ASSESSMENT OF DIAGNOSTICS TOOLS FOR SARCOPENIA SEVERITY USING THE ITEM RESPONSE THEORY (IRT)

M. STEFFL[1], M. MUSALEK[1], V. KRAMPEROVA[1], M. PETR[1], E. KOHLIKOVA[1], I. HOLMEROVA[2], L. VOLICER[3]

1. Faculty of Physical Education and Sport, Charles University, Prague, Czech Republic; 2. Centre of Expertise in Longevity and Long-Term Care, Faculty of Humanities, Charles University, Prague, Czech Republic; 3. School of Aging Studies, University of South Florida, Tampa, USA and 3rd Medical Faculty, Charles University, Prague, Czech Republic. Corresponding author: M. Steffl, Charles University in Prague Faculty of Physical Education and Sport, Department of Physiology and Biochemistry, Jose Martiho 31, Prague, 16252, Czech Republic, steffl@ftvs.cuni.cz

Abstract: *Objective:* To use the item response theory (IRT) methods to examine the degree to which the four selected tools reflect sarcopenia and to arrange them according to their ability to estimate sarcopenia severity. *Design:* A cross-sectional study aimed at verifying the possibilities of using diagnostic tools for sarcopenia. *Setting and Participants:* The study included residents living in an assisted living unit at the Senior Centre in Blansko (South Moravia, Czech Republic) (n=77). Sarcopenia was estimated according to the proposals of the European Working Group on Sarcopenia in Older People (EWGSOP) using calf circumference, the EWGSOP algorithm, hand grip strength, and the Short Physical Performance Battery (SPPB). *Results:* The results from the IRT model showed that these four methods indicate strong unidimensionality so that they measure the same latent variable. The methods ranked according to the discrimination level ranging from high to low discrimination where the calf circumference was the most discriminatory (Hi = 0.86) and the SPPB together with hand grip strength were the least discriminatory (both Hi = 0.44). *Conclusion:* We are recommending to identify mild sarcopenia by SPPB or hand grip strength, moderate sarcopenia by the EWGSOP algorithm and severe sarcopenia by the calf circumference.

Key words: Sarcopenia diagnostics, aging, institutional care facilities.

Introduction

Sarcopenia is the age-related loss of muscle mass and strength defined by Rosenberg in 1989 (1). Even though sarcopenia treatment has been an acute problem, development of therapies of sarcopenia has lagged, partly because of the lack of consensus about definitions for this condition (2). Sarcopenia is highly prevalent in long-term care settings such as the assisted living unit where we did our research. For example about 85% of patients suffered from sarcopenia among institutionalized male elderly individuals in Turkey (3) and 32.8% were identified as affected by sarcopenia in the teaching nursing homes of Catholic University of Rome (4). Nevertheless, the prevalence of sarcopenia in older adults varied significantly when different diagnostic criteria have been applied (5). These variations may have been caused by different levels of its severity. Depending on the severity we would treat sarcopenia by different approaches. For that reason, estimation of sarcopenia severity is an important issue for its treatment in clinical practice. A few diagnostic tools which have been described previously (6) could be used effectively in practice. In this study we proceeded to divide them according to their ability to recognize sarcopenia severity.

When results of diagnostics methods are considered as directly measurable variables (items), we can effectively calculate their measuring ability to estimate condition according to severity by using the item response theory (IRT) (7-9). The essence of the IRT lies in the possibility to determinate the diagnostic property of all dichotomously or polytomously scored indicators and scales (10). IRT provides two very important parameters: 1) discrimination coefficient and 2) difficulty coefficient with a possible equivalent parameter of strictness. In this case the parameter of difficulty/strictness refers to what structural and physical condition rated on Likert scale the tested person must have in order to not be declared as sarcopenic. Therefore, specifically, the IRT approach allows us to assign for each examined person the highest probability degree of investigated latent construct/trait level (7, 11) - i.e. severity of sarcopenia. From a clinical point of view, less difficult/strict tests which patients are not able to perform may detect severe functional or structural impairment and by using highly difficult/strict tests we are better able to recognize impairment when it is still mild and consequently we can start the treatment on time (10, 12). Nonparametric models of the Moken scales (13, 14) represent an efficient IRT approach for detecting both the level of unidimensionality of tests and their difficulty and discrimination properties. These models are, unlike other approaches such as factor analysis or IRT parametric models, free of some strong preconditions (linearity between directly and indirectly measurable variables, the logistic shape of the item response function). For more on the principles and preconditions of their use see the section Statistics. The Mokken scales, which have already been used in previous studies assessing the aforementioned properties of the indicators, such as in IADL (15, 16), showed that more information concerning functional impairment can be obtained using the IRT methods than using a summed score (16, 17).

