Original Article

Danish translation and validation of the Parotidectomy Outcome Inventory 8

Carolina Maria Helena Hilton[1, 2], Jakob Foghsgaard[1], Tejs Klug[3] & Lars Morsø[4]

1) Department of Oto-Rhino-Laryngology, Head and Neck Surgery, Copenhagen University Hospital – North Zealand Hospital, Hilleroed, 2) Department of Employment: Department of Plastic Surgery and Breast Surgery, Zealand University Hospital, Roskilde, 3) Department of Oto-Rhino-Laryngology, Head and neck Surgery, Aarhus University Hospital, 4) OPEN Research Unit, Odense University Hospital, Denmark

ABSTRACT

INTRODUCTION. Complications and recurrence within benign salivary gland surgery are not systematically registered in Denmark. Patient-reported outcome measures are increasingly included in clinical and health policy decision-making, and therefore it is crucial that this type of data is valid. A patient-reported questionnaire regarding outcome after benign parotid gland surgery has been developed and implemented in a national German database. We aimed to translate the Parotidectomy Outcome Inventory 8 (POI-8) into Danish and validate it.

METHODS. The questionnaire was translated. The study population was recruited from a single centre from 6 December 2019 to 1 June 2022. Patients > 18 years of age who had undergone their first parotid salivary gland surgery for a benign tumour were included. The questionnaire underwent pilot-testing and test-retesting; it was sent to respondents twice at a 14-day interval. For the categorical variables, the reliability of the items was tested using the weighted kappa-coefficient.

RESULTS. A weighted kappa coefficient of 0.74 and Cronbach's alpha of 0.78 were found. No significant difference was found between testing at day 0 and 14.

CONCLUSIONS. We have translated and validated the Danish version of the POI-8, finding acceptable levels of the weighted kappa coefficient and Cronbach's alpha. We suggest the systematic use of PROMs in Danish healthcare and specifically in parotidectomy for benign neoplasms.

FUNDING. No funding.

TRIAL REGISTRATION. Not relevant.

Treatment of benign salivary gland tumours is provided in 12 hospitals in Denmark. The specialist plan for ear, nose and throat surgery estimates that 900 annual operations are performed (including sialoendoscopy of the major salivary glands) [1]. Complications and recurrence within benign salivary gland surgery are currently not systematically registered in Denmark. This is an area of interest as the trend – internationally and nationally – within benign salivary gland surgery is moving towards minimally invasive interventions rather than conventional salivary gland surgery [2]. When subjective concepts such as pain and quality of life are evaluated, the patient's own assessment is increasingly taken into account [3]. To better understand and measure these concepts, patient-reported data are essential. Patient-reported outcome (PRO) refers to any form of estimation of a given aspect of health obtained from the patient [4]. The term patient-reported outcome measures (PROMs)

often refers to a tool such as a questionnaire that addresses the patient's own perspective and perceived health status. PROMs allow for evaluation of specific treatments [5, 6] and are increasingly included in clinical and health policy decision-making. It is essential that the PROMs applied are evidence based and validated.

Validated PROMs form the basis for obtaining PROs such as complications. The Parotidectomy Outcome Inventory 8 (POI-8) was developed in Germany to monitor complications [7]. The questionnaire consisted of independent items and was structured according to a formative model [8]. It was evaluative but neither prognostic, diagnostic, nor part of a shared decision tool [9, 10]. A panel of experts identified symptoms related to quality of life after parotidectomy based on which 20 items were prepared. Item reduction left eight questions that were subsequently validated.

The aim of this article was to translate the POI-8 into Danish and validate the POI-8.

METHODS

The study population was recruited from the Department of Otorhinolaryngology, Head and Neck Surgery & Audiology at the North Zealand Hospital, Hilleroed, Denmark, from 6 December 2019 to 1 June 2022. The inclusion criteria were: patients > 18 years of age who underwent first parotidectomy (superficial or total) for a benign neoplasm. Patients who underwent sialoendoscopy and parotidectomy due to infection and salivary gland stones were excluded.

