RESEARCH



Inconclusiveness of psychometric testing of medication adherence questionnaires

Isabelle Arnet¹ · Christiane Eickhoff² · Laura J Sahm³ · Sabine Caloz¹ · Michael Mittag⁴ · Martin Schulz^{2,5} · Samuel S Allemann¹

Received: 12 February 2024 / Accepted: 27 March 2024 © The Author(s) 2024

Abstract

Purpose To propose a paradigm change for the validation procedures of medication adherence questionnaires.

Methods A total of 121 validation procedures of unique questionnaires for medication adherence were analyzed.

Results "Construct validity" and "internal consistency" were most often assessed, and test results varied largely. A more in-depth analysis indicated that the assessment of medication non-adherence included distinct but related constructs, such as the extent to which doses are missed, and the attempt to identify different facets of medication-taking behavior. Consequently, each construct requires a different measurement approach with different psychometric tests for establishing its validity and reliability.

Conclusion Results show that assessing the validity and reliability of adherence questionnaires with standard procedures including statistical tests is inconclusive. Refinement of the constructs of non-adherence is needed in pharmacy and medical practice. We suggest a distinction between the (i) *extent* of missed doses over the past 2 weeks, (ii) *modifiable reasons* for non-adherence behavior, and (iii) *unmodifiable factors* of non-adherence. Validation procedures and corresponding statistical methods should be selected according to the specific single constructs.

Keywords Medication adherence · Questionnaires · Reliability · Validity · Psychometric testing · Self-report

Introduction

According to its most recent definition, medication adherence is a process by which patients use their medicines according to the recommendations [1], that is, a behavior that corresponds to the agreed regimen. Medication nonadherence is a well-described and huge challenge in the prevention, management, and treatment of patients. It is

 ☑ Isabelle Arnet isabelle.arnet@unibas.ch
 Christiane Eickhoff

C.Eickhoff@abda.de

Laura J Sahm l.sahm@ucc.ie

Sabine Caloz sabine.caloz@bluewin.ch

Michael Mittag michael.mittag@fhnw.ch

Martin Schulz m.schulz@abda.de associated with increased healthcare costs [2], morbidity [3], and mortality [4]. As an example, poor adherence to antibiotic therapy contributes to the risk of increasing antimicrobial resistance with all of its associated consequences [5]. Thus, healthcare providers need accurate estimates of individual patients' medication-taking behavior to ameliorate this problem and develop tailored interventions aimed at improving medication adherence. Three different phases of

Samuel S Allemann s.allemann@unibas.ch

- ¹ Pharmaceutical Care, Pharmaceutical Sciences, University of Basel, Basel, Switzerland
- ² Department of Medicine, ABDA–Federal Union of German Associations of Pharmacists, Berlin, Germany
- ³ Pharmaceutical Care Research Group, School of Pharmacy, University College Cork, Cork, Ireland
- ⁴ Fachhochschule Nordwestschweiz, Windisch, Switzerland
- ⁵ Institute of Pharmacy, Freie Universität Berlin, Berlin, Germany

Table 1The nine COSMIN
measurement properties with
corresponding definitions. The
domain "validity" contains
three measurement properties:
content validity, construct
validity, and criterion validity.
The domain "reliability" also
contains three measurement
properties: internal consistency,
reliability, and measurement
error. Adapted from [21] and
[22]