ASSESSMENT OF DIAGNOSTICS TOOLS FOR SARCOPENIA SEVERITY USING THE ITEM RESPONSE THEORY (IRT)

Therefore, the aim of the study was to examine if the four diagnostics tools reflect the same construct (sarcopenia) and how strict they are in their identification of sarcopenia and then to arrange them according to their abilities to estimate sarcopenia severity.

Materials and methods

Participants

The current study involved the participation of 77 elderly subjects (17 males and 60 females). Participants were selected by using purposeful sampling from population living in an assisted living unit at the Senior Centre in Blansko. The criteria of exclusion for participation in the study included: a cardio-stimulator; metal implants; walking aids including canes; use of diuretic drugs; hand and wrist injuries. This study was approved by the Ethics Committee of the Charles University in Prague, and written informed consent was obtained from all participants.

Procedure

Sarcopenia was diagnosed by using calf circumference, EWGSOP algorithm, hand grip strength and SPPB according to the proposals of EWGSOP (6) as follow. Calf circumference measurements were obtained with a cloth tape at the location of the greatest circumference. Calf circumference <31 cm has been associated with sarcopenia (18). The EWGSOP algorithm is shown in Figure 1 (6). To measure body composition, the Professional Body Composition Analyser InBody 720 - Biospace Co., Ltd. Korea was used. Skeletal muscle mass (SMM), which was determined by InBody 720 as a part of resulting protocol, was converted to the skeletal muscle mass index (SMI), by being divided by squared height (kg/m2). The cut-off point proposed by Janssen et al (19) for SMI using bioimpedance analysis BIA was for men: ≤10.75 kg/m2 and for women ≤6.75 kg/m2. Hand grip strength was measured by hand grip dynamometer (Takei TKK A5401 Digital Hand-grip Dynamometer). Cut-off points by gender were <30 Kg for men and <20 Kg for women according to Lauretani et al (20). A 4-m course test was used to measure gait speed. The absolute time(s) of completion the 4-m course was converted to a gait speed (m/s). The cut-off point was <0.8 m/s (18).

There are three components of SPPB balance; gait speed and chair stand tests (21). The balance test included three timed standing positions with feet together, semi-tandem and tandem. Gait speed (s) was measured on 4-m course at a preferred/comfortable walking speed, and in the chair stand test the participants completed five repetitive chair stands as quickly as possible without assistance from the upper limbs. SPPB score was the sum of the scores on these tests where every participant could reach a score between 0 - 4 points for each test. The results were defined as follows: sarcopenia 0-6; pre-sarcopenia 7-9 and no sarcopenia 10-12 points in SPPB total score (21).

Statistics

The nonparametric item response theory (NIRT) was used to evaluate the difficulty/strictness of four selected tests and for assessment of their discrimination property. Data were analyzed through the Mokken scaling analysis (MSA) package of freeware software «R» (22). The Mokken scaling analysis is a nonparametric method based on the principles of the IRT, the models produce an ordinal scale for comparison of tested persons with regard to the level of latent trait, which also allows to specify whether items are on hierarchy based means scores (16). In the Mokken scaling analysis, two probabilistic models can be used: either monotone homogeneous model (MHO) or double monotonous model (DMM) (13, 14, 23). The main difference between NIRT and IRT parametric models is that some strict rules are released from the NIRT models (24).

The MHO model is based on the assumption of unidimensionality, local independence and monotony of the item response function (IRF) that must not decline on a scale in progress at any point. If these conditions are met, the person can be sorted according to their level of trait, only on the basis of a simple sum of the test scores (25). The MHO detects unidimensionality items by the coefficient of scalability, which was introduced by Loevinger (26). For each item, coefficients discrimination (Hi) within a defined scale are computed, which indicate whether a given item has a sufficient coherence on the scale. This Hi has values in the interval (0-1) wherein the lower value of this coefficient means a less discriminatory test (27). If each of the HiS coefficients is >0.3, then we can say, according to Mokken (11), that items create a unidimensional scale. From these HiS, the overall scale factor H is calculated, which measures the quality and strength of the Mokken scale. The scale with H≤0.3 is not considered unidimensional, H≤0.4 shows weak unidimensionality, H≤0.5 represents medium scale and H>0.5 is already considered as a very homogeneous scale (28, 29).