The questionnaire was translated by two native Danish speakers, fluent in German and with experience in translating PROMs. One translator had a medical background. The translations were compared, differences discussed and a final version was produced. The validation process was structured as follows: To test the face validity, we asked a non-expert pilot group of ten patients to do an informal review of the questionnaire regarding clarity, comprehensibility and appropriateness for the target group. Furthermore, patients commented on wording relevance, understandability and feasibility. Additionally, individual interviews were conducted to elaborate the feedback, which was then used to rephrase the questionnaire. In the pilot test, we assessed the content validity by involving subject experts in a formal assessment to determine appropriateness of contents and identify any misunderstandings or omissions.

Subsequently, the questionnaire was sent to a larger group of patients who had undergone parotidectomy. The Odense Patient data Explorative Network (OPEN, a research infrastructure, biobank and research unit) used the patients' secure email to distribute the questionnaire and collect responses in a Research Electronic Data Capture (REDCap, project database (a secure web application for building and managing online surveys and databases)).

A written consent form and the questionnaire were sent twice at a 14-day interval. We asked the patients to complete the questionnaire at approximately the same time of day and at the same location to minimise interruptions. The English language questions are available in [Appendix 1, the German and Danish versions in Appendix 2](#).

Data analysis compared characteristics of responders and non-responders, along with the individual questions between the original and translated versions. The questionnaire contained categorical variables ranked from "no problem" to "major problem". The reliability of the questions was tested using weighted kappa. Weighted kappa tests the agreement across respondents and should be used for agreement regarding ordinal scales as it takes minor and major disagreement into account. Landis & Koch's scale was used in an attempt to reach "moderate" agreement. Furthermore, Cronbach's alpha was used to test for internal consistency. Cronbach's alpha compares the amount of variance within an instrument. Though methodologically not entirely applicable, we

calculated intraclass correlation coefficients (ICC) for the instrument to be able to compare it directly with that of the German developers. All data management and analyses were performed in Stata, version 15 (StataCorp LP, College Station, Texas).

Trail registraton: Not relevant.

RESULTS

Translation

All patients agreed that the questionnaire was understandable, relevant, logical and easy to answer. Patients had comments regarding missing questions, but none were relevant to the aim of this study. Similar questions were removed during item reduction in the German study. They were related to time to diagnosis, follow-up or the time-course/information given during the day of surgery. Based on patients' comments, the phrase "bange for" (afraid of) was changed to "frygter" (fearing) in question 8. Two patients suggested using the term "udfordring" (challenge) instead of "problem" (problem) as answer option. However, the translators did not agree, and the wording remained unchanged.

