

Inter-rater Reliability of Items of the Braden Scale, the Norton Scale and Waterlow Scale**Lenka Šáteková¹, Katarína Žiaková²**¹Department of Nursing and Midwifery, Faculty of Medicine, University of Ostrava in Ostrava, Czech Republic²Department of Nursing, Jessenius Medical Faculty in Martin, Comenius University in Bratislava, Slovak Republic**ABSTRACT**

Background: Worldwide, approximately 40 pressure ulcer risk assessment scales are available. Despite of this amount, the psychometric properties were tested only for some.

Aim: To determine inter-rater reliability of items of selected pressure ulcer risk assessment scales (Braden Scale, Norton Scale and Waterlow Scale).

Methods: The data were collected from April to August 2014 in one long-term care department. The sample consisted of 32 patients. An intra-class correlation coefficient (ICC) was used to determine the inter-rater reliability.

Results: The inter-rater reliability of Braden Scale ranged between $ICC = 0.846$ for the item "activity" and $ICC = 0.645$ for the item "nutrition". The highest inter-rater reliability of Norton Scale reached for item "incontinence" ($ICC = 0.931$), the lowest for item "physical condition" ($ICC = 0.849$). The highest inter-rater reliability of Waterlow Scale is observed for items "sex" ($ICC = 1$), "surgery/trauma" ($ICC = 1$). The lowest inter-rater reliability reached item "weight loss score" ($ICC = 0.497$).

Conclusions: The highest inter-rater reliability reached items of Norton Scale, followed by items of Braden Scale. The lowest inter-rater reliability reached items of Waterlow Scale. We recommend further testing of pressure ulcer risk assessment scales in Czech clinical settings.

KEY WORDS

pressure ulcer, inter-rater reliability, pressure ulcer risk assessment scale, intra-class correlation coefficient, long-term care department

INTRODUCTION

Identification of people at risk of pressure ulcers and initiation of preventive interventions is an important method to reduce the prevalence and incidence of decubitus (1). Such identification is carried out through the implementation of scales for assessing the risk of pressure ulcers. Currently, there are almost 40 scales to assess the risk of decubitus throughout the world (2, 3). Despite their quantity, psychometric properties have been tested only in some of them (4). Use of scales for assessing the risk of decubitus in clinical practice requires them to be valid and reliable, to demonstrate an improvement of health care quality and they need to improve results of the patients (5, 6). Several clinically recommended methods, together with the authors of research studies, suggest the use of scales for assessing the risk of decubitus as a first step in effective prevention of pressure ulcer development (7,

8, 9, 10, 11, 12, 13). The Braden scale, the Norton scale and the Waterlow scale are the most frequently tested scales for assessing the risk of decubitus abroad (11, 14, 15, 16). Research studies abroad focus on determining the validity of these scales (2, 3, 17). However, the inter-rater reliability of the scale is also an important indicator of the use of psychometric measurement tool. It is defined as degree of compliance between the two assessors, who independently assigned the same score to the observed object or values that have been measured or observed (18). It is necessary that a scale for assessing the risk of decubitus has a good validity as well as high degree of inter-rater reliability. In case of a complicated inter-rater reliability, the patient may be identified as risky by one nurse, but as risk free by another one at the same time. This may result in differently planned nursing interventions. On one hand, a waste of resources may occur; on the other hand,

necessary and required care may be neglected (19). Within the framework of psychometric testing, we can evaluate the inter-rater reliability of the total score of the scale as well as its particular items. There has been the Braden scale, the Norton scale and the Waterlow scale tested for the purpose of inter-rater reliability in the Czech Republic to this day (19, 20, 21). Authors Mandysová et al. (20) state the inter-rater reliability of the total score of the Braden scale $\kappa = 0.474$. The inter-rater reliability of particular items of the Braden scale varied (from $\kappa = 0.232$ to $\kappa = 0.470$). The authors Mandysová et al. repeated the research study in 2012 on a larger number of respondents in 2013. The inter-rater reliability of the total score of the Braden scale reached $\kappa = 0.564$. The inter-rater reliability of particular items of the Braden scale varied from $\kappa = 0.091$ to $\kappa = 0.613$ (19). Authors Šáteková and Žiaková (21) state the inter-rater reliability of the total score $ICC = 0.775$ for the Braden scale, $ICC = 0.837$ for the Norton scale and $ICC = 0.914$ for the Waterlow scale. The inter-rater reliability of particular items of the Waterlow and the Norton scale has not been tested in the Czech clinical environment yet.

