THE INTERNATIONAL JOURNAL OF
CLINICAL PRACTICE WILEY

How to identify the most suitable questionnaires and rating scales for your clinical practice or research?

Yara Shoman[1] | Nicole Majery[2] | Marina Otelea[3] | Charlotte Lambreghts[4,5] | Irina Guseva Canu[1]

[1]Center of primary care and public health (Unisanté), University of Lausanne, Lausanne, Switzerland

[2]Service de Santé au Travail Multisectoriel (STM), Strassen, Luxembourg

[3]Carol Davila, University of Medicine and Pharmacy, Bucharest, Romania

[4]Department of Primary Care and Public Health, University of Leuven, Leuven, Belgium

[5]IDEWE, External Service for Prevention and Protection at Work, Heverlee, Belgium

Correspondence
Yara Shoman, Center of primary care and public health (Unisanté), University of Lausanne, Lausanne, Switzerland.
Email: yara.shoman@unisante.ch

Abstract

Background: Selection of the most suitable instrument for a health outcome or exposure assessment is challenging, as there are many different instruments and their versions, most with unknown validity.

Aims: To develop guidelines facilitating the search for the most suitable instrument.

Materials and Methods: Based on our experience, we formalised a five-step process. The first step is the search for systematic reviews of available instruments validity in COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN), International prospective register of systematic reviews (PROSPERO), or conventional (eg, Medline and Web of Science) databases. If there is no systematic review, the clinician should look for original validation studies and assess them critically. We presented two alternatives of this assessment: qualitative using COSMIN and quantitative using our methodological framework. The latter helps to decide upon the instrument validity completeness and interpret the statistical results from original studies objectively. This process was then transformed into guidelines, which were tested by three external clinicians to select the most appropriate instrument to measure depression, occupational stress and daily fatigue.

Results: The guidelines were proved to facilitate the instrument search and selection, practical and time-saving.

Discussion: The guidelines assessment highlighted that clinicians should check whether the instrument that they are looking for was developed for screening or diagnosing purposes, whether it can be self-administered or not, and for which setting it was validated (academic vs clinical).

Conclusion: These guidelines facilitate the objective choice of the most suitable instrument in clinical practice by making the search simple, systematic and time-effective.

1 | INTRODUCTION

The use of questionnaires and rating scales in the clinical setting is increasing, particularly for subjective health measures, such as the quality of life in patients with multiple sclerosis.[1-3] If developed and used correctly, instruments for subjective measures can be as objective and valid as physical measures such as temperature or lung function,[4,5] but easier and less expensive.[6] To improve the quality of subjective

health measures and foster the evidence-based medicine (EBM) paradigm, the term "Patient-Reported Outcome Measure" (PROM) was introduced to designate standardised and validated instruments that are completed by patients to capture their perceptions of their health, exposure and quality of life.[7] However, identification of the best standardised and validated instrument for an outcome of interest rises multiple concerns,[8] as several instruments can be available for this outcome, and even several versions of each instrument, with a few data on their respective validity.[9] The challenge is, thus, twofold: How to find the best PROM? and How to make sure that it is valid?

These questions are essential from the EBM perspective. First, because acquiring adequate search skills strengthens EBM implementation in practice.[10] Second, because the assessment of the weight of evidence of PROM's validity before its use in medicine is a key principle of the EBM because PROMs are a key element in clinimetrics (the domain of rating scales, indices and what measure clinical phenomena such as symptoms and signs[11]).

As far as we know, the whole process of search, critical validity assessment and the choice of the best PROM has never been formalised, although using standardised and valid measures is crucial for EBM.[12] The validation process of questionnaires/rating scales is time-consuming and difficult, especially for those without training in clinimetrics. However, it is fundamental to guarantee that the PROM is clinically or psychometrically sound. It can be performed from a qualitative approach; such as using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN)[9,13] or a quantitative approach following a comprehensive methodological framework.[14]

According to COSMIN, the most crucial psychometric property is PROM's content validity. Nevertheless, COSMIN also helps to assess eight other psychometric properties (ie, criterion validity, structural validity, internal consistency, reliability, measurement error, hypotheses testing and responsiveness and cross-sectional validity). According to the methodological framework, a complete PROM validation encompasses 11 main validity assessment steps, including face validity, content validity, predictive validity, concurrent validity, convergent validity, discriminant validity, exploratory factorial validity, confirmatory factorial validity, stability, homogeneity and sensitivity. For each of these validation steps, the methodological framework provides the definition, the most appropriate analytical method and criteria for objective interpretation of resulting statistics.