Validation

A total of 93 patients met the inclusion criteria (**Figure** 1), 52 patients (56%) answered the questionnaire twice. No significant differences were found between responders and non-responders with regard to age and sex (**Table 1**).
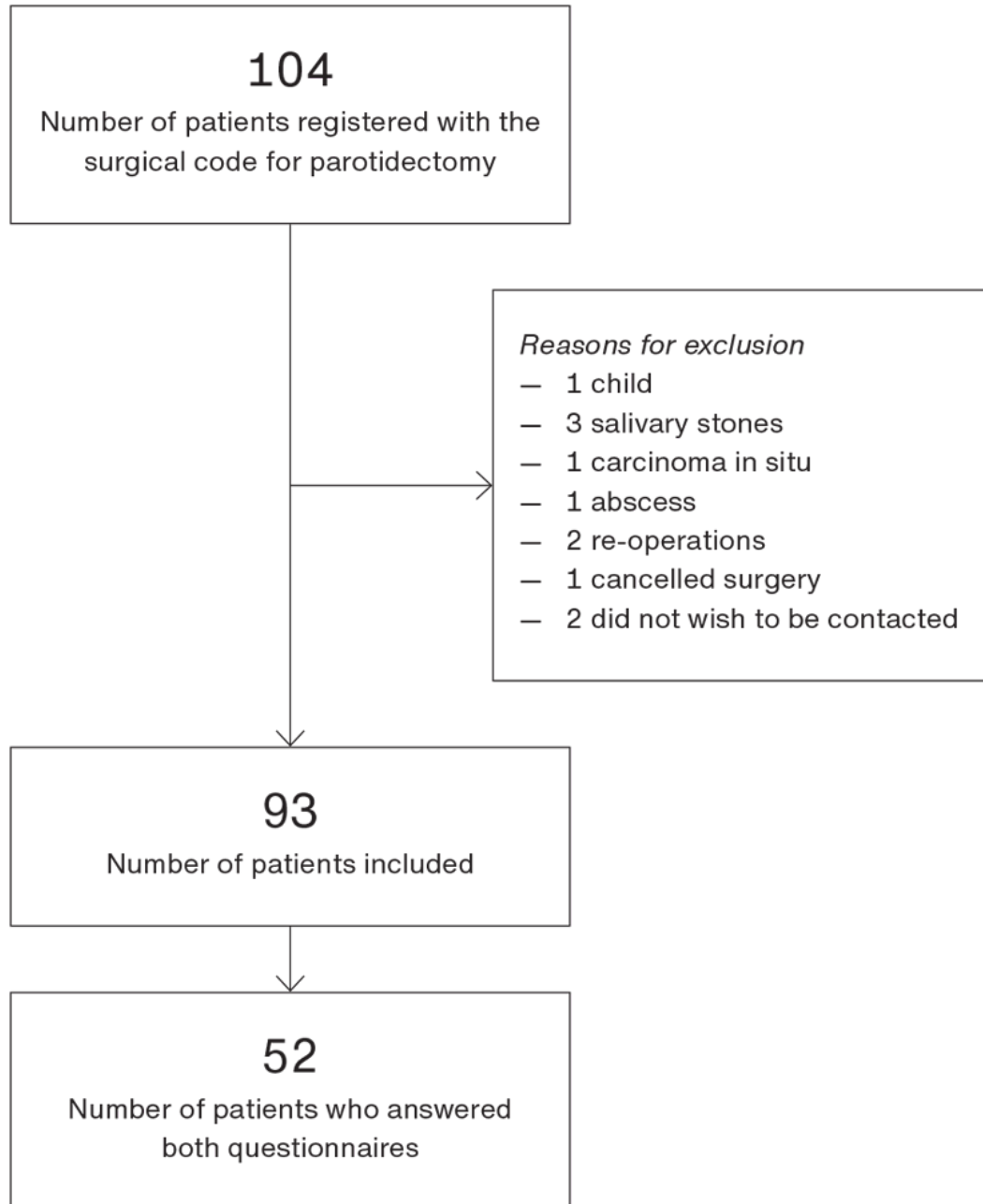
FIGURE 1 Inclusion/exclusion flow chart.

TABLE 1 Age & sex.

	Responders	Non-responders	p value
Age, median (range), yrs	63 (22-96)	57 (24-83)	0.67[a]
Sex, n			
Female	27	13	0.06[b]
Male	25	28	0.06[b]

CI = confidence interval.
a) U-test.
b) Fisher's exact test with 95% CI.

Danish patients answered > 50% of the questions with either "no problem" or "very minor problem" (Appendix 3) but scored higher on all questions than did the German patients [7] (Table 1). The mean Danish questionnaire results (using a scale of 0-5) are reported for baseline and at 14 days in **Table 2**. The mean of the German results (using a scale of 0-5) are also displayed, reported from the alfa-version of their questionnaire (for the questions used in the POI-8) [7].

TABLE 2 Mean of German vs Danish questionnaire results (using a scale of 0-5, see appendix 1).

Question no.	German	Danish (95% CI) baseline	Danish (95% CI) 14 days	Combined mean
1	0.95	1.7 (1.5-1.99)	1.6 (1.38-1.89)	1.7
2	1.84	2.3 (2.02-2.67)	2.2 (1.85-2.54)	2.3
3	1.04	2.0 (1.75-2.32)	2.0 (1.72-2.35)	2.0
4	0.22	1.3 (1.07-1.52)	1.1 (1.01-1.29)	1.2
5	0.80	1.3 (1.09-1.47)	1.3 (1.1-1.52)	1.3
6	0.88	1.2 (1.07-1.41)	1.3 (1.07-1.47)	1.3
7	0.94	1.4 (1.17-1.69)	1.5 (1.19-1.78)	1.5
8	1.03	1.6 (1.32-1.82)	1.6 (1.27-1.84)	1.6

CI = confidence interval.