Image: Note of the second se	Domain		COSMIN measurement	COSMIN definition			
Velidity) an instrument is an adequate reflection of the construct to be measured Appint VOIDEN Structural validity The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured Hypotheses testing for construct validity Item construct validity Cross-cultural validity The degree to which the performance of the items on a translated or culturally adapted instrument is an adequate reflection of the performance of the items of the original version of the instrument Image: Provide the intervalue of a "gold standard" The degree to which the scores of an instrument are an adequate reflection of a "gold standard" Image: Provide the intervalue of the performance of the intervalue of a "gold standard" The degree of the intervalue items Image: Provide the intervalue over time The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured			property				
Image: Section of the construct to be measured reflection of the construct to be measured Image: Section of the construct validity The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured Hypotheses testing for construct validity Item construct validity Cross-cultural validity The degree to which the performance of the items on a translated or culturally adapted instrument is an adequate reflection of the performance of the items of the original version of the instrument Image: Section validity The degree to which the scores of an instrument are an adequate reflection of the performance of the items of the original version of the instrument Image: Section validity The degree of the interrelatedness among the items Test-retest reliability The degree of the interrelatedness among the items Test-retest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured			Content validity (including face	The degree to which the content of			
Image: Properties of the second sec			validity)	an instrument is an adequate			
LINE Structural validity The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured Hypotheses testing for construct validity Hypotheses testing for construct to be measured Item construct validity Cross-cultural validity The degree to which the performance of the items on a translated or culturally adapted instrument is an adequate reflection of the performance of the items of the original version of the instrument Criterion validity The degree to which the scores of an instrument are an adequate reflection of the performance of the items of the original version of the instrument The degree to which the scores of an instrument are an adequate reflection of a "gold standard" Test-retest reliability The degree of the interrelatedness among the items Test-retest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured				reflection of the construct to be			
LICITY Instrument are an adequate reflection of the dimensionality of the construct to be measured Hypotheses testing for construct validity Item construct validity Validity Cross-cultural validity The degree to which the performance of the items on a translated or culturally adapted instrument is an adequate reflection of the performance of the items of the original version of the instrument Image: Construct Validity Criterion validity The degree to which the scores of an instrument are an adequate reflection of a "gold standard" Image: Value Val				measured			
LINE Image:			Structural validity	The degree to which the scores of an			
LIDION Image: section of the section of a "gold standard" LINERAL Internal consistency The degree of the interrelatedness among the items Test-retest reliability The degree of the section of a section over time Measurement error Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured				instrument are an adequate reflection			
Light Hypotheses testing for construct validity Item construct validity Cross-cultural validity The degree to which the performance of the items on a translated or culturally adapted instrument is an adequate reflection of the performance of the items of the original version of the instrument Criterion validity The degree to which the scores of an instrument are an adequate reflection of a "gold standard" Internal consistency The degree of the interrelatedness among the items Test-retest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured				of the dimensionality of the construct			
Light validity The degree to which the performance of the items on a translated or culturally adapted instrument is an adequate reflection of the performance of the items of the original version of the instrument Image: Criterion validity Criterion validity The degree to which the scores of an instrument are an adequate reflection of a "gold standard" Image: Criterion validity The degree of the interrelatedness among the items The degree of the interrelatedness among the items Internal consistency The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured				to be measured			
Image: Content of the image: Content on the image: Conten	~	ity	Hypotheses testing for construct	Item construct validity			
Image: Content of the image: Content on the image: Conten	-IDIT	valid	validity				
Image: Content of the image: Content on the image: Conten	VAI	Construct	Cross-cultural validity	The degree to which the performance			
Image: Participation of the service				of the items on a translated or			
Image: Performance of the items of the original version of the instrument original version of the instrument Image: Performance of the items of the original version of the instrument The degree to which the scores of an instrument are an adequate reflection of a "gold standard" Image: Performance of the items of the instrument are an adequate reflection of a "gold standard" Internal consistency The degree of the interrelatedness among the items Image: Performance of the interrelated rest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Image: Performance of the interrelated rest reliability The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured				culturally adapted instrument is an			
Image: Provide the i				adequate reflection of the			
Image: Properties Criterion validity The degree to which the scores of an instrument are an adequate reflection of a "gold standard" Image: Properties Internal consistency The degree of the interrelatedness among the items Test-retest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured Properties Responsiveness The ability of an instrument to detect				performance of the items of the			
Image: Provide the internal consistency Internal consistency The degree of the interrelatedness among the items Test-retest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured				original version of the instrument			
Image: Provide the internal consistency The degree of the interrelatedness among the items Internal consistency The degree of the interrelatedness among the items Test-retest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured Responsiveness The ability of an instrument to detect			Criterion validity	The degree to which the scores of an			
Internal consistency The degree of the interrelatedness among the items Test-retest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured Patient's p				instrument are an adequate reflection			
Image: Provide the second structure among the items Test-retest reliability The extent to which scores for patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured Responsiveness The ability of an instrument to detect				of a "gold standard"			
Image: Sector of the sector of the same for repeated measurement over time Image: Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured Image: Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured			Internal consistency	The degree of the interrelatedness			
Patients who have not changed are the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured Responsiveness The ability of an instrument to detect				among the items			
Here the same for repeated measurement over time Measurement error The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured Responsiveness The ability of an instrument to detect			Test-retest reliability	The extent to which scores for			
patient's score that is not attributed to true changes in the construct to be measured				patients who have not changed are			
patient's score that is not attributed to true changes in the construct to be measured	ELIABILITY			the same for repeated measurement			
patient's score that is not attributed to true changes in the construct to be measured				over time			
true changes in the construct to be measured	8		Measurement error	The systematic and random error of a			
measured The ability of an instrument to detect				patient's score that is not attributed to			
Responsiveness The ability of an instrument to detect				true changes in the construct to be			
Responsiveness The ability of an instrument to detect change over time in the construct to be measured				measured			
No No Change over time in the construct to No No No No No No <	2IC		Responsiveness	The ability of an instrument to detect			
be measured	PON	VESS		change over time in the construct to			
	RESI	ш		be measured			

