*. 2008 Mar 1; 17(5): 646-53.
33. Kelly J. Inter-rater reliability and Waterlow's pressure ulcer risk assessment tool. *Nursing Standard (through 2013)*. 2005 Apr 20; 19(32): 86.
34. Polit DF, Beck CT. *Nursing research: Principles and methods*. Lippincott Williams & Wilkins, 2004.
35. Kottner J, Dassen T, Tannen A. Inter- and intrarater reliability of the Waterlow pressure sore risk scale: a systematic review. *International Journal of Nursing Studies*. 2009 Mar 1; 46(3): 369-79.
36. Edwards M, Feightner J, Goldsmith CH. Inter-rater reliability of assessments administered by individuals with and without a background in health care. *The Occupational Therapy Journal of Research*. 1995 Apr; 15(2): 103-10.
37. Watkinson C. Inter-rater reliability of risk-assessment scales. *Professional nurse (London, England)*. 1996 Aug; 11(11): 751-2.
38. Cook M, Hale C, Watson B. Interrater reliability and the assessment of pressure-sore risk using an adapted Waterlow Scale. *Clinical Effectiveness in Nursing*. 1999 Jun 1; 3(2): 66-74.
39. Bridel, J. Pressure sore risk in operating theatres. *Nursing Standard*. 1993; 7(32): 4-10.
40. Banks V, Bale S. Improved assessment: is it the key to pressure sore prevention in the community. In *Proceedings of the Second European Conference on Advances in Wound Management*. Macmillan, London 1992: 163-5.
41. Maylor M, Roberts A. A comparison of three risk assessment scales. *Professional nurse, (London, England)*. 1999 Jun; 14(9): 629-32.