For each exposure or outcome, there are often several PROMs available. For example, for occupational burnout assessment among mental health professionals, O'Connor identified eight "validated" PROMs.[15] Nevertheless after a critical assessment of their validity, according to a standardised protocol,[14] we found a moderate quality of evidence of validity for only two occupational burnout PROMs.[16] This research, which took more than a year to be completed, informed us of the need for clear guidance to simplify and shorten the performance of similar tasks in everyday practice. This need is currently even more essential with the increased use of PROMs in telemedicine and COVID-19 context.[17-20] Hence, we developed guidelines that help to make the search for the best instrument systematic, simple and time-effective and, therefore, meet the EBM requirements.

What is known?

- Questionnaires/rating scales are common in both research and clinical practice. However, their validity is often unknown. Clinicians and researchers should therefore assess it to make the best choice of the instrument for their need.
- The validation process of questionnaires/rating scales is difficult and time-consuming, especially for those without training in clinimetrics.

What is new?

- We developed guidelines to facilitate the choice of questionnaires/rating scales in clinical practice. The guidelines have been tested independently by three clinicians on different outcomes/exposures and judged helpful.

2 | MATERIALS AND METHODS

Based on our previous experience,[14,16] we formalised a five-step methodological process that we transformed into user-friendly guidelines to facilitate the selection of the best PROM for a health outcome or exposure of interest. To evaluate these guidelines, we asked three clinicians from three different countries to test them and to comment on their usefulness and clarity. Each clinician chose one outcome or exposure of their choice for this test. Clinicians completed the evaluation to find the most suitable PROM and commented on the guidelines. Although it is sometimes recommended to perform the assessment using tables of comparison between different PROMs,[21] we aimed to provide practical guidelines by simulating a real clinician's situation, and thus, we did not set any controlled conditions.

3 | RESULTS

3.1 | Guidelines presentation

The guidelines are as follows. The search for a specific exposure or health outcome can be conducted following five main steps (Figure 1). The search can be stopped at any step if the users achieved their goal and found a suitable PROM with an acceptable validity that they assessed qualitatively and quantitatively.