OBJECTIVE OF STUDY

The objective of this study was to investigate the inter-rater reliability of items of the Norton Scale, the Braden Scale and the Waterlow Scale.

SAMPLE

Assessors

There were two assessors evaluating the research group of patients. Following evaluation criteria were identified: professional education, at least one year of experience, and working position of nurse at selected department. Exclusion criteria: less than one year of experience.

A nurse with the title of certified specialist and seven years of experience was the A assessor. A nurse with the Bachelor's degree and seven years of experience was the B assessor.

Research group

The research group consisted of 32 patients hospitalized in the department of long-term nursing care. The inclusion criteria were: hospitalization in the given department at the time of research, absence of pressure ulcers, signed informed consent. Exclusion criteria: unsigned informed consent, patients with pressure ulcers. The average age of patients was 74.31 years in the range of 58–95 years.

METHODS

We selected a cross-sectional type of study. Data collection took place at one department with long-term care from April to August 2014. Both assessors passed one hour training directly at the department before the implementation of research. The training focused on the area of evaluation of the risk of pressure ulcers (items of individual tools to assess the risk of decubitus). At the beginning, the A assessor evaluated the risk of decubitus in patients using three selected scales to assess the risk of decubitus. The B assessor evaluated the risk of decubitus the same way as the A assessor within 24 hours. Both evaluations were carried out independently, which means that the assessors could not see the evaluation results of each other. The Braden scale, the Norton scale and the Waterlow scale were used to assess the risk of decubitus.

The Braden scale consists of 6 items: Sensory perception, Moisture, Activity, Mobility, Nutrition, Friction and Shear. All items are rated using numbers 1–4, except Friction and Shear. This item is evaluated using numbers 1–3. Lower score indicates a higher risk of decubitus. The Norton scale consists of the evaluation of the following fields: Physical condition, Mental condition, Activity, Mobility and Incontinence. Each of the given fields is divided into four levels according to importance. Lower score indicates a higher risk of decubitus.

The Waterlow scale consists of 8 items: BMI, Skin type/visual risk areas, Sex and age, Special risks, Continence, Mobility, Malnutrition screening tool, Neurological deficit. Higher score indicates a higher risk of decubitus.

The evaluation of inter-rater reliability of selected scales to assess risk of decubitus was elaborated using the Intraclass Correlation Coefficient method (ICC). This statistical method is recommended for evaluating the inter-rater reliability of scales to assess risk of decubitus (22). The Intraclass Correlation Coefficient (ICC) is a generic tool to measure conformity or consensus. It is used for the measurement of parameter data. It demonstrates a compliance between two or more assessors. Individual results of the ICC statistical method were evaluated as follows: $0–0.20$ = very poor compliance, $0.21–0.40$ = poor compliance, $0.41–0.60$ = fair compliance, $0.61–0.80$ = good compliance, and $0.81–1.0$ = excellent compliance (23).

Before the implementation of our research study, we have obtained the consent of the Ethics Committee from the University of Ostrava and the consent a medical facility, where the research study was conducted. We have obtained the consent from Mrs Mandysová,

the author of linguistic validation of the Braden Scale into Czech. The respondents participated on the study voluntarily and all data has been processed confidentially. The authors declare that the study has no conflict of interest.

Linguistic validation of Braden Scale, Norton Scale and Waterlow Scale

The Braden Scale had been officially translated into Czech in 2013 by authors Mandysová et al. (19) and these translation was used in our research. The Norton and Waterlow scales were officially translated using the back-translation approach. In first step: A translation of the scale from English into Czech was produced by a translator specialized in nursing. In second step: The scale was translated back into English by another translator specialized in nursing. In third step: Both translators and a small working group of experts (two doctors of nursing practice and two PhD students in nursing) compared the two English versions, analyzed differences and ambiguities, looked for errors and made corrections to the Czech version. Based on their consensus and agreement, the final Czech versions of the two scales were produced.

RESULTS

The compliance rate of the Braden Scale ranges from $ICC=0.846$ for "Activity" and $ICC=0.645$ for "Nutrition". Two of the six items of the Braden Scale reached excellent compliance and four of them achieved good compliance among the assessors (Table 1).

The compliance rate of the Norton Scale is shown in Table 2. The highest compliance was reached by "Incontinence" ($ICC = 0.931$), the lowest one was achieved by "Physical condition" ($ICC = 0.849$). All items of the Norton Scale reached excellent compliance among the assessors.