adherence that have been defined are initiation, implementation, and persistence [1], each is linked to different behaviors and thus requires different interventions when insufficient. There is no "gold standard" for assessing adherence [6]. Nonetheless, scales and self-report questionnaires remain a widely used method since they are inexpensive, easy to administer in any setting, and deliver immediate results. A large number of questionnaires to assess adherence have been developed in the past decades [7]. For all these instruments, the quality should be assured as a precondition to gather reliable data that can be used as a robust base for findings and interpretations.

The psychometric quality of a scale is defined by its validity and reliability [8]. One prerequisite to validity is that concepts exist and are represented by factors, and that variation in these factors affects the resulting scores on the questionnaire [9]. Depending upon the purpose of a questionnaire and how the items were developed, different statistical methods are used to establish validity and reliability [10].

In addition to statistical considerations, analytical approaches can be misleading, not least because a correlation between two variables does not mean that one is a measure of the other [11]. Thus, the appropriate statistical methods remain unclear [12], and results (in this context psychometric property) may not always be robust. For some older scales for example [13, 14], shortcomings have been detected including overly simplistic items or a scoring procedure that has not been subjected to adequate testing [15] making the use of these scales less reliable. Consequently, various frameworks and procedures to validate questionnaires have emerged [16]. Templates [17] or best practice guidelines [18] have been proposed, and standards such as the consensus-based COSMIN guide-

Table 2 Frequency of the psychometric analysis for validity, reliabil	-
ity, and responsiveness according to COSMIN used in 121 question	-
naires on adherence. Adapted from [24]. Note that "criterion valid	-

lines (COnsensus-based Standards for the selection of health Measurement INstruments) [19] have been developed. To facilitate further standardization, a COSMIN glossary of terms with their respective measurement properties has been defined [20] (Table 1).

However, despite these efforts, consensus among stakeholders is lacking in the selection of the optimal (self-report) instrument to assess medication adherence, although it is widely agreed that this measure should be included routinely in health records for primary care [23]. Unsurprisingly, many existing adherence questionnaires lack sufficient validity and reliability, as shown by Kwan [24]. This systematic review presents a summary of 121 validation procedures of unique questionnaires for medication adherence, published in 214 studies in 32 languages from 48 countries [24]. Based on the COSMIN guidelines, Kwan evaluated the methodological quality of each study and the psychometric results of each questionnaire to determine the level of evidence of the questionnaires and to recommend their use. However, Kwan did not summarize the measurement properties. Thus, our aims were (i) to calculate the frequency of the psychometric analysis of 121 validation procedures from Kwan's review and (ii) to propose a paradigm change for the validation procedures of adherence questionnaires.