3.1.1 | Step 1

The first step is to use the COSMIN database. This database aims to enhance the choice of the most suitable PROMs both in clinical practice and in research. Hence, starting with this step is time saving
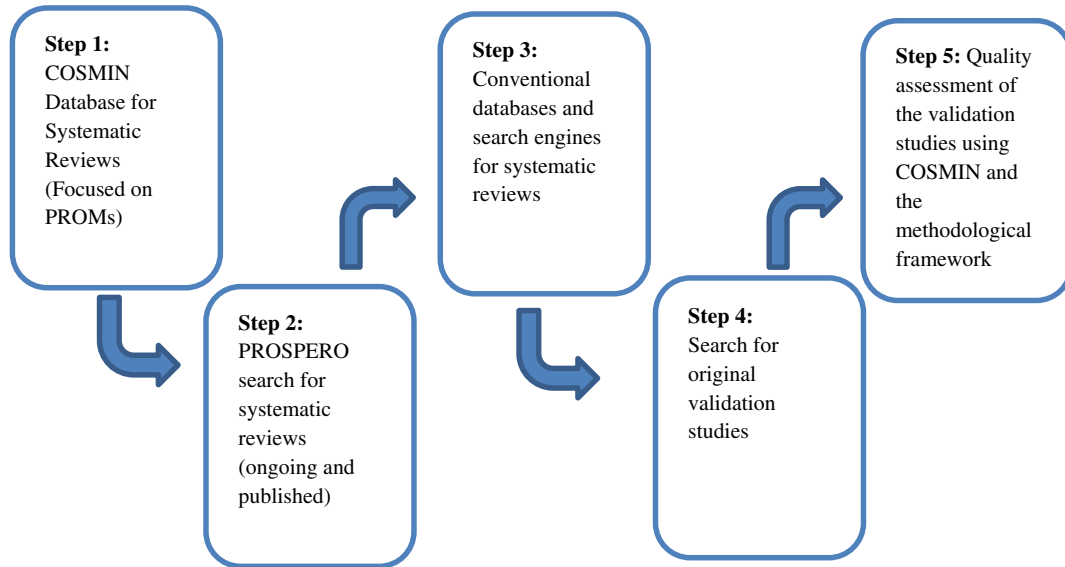
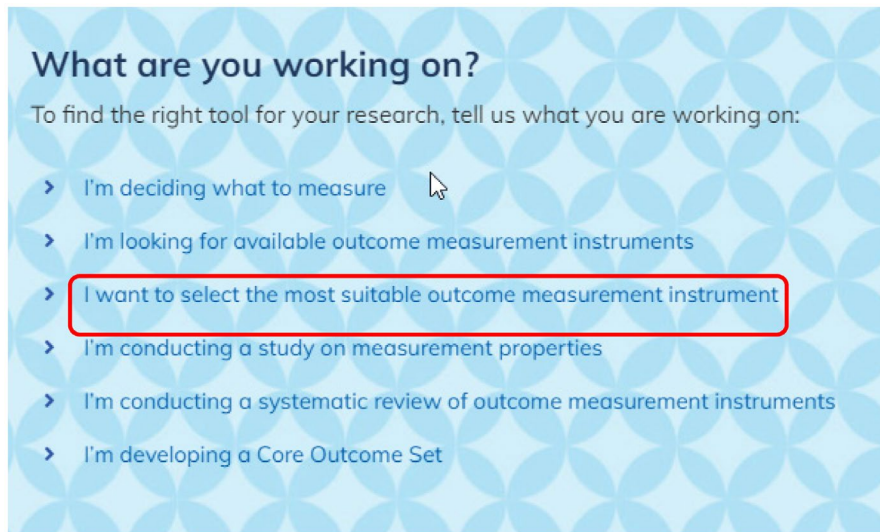**FIGURE 1** Flowchart of the guidelines to find the most suitable PROM

FIGURE 2 COSMIN user interface (adapted from https://www.cosmin.nl/)

especially if the clinician/researcher finds a convenient systematic review of PROMs for the exposure/outcome of interest. The user interface is practical and simple (Figure 2), as it offers guidance with user manuals for any option you select. By selecting the option: "I want to select the most suitable outcome measurement instrument", you will be directed to another page with a thorough explanation of further steps.

We suggest starting with the first option "Have a look in the COSMIN Database for Systematic Reviews on outcome measurement instruments to find the review of your interest." Here you can download the user manual before starting the search in the database of systematic reviews. After checking the manual, you can type your exposure/outcome in the search box and then select your filters.

In case you find a systematic review, you can directly move to step 5 and assess the quality of this review. If the systematic review

you found was not of adequate quality or you did not find any systematic review, please proceed to step 2.

3.1.2 | Step 2

The following step is to search for ongoing or published systematic reviews in the International prospective register of systematic reviews (PROSPERO) (Figure 3). You can choose your filters here as well and type the name of the exposure/outcome in the search box, then check the results of the search. If you just type the name (eg, depression), you will have a large number of results because, unlike COSMIN, PROSPERO is not focused on PROMs. Therefore, it is better to include the word "measure," "measures" and "measurement" with the function "AND," as shown in Figure 3. If you

FIGURE 3   PROSPERO interface and search results (adapted from https://www.crd.york.ac.uk/prospero/#searchadvanced)

did not find any relevant systematic review, you should go to the next step.

3.1.3 | Step 3

This step consists of checking the conventional databases and search engines (eg, Medline, Web of Science) for systematic reviews. You can type your search query and check the results for systematic reviews on the exposure/outcome of interest. For an effective search in these databases, you can use these guidelines.[22,23]

3.1.4 | Step 4

If there is no systematic review of PROM(s) for assessing the exposure/outcome of your interest, you should look for original validation studies of every identified PROM through the conventional databases cited in step 3. Compared with systematic reviews, assessing the original validation studies is more time-consuming and difficult to perform even though some methods[22,23] can facilitate this process.

