VALIDITY OF PRESSURE ULCER RISK ASSESSMENT SCALES: REVIEW

Lenka Šáteková¹, Katarína Žiaková²

¹Department of Nursing and Midwifery, Medical Faculty, University of Ostrava
²Department of Nursing, Jessenius Medical Faculty in Martin, Comenius University in Bratislava, Slovakia

Abstract

Aim: The aim of the review was to determine the predictive validity of selected pressure ulcer risk assessment scales. Prediction of a risk of pressure ulcers is a priority issue in nursing. In the foreign literature, there are currently lots of research studies dealing with the validity of pressure ulcer risk assessment scales. Methods: The data sources were articles in three licenced, free and electronic databases (EBSCO, CINAHL and PubMed). The data were retrieved for the period 2003-2013. The levels of evidence were evaluated according to Haynes’s pyramid of information sources, distinguishing seven categories of studies. Included in the review were studies with levels of evidence of 1-4. The inclusion criteria were met by fifteen studies on the validity of the Braden, Norton, Waterlow Scale, Song and Choi, Cubbin and Jackson, Modified Norton, EVARUCI, Suriadi and Sanada and Modified Braden scales. The most frequently tested scales included the Braden, Waterlow and Norton scales. The Braden Scale showed optimal predictive validity. There is a need for further tests of the Waterlow and Norton scales. Testing of new pressure ulcer risk assessment scales such as the Suriadi and Sanada or EVARUCI scales is underway. The studies in this review showed considerable variations caused by differences in ages and numbers of subjects, settings, cut-off points (i.e. limits defining the risk or its size – low, medium or high), length of study and preventive measures used. Conclusion: There is a need for examining the predictive validity of pressure ulcer risk assessment scales in our clinical setting and comparing the results with foreign studies. Before examining the predictive validity of pressure ulcer risk assessment scales, their proper and consistent translation is needed, according to recognized methodology. After high-quality translations are made, validity tests may be started, contributing to reduction of scientific ambiguity of pressure ulcer risk assessment scales.

Key words: risk assessment, pressure ulcer, scale, nurse, validity.

Introduction

Pressure ulcers represent a nursing and social problem having a high priority in clinical practice. They cause suffering and pain to patients, evoke feelings of failure in nurses, prolong hospital stay and significantly increase additional costs of care (Agrawal, Chauhan, 2012, p. 246; Dealey, Posnett, Walker, 2012, p. 262). The prevalence rates of pressure ulcers are 10-17% in acute care, 0-29% in home care and 2.3-23.9% in long-term care facilities (Ayello, 2007). In the USA, Canada and some parts of Europe, the prevalence of pressure ulcers ranges from 14% to 25% (Stotts, Gunningber, 2007, p. 43). According to Bóriková (2006, p. 20), the prevention of pressure ulcers is the oldest preventive method in the nurses’ work ever since F. Nightingale. Predicting the risk of pressure ulcers is one of the priorities in nursing. Targeted and high-quality preventive approach is cheaper than therapy alone (Lyder, 2003, p. 223-226). Organizations such as the National Pressure Ulcer Advisory Panel (NPUAP) and European Pressure Ulcer Advisory Panel (EPUAP) indicate the prevention of pressure ulcers as the best and effective method. Prevention is divided into five key areas: risk assessment, skin assessment, nutrition, repositioning and use of support surfaces (NPUAP, 2009, p. 4). The first area of prevention, as recommended by the NPUAP and EPUAP, is to assess the risk of pressure ulcers. For predicting pressure ulcers, various scales have been developed, allowing consistent mapping of the occurrence of risk factors, that is, indicators supporting pressure ulcers (Bóriková, 2006, p. 20). In Czech health care facilities, the most common tool currently used is the Norton Scale (Mikula, Müller, 2008, p. 18). In other countries, these are the Braden Scale, also Waterlow Scale in the UK or Modified Norton Scale in Sweden (Lindgren et al., 2002, p. 191). The NPUAP