Results

Out of the 121 validation procedures of Kwan's review, we compiled a list of those measurement properties that were assessed (independent of quality) and how often the results were rated as sufficient (Table 2).

ity" was not assessed because there is no gold standard in the field of medication adherence. Also, note that "sufficient" was determined against COSMIN criteria for good measurement properties

	Validity				Reliability			Responsiveness
	Content validity	Hypotheses testing for construct validity	Structural validity	Cross- cultural validity	Test-retest reliability	Measurement errors	Internal consistency	
Frequency of psychometric analysis [%]	67.8%	88.4%	50.4%	1.6%	42.2%	0%	67.8%	38.8%
Frequency of psychometric analysis with sufficient results [%]	42.2%	46.3%	6.6%	0%	17.4%	0%	52.2%	18.2%

The measurement properties most often assessed were "construct validity" (that is, hypotheses testing for construct validity, structural validity, and cross-cultural validity) and "internal consistency" although "content validity" is considered to be the most important [25]. Crosscultural validity was assessed for 1.6% of the questionnaires, although they were developed in 32 different languages. However, test results varied largely, and thus, a more in-depth interpretation is warranted.

If the assessment of non-adherence is reduced to the extent to which doses are missed, then the multitude of factors that underpin that behavior, as evidenced by 40 years of adherence research [26], are overlooked. Thus, in addition to an inquiry about the number of missed doses, questionnaires attempted to identify different facets of medication-taking behavior, and items elicited information regarding barriers to good adherence or beliefs about medicines [27]. Consequently, distinct but related constructs of non-adherence are measured, which have been summarized as the *extent* to which doses are missed and the reasons for missing doses [28]. Based on this dual conceptualization of non-adherence, each construct requires a different measurement approach with different statistical tests for its evaluation [29], i.e., for establishing its validity and reliability. Questionnaires that identify reasons for non-adherence mostly use validation methods that focus on content and construct validity [27]. However, because each reason for non-adherence is expected to stand alone and can only be correlated to others in case of redundancy [29], some authors propose to validate the *reasons* items individually and with fewer tests, e.g., reliability with test-retest and validity with cognitive interviews [28, 29].

Discussion

In the current landscape, it is reasonable to claim that assessing the validity and reliability of medication adherence scales with standard procedures including statistical tests is inconclusive. Test results must be interpreted considering context and culture (among other variables) so that an oversimplified interpretation is avoided, particularly in the case of questionnaires that have not undergone rigorous methodological testing. Furthermore, suitable statistical methods must be chosen in line with the desired outcomes. Finally, we claim that a refinement of the constructs of non-adherence is needed in pharmacy and medical practice and research since questionnaire results should inform treatment decisions or enable tailored interventions to increase adherence.

Moreover, and following the tenets of Vrijens et al. [1], we suggest separating initiation and persistence from implementation. Not starting the treatment (non-initiation) or stopping the treatment too early (non-persistence) results in