3.1.5 | Step 5

As a starting point, it is possible to use COSMIN for a qualitative evidence appraisal of the studies with the help of the manuals provided on the COSMIN website (Figure 4). The two user manuals entitled "COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs)" and "COSMIN methodology for assessing the content validity of PROMs" are particularly helpful and worth reading attentively the first time you perform this task.

Nevertheless, from a quantitative approach, the methodological framework can facilitate estimating the completeness of validity assessment. Once you open the link to the framework (https://www.medrxiv.org/content/10.1101/2020.06.24.20138115v1.full.pdf), please scroll down to the end of the document and you will find a large table (this is the methodological framework). Using this framework is time-saving as you can use the function "Ctrl+F" on your keyboard to find the validity tests or statistics you are looking for. For instance, if you type "content validity index" in the search box (Figure 5), you can see to which validation step this statistic belongs, the definition of the validation step, the description of the statistic, the interpretation of the indices to decide the quality of the results.

To help you conducting a systematic review of PROMs, we developed several tools for you to use

A comprehensive user manual is written on how to conduct these ten steps in a systematic review of PROMs. In step 5 the content validity of an PROM is determined. As this is a very important and complex step to take, we wrote a separate user manual for assessing content validity of PROMs.

- COSMIN METHODOLOGY FOR SYSTEMATIC REVIEWS OF PATIENT-REPORTED OUTCOME MEASURES (PROMS) – USER MANUAL
- COSMIN METHODOLOGY FOR ASSESSING THE CONTENT VALIDITY OF PROMS – USER MANUAL

Additional tools to download:
- SEARCH FILTER FOR FINDING STUDIES ON MEASUREMENT PROPERTIES
- RISK OF BIAS CHECKLIST FOR ASSESSING THE QUALITY OF STUDIES
- HELP ORGANIZING YOUR COSMIN RISK OF BIAS RATINGS
- HELP EXTRACTING THE DATA AND REPORTING ALL YOUR RESULTS
- COSMIN RATINGS FOR DEVELOPMENT STUDIES OF PROMS

FIGURE 4 COSMIN user manuals (adapted from https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/)

By applying the former steps, you should be able to assess the validity of different PROMs available for the outcome of your interest or at least to correctly understand the validity study results to make a reasonable choice of the most suitable instrument.

3.2 | Guidelines evaluation

Three clinicians tested the guidelines by searching for the most suitable PROM for measuring two health outcomes and one exposure. The first outcome was depression. The search resulted in choosing the second version of the Beck Depression Inventory (BDI-II)[24] as the most validated available PROM for screening because its validity was assessed in four reviews, including one systematic review.[25] Nevertheless, the Patient Health Questionnaire-9[26] was found to be a better option for clinical practice as it is a free PROM and its completion takes less time compared with BDI-II. The second outcome was the daily fatigue in patients with sleep apnoea. The clinician chose the Sleep Apnoea Quality of Life Index (SAQLI)[27] based on the results of a systematic review that assessed the validity of 22 PROMs.[28] Occupational stress was chosen by the third clinician who identified the General Work Stress Scale[29] as the most available valid PROM for her research.

After the clinicians familiarised themselves with the guidelines, the search process for the best PROM following the guidelines

took 12-36 hours overall. Each step separately took between 1 and 6 hours except the fifth step (from 2 to 20 hours). Among the first four steps, the first, third and fourth steps took between 1 and 4 hours, whereas the second step took between 4 and 6 hours.