© 2014 Central European Journal of Nursing and Midwifery
emphasizes that ideal assessment scales should be easy to use with acceptable predictive validity, that is, high sensitivity and specificity (Defloor, Grypdonck, 2005, p. 374). At present, many scales are used in clinical practice but most of them have not been properly validated yet (Thomas, 2001, p. 328). Several scales are just modified originals, their reliability and validity have not been tested and they have poor methodological quality and low number of respondents (Pressure Ulcer Risk Assessment and Prevention, 2001, p. 14). Ideal validity is expressed as sensitivity and specificity. These values are widely accepted as means for evaluating the validity of scales. An ideal scenario would be 100% sensitivity and 100% specificity of a scale. No agreed gold standard has been established (Defloor, Grypdonck, 2005, p. 374). So far, a scale with 100% sensitivity and specificity has not been developed (Bóriková, 2006, p. 24). Research studies dealing with the predictive validity of scales for assessing the risk of pressure ulcers also show positive and negative predictive values and values of the area under the ROC curve as diagnostic parameters to evaluate the validity. Research efforts are targeted at the predictive validity of scales for assessing the risk of the development of pressure ulcers, development of new risk factors and their clinical applicability (Bóriková, 2006, p. 20). Worldwide, testing of the Norton, Braden and Waterlow scales is currently underway. In the Czech Republic, only the Braden scale has been validated so far (Mandysová, Pechová and Ehler, 2013, p. 609). Validity is a correlation between the measurement tool and the external criterion, or the degree of positivity of the answer to the question “Do we actually measure what we are supposed to measure?” Validity is the basic psychometric indicator of a measuring instrument (Gurková, 2009a, p. 171). A measuring instrument is considered valid if it accurately measures the phenomena for which it has been designed. The process leading to verification of validity is referred to as validation (Reiterová, 2003, p. 89). Validity is most often expressed by sensitivity, specificity, positive and negative predictive values.

Sensitivity is an ability of the instrument to give a positive result if the risk actually exists. It expresses the percentage of patients who were predicted to have pressure ulcers and also developed them. For example, 80% sensitivity of the instrument finds out the risk in 80 patients out of 100. Specificity is expressed as the percentage of patients whose pressure ulcers neither developed nor were predicted (Bóriková, Žiaková, Gurková, 2009b, p. 220). The positive predictive value reflects the probability that the object is indeed positive when the test was positive. The negative predictive value reflects the probability that the object is actually negative when the result is negative (Bortlíček, 2008, p. 8). The area under the ROC curve determines the predictive usefulness of an instrument. It is the main criterion of the performance of the test. It assumes values from 0.5 to 1.0. Values below 0.5 are inadequate and the test is no better than a coin toss. The bigger the area under the ROC curve, the more effective the test (Balla et al., 2004, p. 88).

**Aim**

The aim of the review was to determine the validity rates of selected pressure ulcer risk assessment scales using sensitivity, specificity, positive and negative predictive value and the area under the ROC curve in foreign research studies.

**Methods**

**Research question**

What is the degree of validity of selected pressure ulcer risk assessment scales as indicated by foreign research studies?

**Eligibility criteria**

The data were retrieved for the period 2003-2013. Research studies in English, Spanish, Slovak and Czech were selected. The inclusion criteria were as follows: a research study examined the validity of a selected pressure ulcer risk assessment scale or the validity was examined as a part of another research project; a research study in adult population, in a health care facility; study types: systematic review, meta-analysis, randomized controlled trial, cohort study, case-control study; and study languages: English, Spanish, Slovak or Czech.

**Sources**

The data come from three licensed, free and electronic databases (EBSCO, CINAHL and PubMed). The following key words were used: risk assessment, pressure ulcer, scale, nurse, validity.

**Study selection**

Included in the review were studies with levels of evidence of 1, 2, 3 or 4 according to Haynes’s pyramid of information sources: Level 1 – evidence from systematic reviews and meta-analyses; Level 2 – evidence from randomized controlled trials; Level 3 – evidence from cohort studies; and Level 4 – evidence from case control studies (Běhounek et al., 2011, p. 12).
The exclusion criteria were as follows: a research study did not mention data related to either specificity or sensitivity; a research study in the paediatric population; a language other than those mentioned above; research studies below Level 4 in Haynes’s pyramid of information sources. The Boolean operators AND and OR were used in searching. Out of 221 documents found, 15 studies matched the intention of the work and were included in the review article.