"missed doses," but patients rarely think of them this way. When speaking with patients, the healthcare practitioners can ask whether the patient has discontinued, or never initiated, treatment. Thus, we suggest distinguishing three different constructs in the context of poor implementation. Firstly, the *extent* of missed doses over the past 2 weeks. In the literature, the time frame ranged from 1 day to 12 months [27]. Although a short recall period is said to minimize recall errors [30], the optimal time frame is unknown [31]. Nevertheless, forgetfulness is most likely to occur on a Saturday and/or Sunday, that is, when the daily rhythm is broken [32]. Thus, we believe that 2 weeks is an appropriate time frame being both short enough that patients will be able to recall and including a weekend. Secondly, modifiable reasons for non-adherence, that is, underlying barriers that may be possible to modify by the patient. They consist of unintentional (e.g., forgetfulness, inherent beliefs about treatment and medicines) or intentional reasons (e.g., lack of motivation, perceived need for medicine, or self-efficacy). Regarding adverse drug reactions, fear of side effects (that is, not actual side effects) is a well-known reason that leads to intentional non-adherence [33]. However, another reason is the presence of minimal but nonetheless uncomfortable side effects (for example feeling nauseous) usually requiring patients to seek advice from their doctor but may result in the patient skipping a dose or simply discontinuing their treatment [34]. Finally, drug shortages, particularly in low and middle-income countries, may put patients at risk of stopping their treatment for a while, independently of the healthcare system characteristics [35]. This may be inaccurately termed "voluntary" non-adherence. As cognitive and emotional factors guide people's action [36], the emotion associated with the treatment, or the treatment-taking behavior, should be taken into consideration when creating items for medication adherence questionnaires (or instruments). To our knowledge, there is no core set of modifiable reasons (cognitions) for medication adherence. However, the theoretical domains framework (TDF) would appear to be both appropriate and comprehensive when describing the determinants of adherence behavior [37]. It captures 33 theories and 84 theoretical constructs related to behavior change [38] and matches behavior change techniques [39] to each of its 14 domains. Thirdly, there are unmodifiable factors (e.g., cognitive or sensory impairments, sociodemographic factors) negatively affecting adherence; these can be personal situational factors over which patients have little to no control. From the literature, it would appear that these unmodifiable factors are poorly correlated to non-adherence behaviors (e.g., suboptimal response to ACE-inhibitors by African Americans due to a low circulating plasma renin profile [40]); it would therefore be better to prioritize the development of interventions aimed at increasing adherence by means of changing modifiable factors. In addition,

streamlining the items could be a practical advantage, that is, starting with the *extent* items could help identify patients with suboptimal levels of adherence. Following with the *modifiable reasons* items could enable to define targeted and feasible interventions.

As healthcare practitioners do not routinely have access to digital tools to be provided to patients, questionnaires are often filled out and scored by hand, on paper, so that every item adds to the burden of the healthcare practitioners, adding to the enormous time pressure. Thus, trade-off considerations are generally made by primary care physicians and pharmacists. However, if they choose to use a shorter questionnaire, they should keep in mind the three broad categories of abovementioned drivers. Then, if the initial questionnaire indicates that the patient is at behavioral risk, a more complete questionnaire might be useful to determine what the cause of non-adherence might be.

It is noteworthy to mention that adherence questionnaires in general can be broadly grouped into (i) those that purely center on inquiring into patient's actual behavior (e.g., missed doses) and (ii) those that rather focus on patient's attitudes (e.g., beliefs about medication). In this context, it is not surprising that none of the 121 questionnaires to measure non-adherence and analyzed by Kwan is satisfactory enough; otherwise, the best one would be the most used, especially after decades of research on non-adherence. The earlier questionnaires tended to consist only of items pertaining to the behavioral category, while more recent questionnaires, after Kwan's analysis, have few or even no items that specifically ask about actual behavior. As an example, the SPUR (Social, Psychological, Usage, and Rational) questionnaire was developed in 2022 to profile type-2 diabetes patients and determine the risk of non-adherence via 27 items on attitudes [41]. However, the ideal adherence questionnaire should assess actual behavior while at the same time provide insights into its drivers, that is, the reasons behind non-adherence, as recently shown in the 15-STARS questionnaire [42].