In this case, the difficulty coefficient is expressed as a mean score which takes values in the interval (0-1), wherein the higher value of this coefficient means a less difficult test (30). The DMM, which is designed for polytomously scored items, is a more restrictive model, because in addition to the three previous conditions, there is another: the condition of the non-intersection of the IRFs. If all the aforementioned conditions are accomplished, then it indicates that the items are invariantly ordered (IIO). The IIO allows the verification of hierarchical scales that are replicable across files of defined population. This means that it allows to confirm whether all tested persons perceive items identically with regard to their «difficulty». The IIO is expressed through Htrsns (Hτ) coefficient, which is equivalent to H in terms of determining the power and weakness of scale (24).

Results

The study participants had a mean age of 83.0 ±6.3 years (79.3 ±6.4 males and 82.1 ±5.6 females), most of them were females (76.6 %), Table 1 shows characteristics of the participants. Sarcopenia prevalence ranged significantly; it was from 19.5% according to calf circumference to 87.0% according to hand grip strength (Table 2).

Table 1
Descriptive Characteristic of All Participants

Characteristic	N = 77
Age (years)	83.0 ± 6.3
Females	59 (76.6%)
Height, cm	157.1 ± 8,8
Weight, kg	69.2 ± 13.5
Calf circumference, cm	35.1 ± 4.6
Hand grip, kg	
Female	14.0 ± 6.2
Males	24.0 ± 9.0
Usual gait speed, m/s	0.5 ± 0.3
SMI, kg/m²	
Female	6.4 ± 1.1
Males	9.2 ± 1.2
Chair stand tests, s	19.1 ± 11.0
Gait speed, time s for 4 m	13.2 ± 15.5
Balance test, points	2.0 ± 1.7

Note. Categorical variables were presented as numbers (percentages), and continuous variables were presented as means ± standard deviations

Table 2
Prevalence of sarcopenia identified by four tools

| Tool	N = 77
Hand grip strength	67 (87.0%)
SPPB	62 (80.5%)
EWGSOP algorithm	34 (44.2%)
Calf circumference	15 (19.5%)

Data analysis resulted in total H = 0.57 that indicated strong unidimensionality; which means that the selected tests actually measured one common latent variable. All Hi values were higher than the threshold of 0.3 and ranged from 0.44 to 0.86, which indicates that the conditions for MHM models were fulfilled. Thus methods could be ranked in terms of the discrimination level from low to very discriminatory – SPPB and hand grip strength were the least discriminatory (Hi = 0.44 for both) while the calf circumference was the most discriminatory with a value of Hi = 0.86. The EWGSOP algorithm was placed between them with a value of Hi = 0.64. Because the order of scales was the same for all levels of latent trait and Hτ = 0.58, the conditions for DMM models were fulfilled. Out of the four tests - calf circumference, hand grip strength, SPPB, and EWGSOP algorithm, which are used to diagnose sarcopenia, "hand grip strength" was the most difficult and strictest evaluation method for the tested population (mean score of 0.13), while calf circumference was the item with the lowest level of difficulty (mean score of 0.81). Discrimination and difficulty properties of used tests are presented in Table 3.

Table 3
Difficulty and discrimination properties of used tests

Items	Item difficulty mean score	Item discrimination
Hand grip strength	0.13	0.44
SPPB	0.27	0.44
EWGSOP algorithm	0.56	0.64
Calf circumference	0.81	0.86
	Hτ = 0.58	H = 0.57

Figure 1
EWGSOP algorithm proposed by Cruz-Jentoft et al (6)
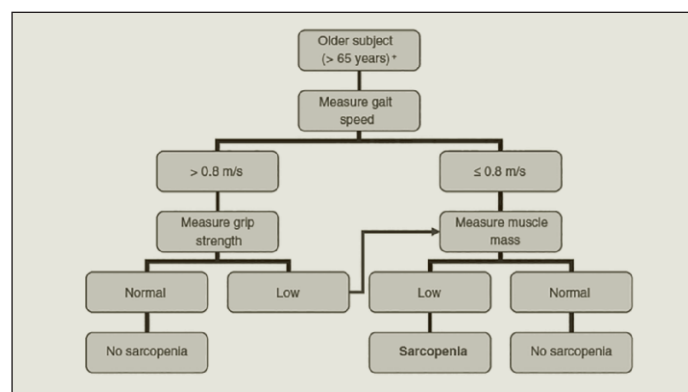
Discussion