**Data analysis**
The following data were analysed in research studies: cut-off score, incidence of pressure ulcers, sensitivity, specificity, positive predictive value, negative predictive value, and the area under the ROC curve.

**Results**
In the foreign literature, there are currently lots of research studies dealing with the predictive validity of pressure ulcer risk assessment scales. Based on the inclusion criteria, fifteen research studies dealing with the validity were found. These studies were analyzed and subsequently data on the validity of the scales were identified (see Table 1). The most frequently tested scale was the Braden Scale (10 studies), followed by the Waterlow Scale (4 studies) and Norton Scale (3 studies).

<table>
<thead>
<tr>
<th>Table 1 Validity of pressure ulcer risk assessment scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Braden Scale</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Norton Scale</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Song and Choi Scale</td>
</tr>
<tr>
<td>Cubbin and Jackson</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Waterlow</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Modified Norton Scale</td>
</tr>
<tr>
<td>EVARUCI Scale</td>
</tr>
<tr>
<td>Suriadi and Sanada Scale</td>
</tr>
<tr>
<td>Modified Braden Scale</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

ND – no data, PPV – positive predictive value, NPV – negative predictive value, AUC – area under the ROC curve

Other scales were tested in small numbers of studies (in descending order): the Modified Braden, Cubbin and Jackson, Song and Choi, Modified Norton, EVARUCI, Suriadi and Sanada scales. Scales were
most commonly tested in intensive care units (ICUs; 7 studies), followed by hospitals (4 studies), long-
term care facilities (2 studies), orthopaedic wards (1
study) and internal medicine wards (1 study). The
sample sizes varied across the studies, from 23
respondents (Costa, Caliri, 2011) to 1172 subjects
(Defloor, Grypdonck, 2005). Follow-up period also
varied, from one month (Defloor, Grypdonck, 2005)
to thirteen months (Chan et al., 2009). The mean age
of respondents in the studies ranged from 55.2 years
(Suriadi et al., 2008) to 84.6 years (Defloor,
Grypdonck, 2005). In three studies, the mean age of
the respondents was not stated. The incidence of
pressure ulcers ranged from 2.1% (Kwong et al.,
2005) to 28.4% (Suriadi et al., 2008). Various studies
reported different cut-off points (i.e. limits defining
the risk or its size – low, medium or high), despite the
fact that the same scale was used and the studies were
conducted in the same units. For the Braden Scale,
authors used cut-off points of 14 and 16 in ICUs and
17 and 18 in long-term care facilities. For the Norton
Scale, the largest range of validity was at a cut-off
point of 14. The Waterlow Scale showed the biggest
value of validity at a cut-off point of 17. Three
studies did not indicate at what cut-off point the
validity was assessed (see Table 2).

Table 2 Design and study characteristics

<table>
<thead>
<tr>
<th>Author</th>
<th>Scale and cut-off score</th>
<th>Setting</th>
<th>N</th>
<th>Mean age (years)</th>
<th>Follow-up period</th>
<th>Minimum stage</th>
<th>PU prevention</th>
<th>Level of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defloor, Grypdonck (2005)</td>
<td>Braden 17 Norton 12</td>
<td>Long-term care facilities for older people</td>
<td>1172</td>
<td>M = 84.6 W = 79.2</td>
<td>4 weeks</td>
<td>I</td>
<td>Small cell alternating and water mattresses, sheepskins, gel cushions.</td>
<td>2</td>
</tr>
<tr>
<td>Kwong et al. (2005)</td>
<td>Modified Braden 16 Norton 14</td>
<td>Hospitals</td>
<td>429</td>
<td>54.07</td>
<td>ND</td>
<td>I</td>
<td>Turning (every 2 hours), use of material to reduce pressure (cushion, small pillow, air ring, etc.), keeping bed linen clean, dry and smooth, keeping skin clean and dry, positioning, use of a draw sheet for lifting patients, massage of PU.</td>
<td>3</td>
</tr>
<tr>
<td>Feuchtinger et al. (2007)</td>
<td>Waterlow - ND Norton - ND Braden - ND</td>
<td>Hospitals</td>
<td>754</td>
<td>ND</td>
<td>ND</td>
<td>I</td>
<td>ND</td>
<td>4</td>
</tr>
<tr>
<td>Suriadi et al. (2008)</td>
<td>Suriadi and Sanada 4</td>
<td>ICU</td>
<td>253</td>
<td>55.2</td>
<td>9 months</td>
<td>I</td>
<td>ND</td>
<td>3</td>
</tr>
<tr>
<td>Chan et al. (2009)</td>
<td>Modified Norton 19 Braden 16</td>
<td>Orthopaedic wards</td>
<td>197</td>
<td>79.4</td>
<td>13 months</td>
<td>I</td>
<td>ND</td>
<td>3</td>
</tr>
<tr>
<td>Kim, Lee, Lee, Eom (2009)</td>
<td>Braden 14 Song, Choi 21</td>
<td>Surgery ICU</td>
<td>219</td>
<td>58.1</td>
<td>4 months</td>
<td>I</td>
<td>Turning (every 2 hours), keeping skin clean and</td>
<td>3</td>
</tr>
</tbody>
</table>
Braden Scale