Conclusions

Assessing non-adherence with minimal efforts and without the need to delve into its reasons, i.e., with a simple questionnaire, may represent a significant step forward and might be acceptable for many healthcare practitioners. However, adherence questionnaires tackle different aspects of medication-taking behavior and therefore different constructs. Consequently, they are seldom comprehensive instruments. Thus, like there is no "one-size-fits-all" adherence questionnaire, there is no standard statistical test battery to determine the validity and reliability of these questionnaires. Validation procedures and corresponding statistical methods should match the single constructs. Thus, practitioners should carefully assess their need to identify the problem (i.e., noninitiation, non-implementation, or non-persistence) and only then determine its drivers with respect to the time and effort they are ready to expend. Finally, the interpretation of the test results deserves caution and pragmatism.

Acknowledgements The authors wish to thank the participants of the Pharmaceutical Care Network Europe (PCNE) Workshop 2022 for their inputs to the manuscript, especially Anita Tuula, Anna Laven, Beatriz Santos, Dominik Stämpfli, Ellen Koster, João Gregório, Manfred Krüger, María Pilar Modamio Charles, Rikke Nørgaard Hansen, Selina Barbati, and Viktoria Wurmbach.

Author contribution I.A. and C.E. conceptualized the manuscript. S.C. performed the analysis. M.M. insured statistical control. I.A. wrote the main manuscript text. All authors reviewed the manuscript.

Funding Open access funding provided by University of Basel The authors did not receive support from any organization for the submitted work.

Availability of data and materials No datasets were generated or analyzed during the current study.

Declarations

Ethical approval Not applicable.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Vrijens B, De Geest S, Hughes DA, Przemysław K, Demonceau J, Ruppar T, Dobbels F, Fargher E, Morrison V, Lewek P, Matyjaszczyk M, Mshelia C, Clyne W, Aronson JK, Urquhart J, for the APT (2012) A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 73(5):691–705. https://doi.org/10.1111/j.1365-2125.2012.04167.x
- Cutler RL, Fernandez-Llimos F, Frommer M, Benrimoj C, Garcia-Cardenas V (2018) Economic impact of medication nonadherence by disease groups: a systematic review. BMJ Open 8(1):e016982. https://doi.org/10.1136/bmjopen-2017-016982
- Walsh CA, Cahir C, Tecklenborg S, Byrne C, Culbertson MA, Bennett KE (2019) The association between medication nonadherence and adverse health outcomes in ageing populations:

a systematic review and meta-analysis. Br J Clin Pharmacol 85(11):2464–2478. https://doi.org/10.1111/bcp.14075

- Van Alsten SC, Harris JK (2020) Cost-related nonadherence and mortality in patients with chronic disease: a multiyear investigation, national health interview survey, 2000–2014. Prev Chronic Dis 17:E151. https://doi.org/10.5888/pcd17.200244
- Bergsholm YKR, Feiring M, Charnock C, Holm LB, Krogstad T (2023) Exploring patients' adherence to antibiotics by understanding their health knowledge and relational communication in encounters with pharmacists and physicians. Explor Res Clin Soc Pharm 12:100372. https://doi.org/10.1016/j.rcsop.2023.100372
- Lam WY, Fresco P (2015) Medication adherence measures: an overview. BioMed Res Int 2015:217047. https://doi.org/10.1155/ 2015/217047
- Nassar RI, Basheti IA, Saini B (2022) Exploring validated selfreported instruments to assess adherence to medications used: a review comparing existing instruments. Patient Prefer Adherence 16:503–513. https://doi.org/10.2147/PPA.S352161
- Karras DJ (1997) Statistical methodology: II. Reliability and validity assessment in study design, Part B. Acad Emerg Med 4(2):144–174. https://doi.org/10.1111/j.1553-2712.1997. tb03723.x
- Borsboom D, Mellenbergh GJ, van Heerden J (2004) The concept of validity. Psychol Rev 111(4):1061–1071. https://doi.org/ 10.1037/0033-295x.111.4.1061
- Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Group tMFP-ROCM (2007) What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value in Health 10(s2):S94–S105. https://doi.org/10.1111/j.1524-4733. 2007.00272.x
- Sechrest L (2005) Validity of measures is no simple matter. Health Services Research 40(5p2):1584–1604. https://doi.org/10.1111/j. 1475-6773.2005.00443.x
- Schmidt ME, Steindorf K (2006) Statistical methods for the validation of questionnaires–discrepancy between theory and practice [abstract]. Methods Inf Med 45(4):409–413
- Kim MT, Hill MN, Bone LR, Levine DM (2000) Development and testing of the Hill-Bone Compliance to High Blood Pressure Therapy Scale. Prog Cardiovasc Nurs 15(3):90–96. https://doi. org/10.1111/j.1751-7117.2000.tb00211.x
- Morisky DE, Green LW, Levine DM (1986) Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 24(1):67–74. https://doi.org/10.1097/00005650-198601000-00007
- Koschack J, Marx G, Schnakenberg J, Kochen MM, Himmel W (2010) Comparison of two self-rating instruments for medication adherence assessment in hypertension revealed insufficient psychometric properties. J Clin Epidemiol 63(3):299–306. https:// doi.org/10.1016/j.jclinepi.2009.06.011
- Hinkin TR (1995) A review of scale development practices in the study of organizations. J Manage 21(5):967–988. https://doi.org/ 10.1016/0149-2063(95)90050-0
- Elangovan N, Sundaravel E (2021) Method of preparing a document for survey instrument validation by experts. MethodsX 8:101326. https://doi.org/10.1016/j.mex.2021.101326
- Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL (2018) Best practices for developing and validating scales for health, social, and behavioral research: a primer. Frontiers in Public Health 6. https://doi.org/10.3389/fpubh.2018. 00149
- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC (2006) Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol 6:2. https:// doi.org/10.1186/1471-2288-6-2

- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patientreported outcomes. J Clin Epidemiol 63(7):737–745. https://doi. org/10.1016/j.jclinepi.2010.02.006
- Mokkink LB, Prinsen C, Patrick D, Alonso J, Bouter LM, de Vet HC, Terwee CB (2018) COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual. Available at URL: https://www.cosminnl/wp-content/ uploads/COSMIN-syst-review-for-PROMs-manual_version-1_ feb-2018pdf (last access 27 July 2022)
- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 10:22. https://doi.org/10.1186/1471-2288-10-22
- National Institutes of Health and Society for Behavioral Medicine (NIH and SBM) (2011) Identifying core behavioral and psychosocial data elements for the Electronic Health Record - executive summary. available at URL: https://www.sbmorg/UserFiles/file/ EHR_Meeting_May_2-3-2011--Executive_Summarypdf (last access 126 July 2022)
- 24. Kwan YH, Weng SD, Loh DHF, Phang JK, Oo LJY, Blalock DV, Chew EH, Yap KZ, Tan CYK, Yoon S, Fong W, Ostbye T, Low LL, Bosworth HB, Thumboo J (2020) Measurement properties of existing patient-reported outcome measures on medication adherence: systematic review. J Med Internet Res 22(10):e19179. https://doi.org/10.2196/19179
- 25. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, de Vet HCW, Mokkink LB (2018) COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 27(5):1159–1170. https://doi.org/10.1007/s11136-018-1829-0
- Kardas P, Lewek P, Matyjaszczyk M (2013) Determinants of patient adherence: a review of systematic reviews. Front Pharmacol 4:91. https://doi.org/10.3389/fphar.2013.00091
- Nguyen T-M-U, Caze AL, Cottrell N (2014) What are validated self-report adherence scales really measuring? a systematic review. Br J Clin Pharmacol 77(3):427–445. https://doi.org/10. 1111/bcp.12194
- Voils CI, Hoyle RH, Thorpe CT, Maciejewski ML, Yancy WS Jr (2011) Improving the measurement of self-reported medication nonadherence. J Clin Epidemiol 64(3):250–254. https://doi.org/ 10.1016/j.jclinepi.2010.07.014
- Voils CI, Maciejewski ML, Hoyle RH, Reeve BB, Gallagher P, Bryson CL, Yancy WS Jr (2012) Initial validation of a self-report measure of the extent of and reasons for medication nonadherence. Med Care 50(12):1013–1019. https://doi.org/10.1097/MLR. 0b013e318269e121
- McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, Thomas R, Harvey E, Garratt A, Bond J (2001) Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. Health Technol Assess 5(31):1–256. https://doi.org/10.3310/hta5310
- 31. Gomes D, Placido AI, Mó R, Simões JL, Amaral O, Fernandes I, Lima F, Morgado M, Figueiras A, Herdeiro MT, Roque F (2019) Daily medication management and adherence in the polymedicated elderly: a cross-sectional study in Portugal. Int J Environ Res Public Health 17 (1). https://doi.org/10.3390/ijerph17010200
- 32. Boucquemont J, Pai ALH, Dharnidharka VR, Hebert D, Zelikovsky N, Amaral S, Furth SL, Foster BJ (2020) Association between day of the week and medication adherence among adolescent and young adult kidney transplant recipients. Am J Transplant 20(1):274–281. https://doi.org/10.1111/ajt.15590