For validating baseline versus 14 days, we used several statistical methods. The categorical data displayed a weighted kappa coefficient of 0.74, corresponding to substantial agreement, and a Cronbach-α of 0.78, which is acceptable. We calculated an ICC, resulting in a value > 0.50 (0.5-0.77), which demonstrated a moderate to good reliability, an expected agreement above 94% (94.23-98.56%), an agreement above 84% (84.98-96.05%), a kappa ranging from 0.49 to 0.81 and a z value ranging from 3.9 to 5.88 (**Table 3**).

TABLE 3 Intraclass correlation coefficients (ICC)/kappa values/agreement.

Question no.	ICC	Expected agreement, %	Agreement, %	Kappa (± standard error)	z	Prob > z[a]
1	0.61	95.91	89.78	0.6002 (± 0.1380)	4.35	0.0000
2	0.81	97.69	87.78	0.8111 (± 0.1380)	5.88	0.0000
3	0.77	96.51	84.98	0.7680 (± 0.1376)	5.58	0.0000
4	0.64	98.56	96.05	0.6351 (± 0.1321)	4.81	0.0000
5	0.53	95.09	89.69	0.5231 (± 0.1340)	3.90	0.0000
6	0.67	97.96	93.98	0.6605 (± 0.1379)	4.79	0.0000
7	0.73	97.62	91.36	0.7242 (± 0.1356)	5.34	0.0000
8	0.5	94.23	88.72	0.4887 (± 0.1375)	3.56	0.0002

a) Probability that the kappa agreement is not random.

DISCUSSION

We relied upon the German article to identify known and relevant side-effects important to the patient population [7]. The questionnaire was translated into Danish and after rephrasing and minor corrections, patients found that the questions were understandable and applicable. The final translation was reviewed by the working group (authors) consisting of experts within the field of parotid surgery and with experience in translation and validation of questionnaires and sufficient knowledge of the German language. Though not adhering completely to the Danish guideline referred to in the Introduction [8], with no back translation, we believe that the review process constitutes sufficient substitute and that the final translation is valid. For question number 6, we used a wording different from the original wording in German; "Svedproduktion omkring operationsarret (særligt i forbindelse med måltider)" (Production of sweat surrounding the surgical scar (especially when eating)). Frey´s syndrome is not limited to the incision area, but affects the skin surrounding the incision. Rhetorically, "the surgical area" and "surrounding the surgical scar" are very similar phrasings. Based on comments from patients confirming our assumption that patients are generally unaware of the area of surgery/the parotid gland but easily understand the area surrounding the surgical scar, we decided to use a different wording.

The questionnaire seemed to be stable over the period used and the timing for delivery/answering did not influence the answers provided. Regarding the questionnaire, mean values showed no significant difference between baseline and 14-day answers. German patients scored lower on all questions than Danish patients. We found no obvious reason for this difference, but consider it unlikely that it is due to differences in the phrasing of the questions of the POI-8. The current and minimal evidence does not point towards Danish surgical outcomes being inferior to international standards. For example, Golding CN et al. registered Frey's syndrome after large salivary gland surgery in the Central Denmark Region and found a relatively low incidence compared to the international literature [11]. The few other existing reports of facial nerve palsies due to post recurrent

pleomorphic adenoma found similar frequencies (14-29%) [12].

Various statistical methods may be used for validation. We calculated the weighted kappa-coefficient, Cronbach-α and ICC. The German study used Cronbach-α and the Pearson correlation coefficient. We applied a weighted kappa coefficient for the analysis as it relays more valid results when working with categorical data on an ordinal scale. Using ordinary kappa analysis would result in an equal weight of disagreement regardless of "the size" of the disagreement. On an ordinal scale, disagreement can be limited or quite large. Using a weighted kappa coefficient takes "the size" of the disagreement into account.

Calculation produced a weighted kappa coefficient of 0.74. The question is whether this result indicates good or poor agreement and how high a weighted kappa may be expected. For interpretation we used the scale by Landis & Koch [13]:

< 0 No agreement

0-0.2 Slight

0.2-0.4 Fair

0.4-0.6 Moderate

0.6-0.8 Substantial

0.8-1 Almost perfect

The scale indicates that an agreement with a weighted kappa coefficient of 0.74 is considered substantial. We consider a substantial agreement combined with z values ranging from 3.9 to 5.88 sufficient to use the POI-8 for clinical monitoring of outcome and side effects after salivary gland operations for benign tumours in a Danish cohort.