Based on the inclusion criteria, 10 studies dealing with the validity of the Braden Scale were found in various clinical settings. The validity indicators were good: sensitivity 95% (Costa, Caliri, 2011) and 67% (Chan et al., 2009), specificity 83.1% (Serpa et al., 2011) and 29% (Feuchtinger et al., 2007), positive predictive value 70% (Feuchtinger et al., 2007) and 5% (Kwong et al., 2005) and negative predictive value between 100% (Kwong et al., 2005) and 38% (Feuchtinger et al., 2007). The area under the ROC curve ranged from a very low value of 0.155 to a very good value of 0.881.

Waterlow Scale

Table 1 gives data from four studies investigating the validity of the Waterlow Scale. The validity indicators for this scale were as follows: sensitivity between 86% (Balzer et al., 2007) and 71.4% (Serpa et al., 2009), specificity between 79% (Webster et al., 2010) and 4.2% (Tannen et al., 2010), positive predictive value 14.3% (Serpa et al., 2009) and 13% (Webster et al., 2010) and negative predictive value 98% (Webster et al., 2010) and 96.8% (Serpa et al., 2009). The area under the ROC curve ranged from usable (0.640) to very good (0.810).

Norton Scale

Table 1 shows data from three studies examining the validity of the Norton Scale. The validity indicators for the Norton Scale were: sensitivity between 89% (Kwong et al., 2005) and 62.3% (Defloor, Grypdonck, 2005) and specificity between 75% (Balzer et al., 2007) and 61% (Kwong et al., 2005). Details on positive and negative predictive values were not mentioned in two out of the three studies. Also, the area under the ROC curve was stated in only one study, with a good value of 0.752.

Other scales

Validity data of six other pressure ulcer risk assessment scales were also included in the review article. Some were modified versions of the originals (Modified Norton, Modified Braden), others were originally developed (Cubbin and Jackson, Suriadi and Sanada, EVARUCI, Song and Choi). The EVARUCI scale showed the best area under the ROC curve (0.938). As there are few research studies dealing with the validity of these scales, further research is needed.