- 33. Unni EJ, Gupta S, Sternbach N (2024) Reasons for non-adherence with antidepressants using the Medication Adherence Reasons Scale in five European countries and United States. J Affect Disord 344:446–450. https://doi.org/10.1016/j.jad.2023.10.010
- 34. Marston MT, Berben L, Dobbels F, Russell CL, de Geest S (2023) Prevalence and patient-level correlates of intentional non-adherence to immunosuppressive medication after hearttransplantation-findings from the international BRIGHT study. Transplant International: Official Journal of the European Society for Organ Transplantation 36:11308. https://doi.org/10.3389/ti. 2023.11308
- 35. Bou Malhab S, Haddad C, Sacre H, Hajj A, Zeenny RM, Akel M, Salameh P (2023) Adherence to treatment and harmful effects of medication shortages in the context of severe crises: scale validation and correlates. J Pharm Policy Pract 16(1):163. https://doi. org/10.1186/s40545-023-00667-5
- 36. Tomljenovic H, Bubic A (2021) Cognitive and emotional factors in health behaviour: dual-process reasoning, cognitive styles and optimism as predictors of healthy lifestyle, healthy behaviours and medical adherence. Curr Psychol 40(7):3256–3264. https:// doi.org/10.1007/s12144-019-00268-z
- Allemann S, Nieuwlaat R, van den Bemt B, Hersberger K, Arnet I (2016) Matching adherence interventions to patient determinants using the Theoretical Domains Framework. Front Pharmacol 7 (429). https://doi.org/10.3389/fphar.2016.00429
- Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, Foy R, Duncan EM, Colquhoun H, Grimshaw JM, Lawton R, Michie

S (2017) A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 12(1):77. https://doi.org/10.1186/s13012-017-0605-9

- Michie S, Johnston M, Francis J, Hardeman W, Eccles M (2008) From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Applied Psychol 57(4):660–680
- Williams SF, Nicholas SB, Vaziri ND, Norris KC (2014) African Americans, hypertension and the renin angiotensin system. World J Cardiol 6(9):878–889. https://doi.org/10.4330/wjc.v6.i9.878
- 41. de Bock E, Dolgin K, Arnould B, Hubert G, Lee A, Piette JD (2022) The SPUR adherence profiling tool: preliminary results of algorithm development. Curr Med Res Opin 38(2):171–179. https://doi.org/10.1080/03007995.2021.2010437
- 42. Arnet I, Sahm LJ, Gregorio J, Tuula A, Kruger M, Wurmbach VS, Hoti K, Schulz M, Eickhoff C (2024) Development and validation of the 15-STARS - a novel self-report pharmacy-based questionnaire to screen for medication non-adherence. Research in social & administrative pharmacy: RSAP 20(3):308–320. https://doi.org/ 10.1016/j.sapharm.2023.11.005

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.