The compliance rate of the Waterlow Scale is shown in Table 3. The highest compliance can be observed in "Sex" ($ICC = 1$), "Major surgery or trauma" ($ICC = 1$). The lowest compliance was reached by "Weight loss score" ($ICC = 0.497$). Four of the twelve items of the Waterlow Scale reached excellent compliance, eight of them achieved good compliance and one reached fair compliance among the assessors.

DISCUSSION

The inter-rater reliability is one of the aspects showing the reliability of tool. It is used to determine level of compliance between two assessors in repeated measurement. The Braden Scale was the first-rated scale to assess the risk of decubitus in our research study. The best resulting items of the Braden scale were

Table 1 Inter-rater reliability of items of Braden Scale

Item of Braden Scale	ICC	95% CI pre ICC		Inter-rater reliability
Sensory perception	0.742	0.535	0.865	good
Moisture	0.663	0.413	0.820	good
Activity	0.846	0.708	0.922	excellent
Mobility	0.812	0.649	0.903	excellent
Nutrition	0.645	0.387	0.810	good
Friction and shear	0.712	0.487	0.848	good

Legend: ICC – The Intraclass Correlation Coefficient, CI – Confidence interval

Table 2 2 Inter-rater reliability of items of Norton Scale

Item of Norton Scale	ICC	95% CI pre ICC		Inter-rater reliability
Physical condition	0.849	0.713	0.923	excellent
Mental condition	0.891	0.788	0.945	excellent
Activity	0.893	0.792	0.946	excellent
Mobility	0.896	0.798	0.948	excellent
Incontinent	0.931	0.864	0.966	excellent

Legend: ICC – The Intraclass Correlation Coefficient, CI – Confidence interval

Table 3 Inter-rater reliability of items of Waterlow Scale

Item of Waterlow Scale	ICC	95% CI pre ICC		Inter-rater reliability
Build/weight for height	0.756	0.557	0.873	good
Skin type visual risk areas	0.717	0.495	0.851	good
Sex	1	1	1	excellent
Age	0.988	0.976	0.994	excellent
Has patient lost weight recently	0.674	0.429	0.826	good
Weight loss score	0.497	0.185	0.718	fair
Patient eating poorly or lack of appetite	0.674	0.429	0.826	good
Continence	0.713	0.49	0.849	good
Mobility	0.890	0.786	0.945	excellent
Tissue malnutrition	0.630	0.366	0.801	good
Neurological deficit	0.773	0.585	0.883	good
Major surgery or trauma	1	1	1	excellent
Medication	0.659	0.408	0.818	good

Legend: ICC – The Intraclass Correlation Coefficient, CI – Confidence interval

“Activity” and “Mobility”. The following items reached substantial compliance, according to the order: “Sensory perception”, “Friction and Shear”, “Moisture” and “Nutrition”. Authors Kottner et al. (23) report similar level of compliance among the items “Sensory perception” (ICC = 0.74), “Moisture” (ICC = 0.64), “Activity” (ICC = 0.88), “Mobility” (ICC = 0.82), “Nutrition” (ICC = 0.79) and “Friction and Shear” (ICC = 0.83). Authors Wang et al. (14) declare a higher level of compliance on items “Sensory perception” (ICC = 0.926) and “Activity” (ICC = 0.964). The rate of compliance among other items of the Braden Scale is consistent with our results. As illustrated in our research, none of the items of the Braden Scale has lower compliance rate than ICC = 0.6. This demonstrates an appropriate inter-rater reliability of the items of this scale. Such inter-rater reliability values may be caused by the fact that the Braden Scale contains more extent characteristics of the individual items of the scale. However, this results in more time consuming burden for the nurses. The results are surprising even for the researchers themselves, as this scale is not used routinely in Czech clinical environment, containing ambiguous meanings of terms like “sometimes”, “rarely” at the same time.

The second evaluated scale is the Norton Scale. Items of the Norton Scale reached the highest compliance rate among the assessors as compared to the Braden Scale and the Waterlow scale. The Norton Scale reported an overall higher compliance rate among the assessors, compared with foreign research. Authors Wang et al. (14) report lower values of the items “Physical condition” (ICC = 0.595), “Incontinence” (ICC = 0.681). Other items of the Norton scale are slightly higher than in our study: “Mental condition” (ICC = 0.929), “Activity” (ICC = 0.975) and “Mobility” (ICC = 0.911). A modified version of the Norton scale has been currently the most used scale in Czech health facilities (25). This scale includes items that are contained in the Norton Scale. Five items of this modified Norton scale may be found also in the original Norton scale. We rate this as the main reason for the high compliance among the assessors in this research study. We suppose that nurses in Czech clinical environment have knowledge of the Norton Scale as well as certain experience in its administration.