Discussion

Recently, several pressure ulcer risk assessment scales have been developed and used around the
The most common tested pressure ulcer risk assessment scales abroad are the Braden and Norton scales in both the original and modified versions (Bóriková, 2006, p. 22). Studies dealing with validity of the Braden, Norton, Waterlow, Song and Choi, Cubbin and Jackson, EVARUCI, Suriadi and Sanada, Modified Norton and Modified Braden scales were identified by this review. Pressure ulcer risk assessment scales are used in clinical practice. Inconsistent and often controversial research design is a major drawback of studies. This article is mainly concerned with the Braden, Waterlow and Norton scales. According to the results of research studies included in the present review, the Braden Scale exhibited optimal validity in different clinical settings (long-term care facilities, ICUs and hospitals). This was confirmed by findings in a survey by Pancorbo-Hidalgo et al. (2006, p. 107). The Braden Scale also showed (except for one study) area under the ROC curve values over 0.600, that is, sufficient to claim that its predictive utility is good. Therefore, the Braden Scale showed optimal validity with a good prediction of the risk, which is consistent with findings by Bóriková (2006, p. 24). The Waterlow Scale showed good sensitivity but low specificity. It means that in many patients, preventive measures are unnecessarily initiated, resulting in a waste of both material and human resources. The results show that the Waterlow Scale had a lower predictive value than the Braden Scale. It is necessary to provide definitive evidence of the validity of the Waterlow Scale by randomized controlled trials with other pressure ulcer risk assessment scales and clinical judgment. This fact is confirmed by arguments of Webster et al. (2010, p. 14). None of four studies validating the Waterlow Scale was higher than Level 3 of evidence according to Haynes’s pyramid of information sources. Finally, the Norton Scale was assessed. Even though this is the first pressure ulcer risk assessment scale, only few research studies have tested its validity. The sensitivity and specificity values were good. In some studies, data on positive and negative predictive values were unavailable and so were area under the ROC curve values. Therefore, it would be appropriate to conduct further studies to confirm its validity with all the indicators. Further verification of the Norton Scale was also recommended in a systematic review by Pancorbo-Hidalgo et al. (2006, p. 107). Validation studies of other pressure ulcer risk assessment scales were also included in the present review. For these scales, however, fewer studies were found (a single study for most of them). The EVARUCI Scale, specially developed for ICUs, showed good values of the validity indicators. And so did the Suriadi and Sanada Scale. However, further verification of their validity is recommended. A controversial area of the research is the methodology for examining their validity. Sensitivity and specificity are the most commonly used and recommended statistical tools to assess the predictive validity of scales. Unfortunately, the design of validity studies is rather inconsistent. There are significant differences between studies that are caused by differences in ages and numbers of subjects, settings, cut-off points, follow-up periods and indicated preventive nursing care. Defloor and Grypdonck (2005, p. 613) performed a critical analysis of methods evaluating the prediction validity of pressure ulcer risk assessment scales. They do not consider sensitivity and specificity as the most appropriate diagnostic parameters. Scales should be evaluated in combination with the preventive measures used. Diagnostic parameters of a scale may be affected by implementation of standard preventive measures. Pressure ulcer risk assessment scales do identify patients in need of preventive measures. If they are started, the probability of developing pressure ulcer at the beginning of the study and at the end will not be constant. Preventive measures change the values of sensitivity and specificity. Moreover, the values of sensitivity and specificity, according to Defloor and Grypdonck (2005, p. 616), are affected by heterogeneity of length of observation, choice of appropriate cut-off points and setting. Therefore, these items are included in the tables in this article. Several research studies did not mention what preventive measures had been implemented or only standard nursing care was mentioned (see Table 2), without the authors specifying the type of care. The inconsistency of methodology for assessing the predictive validity of scales makes comparison with other studies impossible. In recent literature, the predictive validity of scales has mostly been studied in intensive care and surgical units. There is an analysis of clinical applicability of these scales in the above wards (Lindgren et al., 2002, p. 190). In the present review, seven studies were concerned with testing the scales in ICUs (see Table 2).

**Conclusion**

In foreign literature and electronic databases, several studies on the predictive validity of scales were identified. Nursing research into validation of pressure ulcer risk assessment scales is at a high level abroad, as the review proves. The most commonly tested pressure ulcer risk assessment instruments include the Braden, Waterlow and Norton scales. The Braden Scale has optimal validity. Further research is needed to study the above scales and their validity. There is a need to examine the predictive validity of
pressure ulcer risk assessment scales in our clinical setting and compare the results with foreign studies. Then we need to determine which of these scales is most valid in our country. However, before examining the validity of scales, proper and consistent translation is needed, according to the established methodology. In Czech professional literature, for example, several versions of the Braden Scale and others are available. That is undesirable as it leads to inconsistencies in nursing diagnoses. After high-quality translations are made, validity tests may be started, contributing to reduction of scientific ambiguity of pressure ulcer risk assessment scales.

**Ethical aspects and conflict of interest**

All bibliographic sources were cited. The authors declare no conflict of interest.

**Acknowledgement**

Supported by grant KE GA No. 049-4UK/2013 E-textbook: Nursing in neurology.

**References**