4 | DISCUSSION

As far as we know, there is no clear guidance on how to search for the best PROMs, nor for assessing their validity as a clinimetric tool. Several standards have been proposed for assessing the methodological quality of original studies focused on some particular psychometric properties of a PROM.[30,31] COSMIN guidelines provide a qualitative assessment of original validation studies when conducting systematic reviews, but it is still recent, not well known, and somehow incomplete. Some authors attempted to guide the choice of PROMs based on their validity but only for research purposes.[32,33] We have presented guidelines that can facilitate the selection of the best and most valid PROM in both research and clinical practice. Three clinicians tested these guidelines and provided their feedback to increase their clarity and effectiveness. However, after familiarising themselves with the guidelines, the process took different time spans depending on the number of steps they had to follow, which differed from a

Validation step	Definition	Analysis/techniques	Indices and interpretation
Content validity	The degree in which the scale content adequately reflects the construct that is being measured; it evaluates how much an item sample represents in a defined universe or content domain[2].	*Scale items cover all aspects of the construct* [4]. Compare test content to the theoretical construct content to see if they are related [4]. 1) Qualitative approach: assessment by an expert committee 2) Quantitative approach: calculation of the content validity index (**CVI**) to measure the proportion of judges who agree on certain aspects of a tool and its items [2].	a) Deductive approach. b) Four points Likert scale where items rated with 1 or 2 points have to be removed. **CVI** = number of answers 3 or 4/total number of answers. .>8 = acceptable >.9 preferable [2].

FIGURE 5 The illustration of the search in the methodological framework (adapted from https://www.medrxiv.org/content/10.1101/2020.06.24.20138115v1)
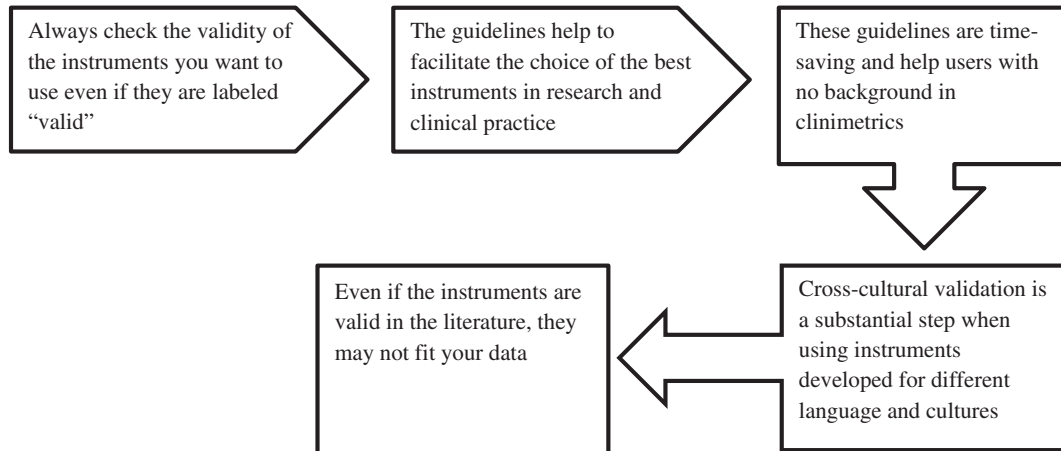

FIGURE 6 Lessons learned

minimum of two steps (eg, the first and fifth steps) to all the five steps. For depression and fatigue, they did not need to search for original validation studies. However, for occupational stress, there were no completed systematic reviews yet, and the clinician had to assess the validity of original studies, which took considerably more time. Consequently, the guidelines provided an effective roadmap and time-saving tool with a comprehensive summary of the most used psychometric tests and the interpretation of their statistical results.

The feedback also highlighted that clinicians should check whether the PROM that they are looking for was developed for screening or diagnosing purposes, whether it can be self-administered or not, and for which setting it was validated (academic versus clinical). In the example of daily fatigue, SAQLI cannot be self-administered,[28] for occupational stress the General Work Stress Scale was validated for research purpose only,[29] and for depression, BDI-II was originally validated for screening.[34]

Another important conclusion from the evaluation of the guidelines is that PROMs labelled as "valid" should also be checked. Many widely used PROMs are not as valid as they may be reported. Therefore, the clinician should always question their validity. Additionally, many PROMs are validated in the original language, but this does not guarantee their cross-cultural validity when they are translated in another language or used in a different population (eg, French-speaking Swiss or Canadians vs French). For cross-cultural validity, you can check these three references.[21,35,36]