The last evaluated scale is the Waterlow scale. In our research study, the Waterlow scale reached the highest level of compliance in items: “Sex”, “Age”, “Mobility” and “Major surgery or trauma”. These are followed by “Neurological deficit”, “Build/weight for height”, “Skin type/visual risk areas with visible risk

areas”, “Continence”, “Has patient lost weight recently”, “Patient eating poorly or lack of appetite”, “Medication” and “Tissue malnutrition”. Item “Weight loss score” shows the lowest compliance rate among the assessors. Authors Wang et al. (14) declare a higher level of compliance on items “Build/weight for height” (ICC = 0.959), “Age” (ICC = 0.990), “Tissue malnutrition” (ICC = 0.737), “Neurological deficit” (ICC = 0.830) and “Medication” (ICC = 0.890). Items “Skin type/visual risk areas with visible risk areas” (ICC = 0.592), “Sex” (ICC = 0.840) and “Major surgery or trauma” (ICC = 0.600) show lower compliance rate. The rate of compliance among other items of the Waterlow Scale is consistent with our results. The Waterlow Scale is not a commonly used scale in Czech clinical environment. Achieving such compliance rate among the assessors may be explained by the fact that the Waterlow Scale contains short and clearly defined items.

Finally, the results of our research study show the highest level of compliance among the assessors using the Norton scale, followed by the Braden scale. The Waterlow scale reached the lowest compliance rate. The following items showed the highest level of compliance in all scales: “Activity”, “Mobility”, “Physical condition”, “Mental condition”, “Incontinence”, “Sex”, “Age” and “Major surgery or trauma”. Item “Weight loss score” reported the lowest compliance rate among all the items. It is necessary to pay attention to items with low compliance among the assessors. The results of our study cannot be compared with other research study in the Czech Republic. This is despite the fact the authors Mandysová et al. (20) and Mandysová et al. (19) are dealing with this issue. The reason for this is the use of different statistical methods of evaluating compliance among the assessors.

The issue of reliability of scales for assessing the risk of decubitus has been addressed only marginally in the Czech clinical environment, therefore we recommend its further testing. Work with the translated scale according to the selected translation methodology is required for the study of inter-rater reliability, as there are several translations of scales to assess the risk of decubitus in Czech scientific literature. However, it is not possible to find the translation method.

LIMITATIONS

This research study is limited by the low number of respondents. The research took place in a single clinical environment. For this reason, it is not possible to generalize our results to the entire population of the Czech Republic. Use of the ICC statistical method in

evaluating the compliance rate among the assessors may indicate another limitation, as it has not been used in the Czech Republic yet.

CONCLUSION

Our research study is focused on detecting the inter-rater reliability of individual items of three scales for assessing the risk of decubitus: the Braden Scale, the Norton Scale and the Waterlow Scale.

The Norton Scale reached the highest level of compliance of the items, followed by the Braden Scale. The Waterlow Scale showed the lowest compliance rate among the items. Item "Weight loss score" reached the lowest level of compliance. The issue of reliability of scales for assessing the risk of decubitus has been addressed only marginally in the Czech clinical environment, therefore we recommend its further testing.