There are different reports about acceptable values of Cronbach´s alpha, ranging from 0.70 to 0.95[14]. The German results had a good internal consistency with a Cronbach's alpha of 0.84, comparable to the 0.78 found in the Danish cohort, both results exceeding 0.70.

Based on the 95% confidence interval of the ICC estimate, poor values are below 0.5, moderate are between 0.5 and 0.75, good between 0.75 and 0.9 and values exceeding 0.90 are indicative of excellent reliability [15]. We calculated an ICC, yielding a value > 0.50 (0.5-0.77), indicating moderate to good reliability.

If registration of complications and recurrence within salivary gland surgery was done systematically, the questionnaire could be expanded. It could then be applied to analyse symptoms at different time points after surgery. It may potentially be useful in improving salivary gland surgery by comparing two surgical options on the basis of PROMs. The reliability of PROMs has been questioned [16], but findings show that PROMs are highly reliable. PROMs are also heterogeneous. Some are used as decision tools [17], some are predictive/prognostic and others are diagnostic tools [18]. The development and quality criteria vary according to the purpose. A prognostic PROM needs to predict a patient's prognosis with a high degree of certainty. The POI-8 is not a predictive, diagnostic nor a shared decision tool. It is a validated, retrospective measuring instrument since it mainly reflects complications to surgery. Its main advantage is that it collects PRO and side effects specifically after parotid gland surgery for benign tumours. Therefore, it is essential to determine if using this instrument on Danish patients is feasible and to explore if answers given by patients are reliable and reproducible. The POI-8 has undergone the described translation and validation process, showing that these demands have been met. Therefore, we argue that the Danish version of POI-8 is a sufficiently reliable source for measuring PRO and collecting data on perceived side effects. We believe that reliable information collected with this patient tool can inform future surgery in the field of parotid gland surgery for benign tumours.

Limitations

It is difficult to compare with the German results as we have used different statistical methods. In our experience, using the weighted kappa coefficient yields more valid results when working with categorical data, but some might argue that the scale can be read as continuous. The measured agreement was high and very close to the expected agreement. It appears that the questionnaire performs as well as the German version does in Germany.

We received 50 responses to both questionnaires. Tsang S et al. proposed guidelines for the respondent-to-item ratio. They ranged from 5:1 (i.e., fifty respondents for a ten-item questionnaire), 10:1, to 15:1 or 30:1. Others suggested that sample sizes of 50 should be considered very poor, 100 as poor, 200 as fair, 300 as good, 500 as very good and 1,000 or more as excellent. The respondent-to-item ratios may be used to further strengthen the rationale for a large sample size when necessary [19]. Anthoine E et al. proposed that the sample size determination for psychometric validation studies is rarely ever justified *a priori*. This emphasises the lack of clear scientifically sound recommendations on this topic [20]. Therefore, our number might seem small, but the specific questionnaire may be thought of as a way to monitor side effects (and health-related quality of life) and therefore we consider that the collected responses are adequate [10] (Page 290, point 3 middle of the page). If, in a future setting, the questionnaire is to be used as a decision tool, further validation might be necessary.

CONCLUSIONS

We have translated and validated the Danish version of the POI-8, finding acceptable levels of the weighted kappa coefficient (0.74) and Cronbach's alpha (0.78). We recommend systematic use of PROMs in the Danish health care and specifically in parotidectomy for benign neoplasms.

Correspondence *Carolina Maria Helena Hilton*. E-mail: carro_hilton@hotmail.com

Supplementary https://content.ugeskriftet.dk/sites/default/files/2024-01/a10230633-supplementary.pdf

REFERENCES

1. Danish Health Authority. Specialevejledning for oto-rhino-laryngologi. Danish Health Authority, 2021. www.sst.dk/-/media/Viden/Specialplaner/Specialeplan-for-oto-rhino-laryngologi/Specialevejledning-for-Oto-rhino-laryngologi-den-15-juni-2021.ashx?sc_lang=da&hash=85FFBCF14AE06666D10472E66EE84A9F (Jan 2024).