It is noteworthy to mention that even if a PROM is valid in the literature, it is recommended to retest it based on the collected data. However, this matter is beyond the scope of this article, and some useful recourses already exist.[37-39]

Finally, compared with over 1 year spent to find the best available PROM for burnout, spending approximately 36 hours following these guidelines to find the best PROM for any exposure or outcome appears to save a significant amount of time. Logically, the fifth step of the guidelines is much longer compared with all other steps because it requests returning to original studies examining PROM's validity. The second step (ie, using the PROSPERO database) can be more time-consuming compared with the search in COSMIN or conventional databases because PROSPERO provides the search results for ongoing and published reviews while the other databases restrict the search for only published reviews. In some cases such as the example of occupational stress, clinicians may not find a completely valid PROM, but at least they will be able to choose the best available one and make an informed choice. These guidelines are particularly helpful when the clinicians or researchers lack the required skills required for literature search and clinimetrics and help to make the search for the best PROM simple, systematic and time-effective.

5 | LESSONS LEARNED

It is essential to always check the validity of the instruments before using them even when they are labelled as "valid" in the literature

(Figure 6). The presented guidelines facilitate the instrument choice based on its clinical and psychometric validity and the usage purpose (ie, research, screening or clinical practice). These guidelines were tested for the instruments for exposure and health outcome assessment and were proved to be time-saving and helpful, particularly for users not familiar with clinimetrics.

6 | CONCLUSIONS

The presented guidelines facilitate the objective choice of the most suitable questionnaire/rating scale in clinical practice and research. The time spent using the guidelines can differ not only based on the quality and availability of reviews but also the original validation studies of questionnaires/rating scales. Nevertheless, following these guidelines saves time and improves clinicians' and researchers' education regarding the most important psychometric properties of a PROM, validation steps and the interpretation of their statistical results. Consequently, these guidelines should help to make the search for the best PROM simple, systematic and time-effective, in compliance with the EBM paradigm.

DISCLOSURES

The authors have declared no disclosures.