REFERENCES

1. Ayello EA, Braden B. How and Why to Do Pressure Ulcer Risk Assessment. *Adv Skin Wound Care*. 2002;15(3):125-33.
2. Costa IG, Caliri MHL. Predictive validity of the Braden Scale for patients in intensive care. *Acta paul. Enferm.* 2011;24(6):772-7.
3. Serpa LF, Santos VL, Campanili TC, Queiroz M. Predictive Validity of the Braden Scale for Pressure Ulcer Risk in Critical Care Patients. *Rev Latino – Am Enfermagen*. 2011;19(1):50-7.
4. Papanikolaou P, Lyne P, Anthony D. Risk assessment scales for pressure ulcer: a methodological review. *Int J Nurs Stud*. 2007;44(2):285-96.
5. Kottner J, Hauss A, Schluer AB, Dassen T. Validation and clinical impact of paediatric pressure ulcer risk assessment scales: A systematic review. *Int J Nurs Stud*. 2013;50(6):807-18.
6. Keller BPJA, Wille J, Van Ramshorst B, Van der Werken CH. Pressure ulcers in intensive care patients: a review of risks and prevention. *Intensive Care Med*. 2002;28(10):1379-88.
7. National Pressure Ulcer Advisory Panel, European Pressure Ulcer Advisory Panel. Prevention and Treatment of Pressure Ulcer: Quick Reference Guide. Washington DC: National Pressure Ulcer Advisory Panel, European Pressure Ulcer Advisory Panel; 2014.
8. National Institute for Health and Care Excellence. Pressure Ulcer Prevention. London: National Institute for Health and Care Excellence; 2014.
9. Deutsches Netzwerk für Qualitätsentwicklung in der Pflege. Expertenstandard Dekubitusprophylaxe in der Pflege. Hochschule Osnabrück: Deutsches Netzwerk für Qualitätsentwicklung in der Pflege (DNQP); 2010.
10. Registered Nurses' Association of Ontario. Risk Assessment & Prevention of Pressure Ulcer. Toronto: Registered Nurses' Association of Ontario; 2005.
11. Chan WS, Pang SM, Kwong EW. Assessing predictive validity of the modified Braden scale for prediction of pressure ulcer risk of orthopaedic patients in an acute care setting. *J Clin Nurs*. 2009;18(11):1565-73.
12. Kwong E, Pang S, Wong T, Ho J, Shao-Ling X, Li-Jun T. Predicting pressure ulcer risk with the modified Braden, Braden, and Norton scales in acute care hospitals in Mainland China. *Appl Nurs Res*. 2005;18(2):122-8.
13. Pancorbo-Hidalgo PL, García-Fernández FP, Soldevilla-Agreda JJ, Martínez-Cuervo F. Pressure ulcer risk assessment: Clinical practice in Spain and a meta-analysis of scales' effectiveness. *Geronomos*. 2008;19(2):84-98.
14. Wang LH, Chen HL, Yan HY, Gao JH, Wang F, Ming Y, Lu L, Ding JJ. Inter-rater reliability of three most commonly used pressure ulcer risk assessment scales in clinical practice. *Int Wound J*. 2015;12(5):590-4.
15. Pancorbo-Hidalgo PL, Garcia-Fernandez FP, Lopez-Medina IM, Alvarez-Nieto C. Risk assessment scales for pressure ulcer prevention: a systematic review. *J Adv Nurs*. 2006;54(1):94-110.
16. Šáteková L, Žiaková K. Validity of pressure ulcer risk assessment scales: review. *Cent Eur J Nurs Midw*. 2014;5(2):85-92.
17. Defloor T, Grypdonck MFH. Pressure ulcers: validation of two risk assessment scales. *J Clin Nurs*. 2005;14(3):373-82.
18. Polit DF, Beck CT. Nursing research: Principles and Methods. Philadelphia: Lippincott Williams & Wilkins; 2004.
19. Mandysová P, Pechová J, Ehler E. Využití škály Bradenové pro predikci rizika vzniku dekubitů: inter – rater reliabilita. Ošetřovatelství a porodní asistence. 2013;4(3):609-13.
20. Mandysová P, Ehler E, Trejbalová L. Česká verze škály Bradenové: metodika překladu a shoda mezi posuzovateli. Ošetřovatelstvo. 2012;2(4):137-42.
21. Šáteková L, Žiaková K. Inter-rater reliabilita Bradenovej škály, Nortonovej škály a Waterlowej škály v Českej republike. In Bužgová R, Sikorová L, editors. Ošetřovatelský výzkum a praxe založená na důkazech: X. mezinárodní sympózium ošetřovatelství; 2016 květen 26; Ostrava, Česká Republika. Ostrava: Ostravská Univerzita v Ostravě; 2016, p. 214-20.
22. Kottner J, Dassen T. Interpreting interrater reliability coefficients of the Braden scale: A discussion paper. *Int J Nurs Stud*. 2008;45(8):1238-46.

23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-74.
24. Kottner J, Halfens R, Dassen T. An interrater reliability study of the assessment of pressure ulcer risk using the Braden scale and the classification of pressure ulcer in a home care setting. *Int J Nurs Stud.* 2009;46(10):1307-12.
25. Mikula J, Müllerová N. Prevence dekubitů. Praha: Grada; 2008.

CONTACT DETAILS OF MAIN AUTHOR

Mgr. Lenka Šáteková
Department of Nursing and Midwifery
Faculty of Medicine
University of Ostrava in Ostrava
Syllabova 19
703 00 OSTRAVA-ZÁBŘEH
lenkasat@gmail.com