2. Witt RL, Rassekh C. Gland-preserving surgery for benign neoplasms. In: Gillespie MB, Walvekar RR, Schaitkin BM et al., eds. Gland-preserving salivary surgery. A problem-based approach, 2018:147-58. doi: https://doi.org/10.1007/978-3-319-58335-8_14

3. Dawson J, Doll H, Fitzpatrick R et al. The routine use of patient reported outcome measures in healthcare settings. BMJ. 2010;340:c186. doi: https://doi.org/10.1136/bmj.c186

4.   Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes. 2006;4:79. doi: https://doi.org/10.1186/1477-7525-4-79

5.   Cano SJ, Klassen A, Pusic AL. The science behind quality-of-life measurement: a primer for plastic surgeons. Plast Reconstr Surg. 2009;123(3):98e-106e. doi: https://doi.org/10.1097/PRS.0b013e31819565c1

6.   Patrick DL, Burke LB, Gwaltney CJ et al. Content validity - establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1 - eliciting concepts for a new PRO instrument. Value Health. 2011;14(8):967-77. doi: https://doi.org/10.1016/j.jval.2011.06.014

7.   Baumann I, Cerman Z, Sertel S et al. [Development and validation of the Parotidectomy Outcome Inventory 8 (POI-8). Measurement of quality of life after parotidectomy in benign diseases]. HNO. 2009;57(9):884-8. doi: https://doi.org/10.1007/s00106-009-1991-3

8.   Willert CB, Hölmich LR, Thorborg K. Udvikling og validering af patientrapporterede spørgeskemaer - del 1. ugeskriftet.dk/videnskab/udvikling-og-validering-af-patientrapporterede-sporgeskemaer-del-1 (12 Jan 2023).

9.   Terwee CB, Bot SDM, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42. doi: https://doi.org/10.1016/j.jclinepi.2006.03.012

10.   De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. Cambridge University Press, 2011. doi: https://doi.org/10.1017/CBO9780511996214

11.   Golding CN, Larsen DG. The incidence of Frey syndrome and treatment with botulinum toxin in the Central Denmark Region. Laryngoscope Investig Otolaryngol. 2022;7(6):1814-9. doi: https://doi.org/10.1002/lio2.926

12.   Nøhr A, Andreasen S, Therkildsen MH, Homøe P. Stationary facial nerve paresis after surgery for recurrent parotid pleomorphic adenoma: a follow-up study of 219 cases in Denmark in the period 1985-2012. Eur Arch Oto-Rhino-Laryngol. 2016;273(10):3313-9. doi: https://doi.org/10.1007/s00405-016-3921-9

13.   Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74. doi: https://doi.org/10.2307/2529310

14.   Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53-55. doi: https://doi.org/10.5116/ijme.4dfb.8dfd

15.   Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-63. doi: https://doi.org/10.1016/j.jcm.2016.02.012

16.   Wennberg S, Amundsen MF, Bugten V. A validation study of the 30-day questionnaire in the national Norwegian Tonsil Surgery Register: can we trust the data reported by the patients? Eur Arch Oto-Rhino-Laryngol. 2024;281(2):977-84. doi: https://doi.org/10.1007/s00405-023-08306-0

17.   Bongers IL, Buitenweg DC, van Kuijk REFM, van Nieuwenhuizen C. I need to know: using the CeHRes Roadmap to develop a treatment feedback tool for youngsters with mental health problems. Int J Environ Res Public Health. 2022;19(17):10834. doi: https://doi.org/10.3390/ijerph191710834

18.   Nativ N, Pincus T, Hill J, Ami NB. Predicting persisting disability in musculoskeletal pain patients with the STarT MSK screening tool: results from a prospective cohort study. Musculoskeletal Care. 2023;21(4):1005-10. doi: https://doi.org/10.1002/msc.1776

19.   Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth. 2017;11(suppl 1):S80-S89. doi: https://doi.org/10.4103/sja.SJA_203_17

20.   Anthoine E, Moret L, Regnault A et al. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes. 2014;12:176. doi: https://doi.org/10.1186/s12955-014-0176-2