REFERENCES

1. Baumstarck K, Boyer L, Boucekine M, Michel P, Pelletier J, Auquier P. Measuring the quality of life in patients with multiple sclerosis in clinical practice: a necessary challenge. *Mult Scler Int*. 2013;2013:524894.
2. Wolfe F. The effective use of questionnaires in clinical practice. *Arthrit Rheum-Arthr*. 2007;57(5):705-706.
3. Wood J, Gupta S. Using rating scales in a clinical setting: A guide for psychiatrists. *Current Psychiatry*. 2017;16(2):21-25.
4. Falissard B. *Mesurer la subjectivité en santé: Perspective méthodologique et statistique*, 2nd Falissard editor. ed. Elsevier; 2008.
5. Hamilton D, Giesinger J, Giesinger K. It is merely subjective opinion that patient-reported outcome measures are not objective tools. *Bone and Joint Research*. 2018;6(12):665-666.
6. Demetriou C, Özer B, Self-Report EC. *Self-report questionnaires. The encyclopedia of clinical psychology*. John Wiley & Sons; 2015.
7. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *Bja Educ*. 2017;17(4):137-144.
8. Paiva SM, Perazzo MF, Ortiz FR, Pordeus IA, Martins-Junior PA. How to select a questionnaire with a good methodological quality? *Braz Dent J*. 2018;29(1):3-6.
9. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-549.
10. Majid S, Foo S, Luyt B, et al. Adopting evidence-based practice in clinical decision making: nurses' perceptions, knowledge, and barriers. *J Med Libr Assoc*. 2011;99(3):229-236.
11. Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. *Int J Clin Pract*. 2012;66(1):11-15.
12. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ*. 1995;310:1122-1126.
13. Mokkink LB, De Vet HC, Prinsen CA, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171-1179.
14. Marca SC, Paatz P, Gyorkos C, Cuneo F, Bugge MD, Godderis L, et al. Validation of questionnaires and rating scales used in medicine: protocol for a systematic review of burnout self-reported measures. 2020.
15. O'Connor K, Neff DM, Pitman S. Burnout in mental health professionals: a systematic review and meta-analysis of prevalence and determinants. *Eur Psychiat*. 2018;53:74-99.
16. Shoman Y, Marca S, Bianchi R, Godderis L, van der Molen H, Guseva CI. Psychometric properties of burnout measures: a systematic review. *Epidemiol Psychiatric Sci*. 2021;30(e8):1-9. 10.1017/S2045 796020001134
17. Hajesmaeel-Gohari S, Bahaadinbeigy K. The most used questionnaires for evaluating telemedicine services. *Bmc Med Inform Decis*. 2021;21(1):1-2.
18. Kumari A, Ranjan P, Vikram NK, et al. A short questionnaire to assess changes in lifestyle-related behaviour during COVID 19 pandemic. *Diabetes Metab Syndr*. 2020;14(6):1697-1701.
19. Luo H, Yongchan L, Frits W, Prinzen F. Surveillance of COVID-19 in the general population using an online questionnaire: report from 18,161 respondents in China. *JMIR Public Health Surveill*. 2020;6(2):e18576.
20. Zurlo MC, Della Volta MFC, Vallone F. COVID-19 student stress questionnaire: development and validation of a questionnaire to evaluate students' stressors related to the coronavirus pandemic lockdown. *Front Psychol*. 2020;11:576758.
21. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25(24):3186-3191.
22. Baumann N. How to use the medical subject headings (MeSH). *Int J Clin Pract*. 2016;70(2):171-174.
23. Citrome L, Moss SV, Graf C. How to search and harvest the medical literature: let the citations come to you, and how to proceed when they do. *Int J Clin Pract*. 2009;63(11):1565-1570.
24. Beck A, Steer R, Brown G. *Manual for the beck depression inventory-II*. Psychological Corporation; 1996.
25. Wang YP, Gorenstein C. Assessment of depression in medical patients: a systematic review of the utility of the Beck Depression Inventory-II. *Clinics (Sao Paulo)*. 2013;68(9):1274-1287.
26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613.
27. Flemons WW, Reimer MA. Development of a disease-specific health-related quality of life questionnaire for sleep apnea. *Am J Respir Crit Care Med*. 1998;158(2):494-503.
28. Abma IL, van der Wees PJ, Veer V, Westert GP, Rovers M. Measurement properties of patient-reported outcome measures (PROMs) in adults with obstructive sleep apnea (OSA): a systematic review. *Sleep Med Rev*. 2016;28:18-31.
29. Stanton JM, Balzer WK, Smith PC, Parra LF, Ironson G. A general measure of work stress: the stress in general scale. *Educ Psychol Meas*. 2001;61(5):866-888.
30. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.
31. Wade R, Corbett M, Eastwood A. Quality assessment of comparative diagnostic accuracy studies: our experience using a modified version of the QUADAS-2 tool. *Res Synth Methods*. 2013;4(3):280-286.
32. Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. *BMJ*. 2004;328(7451):1312-1315.

33. Taherdoost H. Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *Int J Acad Res Manag.* 2016;5(3):28-36.

34. Maruish ME. *Handbook of pediatric psychological screening and assessment in primary care introduction.* 2018, 3-56.

35. Fasi J. Questionnaires: selecting, translating and validating In: Research ElfHaC, editor. the Netherlands; 2010. p. 1-5.

36. Yasir A. Cross cultural adaptation & psychometric validation of instruments: step-wise description. *Int J Psychiatry.* 2016;1(1):1-4.

37. Boynton PM. Administering, analysing, and reporting your questionnaire. *BMJ.* 2004;328(7452):1372-1375.

38. European Commission. Qualitative methodologies for questionnaire assessment In: ISTAT, editor. Luxembourg. 2017. https://circabc.europa.eu/sd/a/7f617c55-1b01-41a5-96a4-966394f28b32/Methodological%20document%20-%20qualitative%20methods%20for%20pretesting.pdf. Accessed February 17, 2017.

39. Patel S. How do you ensure your survey questions are valid and reliable? 2019. http://salmapatel.co.uk/academia/survey-questions-valid-reliable/. Accessed April 17, 2019.