The prevalence of sarcopenia depends on how the condition is defined (31). Different strategies influence the ability to diagnose sarcopenia, the optimal tool has not yet been proposed. For example, hand grip strength, which has been declared as a suitable tool for practical use, has various cut-off points. For example, there have been proposed the following values: grip strength <26 Kg for men; <16 Kg for women by McLean et al (32), and <37.0 for men and <21.0 for women by Sallinen et al (33). The prevalence of sarcopenia according to hand grip strength could vary according the the cut-off point used. All those cut-off points have validity in sarcopenia diagnostics, they all could estimate sarcopenia, but each on

ASSESSMENT OF DIAGNOSTICS TOOLS FOR SARCOPENIA SEVERITY USING THE ITEM RESPONSE THEORY (IRT)

a different level. We tried to solve the problem precisely in our study by IRT method. In the diagnosis of sarcopenia, we currently have only a minimal information about the degree of uniformity (unidimensionality) of the diagnostic tests towards the evaluated trait or verification of their difficulty/strictness and discrimination. For this reason, we decided to take advantage of the IRT and its benefits to demonstrate 1) whether the selected tests recommended for the sarcopenia diagnosis in clinical practice actually measure one common latent trait and 2) what is the hierarchy of the tests according to their difficulty/strictness as well as discrimination among the elderly living in institutional long-term care facilities.

The IRT is concerned with the analysis of scoring tests, questionnaires and similar tools for the measurement of various capabilities. It is based on the application of mathematical models to test data. The main idea of the IRT is that the likelihood of passing the test item is the mathematical function of the respondent and item parameters. The discrimination coefficient provides us an important piece of information on how well or finely each item differentiates the participants with relatively low levels of ability (trait) from the others with a relatively high level of ability (trait). For example, an item with a low level of discrimination coefficient will distinguish with great difficulty between the patients with moderate and severe functional impairment. The result is that the likelihood of this item being positive is virtually the same for all patients. Nevertheless, determining between severe or moderate impairment is an important issue for diagnosis. The higher is the discrimination coefficient of the item, the more finely it distinguishes between the levels of the assessed traits and thereby increases the likelihood of the adequate determination of actual diagnosis.

In the past, the IRT was used to analyze items in activities of daily living (ADLs) or instrumental activities of daily living (IADLs) (15, 16) and Mini-Mental State Examination (MMSE) (34). For example, in the study McGrory, Shenkin, Austin and Starr (14), the authors concluded that in IADL, the most discriminatory item was «Shopping» while the item «Travel» had the lowest discrimination. In the area of evaluation studies of sarcopenia, Steffl et al. (35) evaluated sarcopenia among the elderly in nursing home using the IRT. They concluded that «Chair Stand» had the highest difficulty level of the tests within EWGSOP proposals, which 78% of subjects in the study were not able to pass.

Risk of sarcopenia remains an unsolved problem mainly in assisted living facilities. Its solution such as early nutritional supplementation, which should be an essential strategy (36), is, among the others, dependent on correct and timely diagnosis. We believed that our analysis presented in this study could bring new focus on the problem of sarcopenia. Our results have shown that the selected diagnostic tools really measure the common latent variable - sarcopenia. Nevertheless, their difficulty/strictness and discrimination vary significantly. While hand grip strength as well as SPPB can identify sarcopenia

when impairment has been still mild, EWGSOP algorithm as a gold standard diagnoses moderate sarcopenia, then calf circumference is suitable to diagnose severe impairment.

Results of this study are limited by the small number of subjects, especially men. It will be necessary to confirm our results by testing a larger cohort. Another limitation is that the data were obtained only at one time and, therefore, we could not evaluate sensitivity of these tools for measuring sarcopenia progression.

References

1. Rosenberg IH. Epidemiologic and Methodologic Problems in Determining Nutritional-Status of Older Persons - Proceedings of a Conference Held in Albuquerque, New Mexico, October 19-21, 1988 - Summary Comments. Am J Clin Nutr 1989;50: 1231-1233
2. Cesari M, Fielding R, Benichou O, Bernabei R et al. Pharmacological Interventions in Frailty and Sarcopenia: Report by the International Conference on Frailty and Sarcopenia Research Task Force. J Frailty Aging 2015;4: 114-120
3. Bahat G, Saka B, Tufan F, Sivrikaya S et al. Prevalence of sarcopenia and its association with functional and nutritional status among male residents in a nursing home in Turkey. Aging Male 2010;13:211-214. doi: 10.3109/13685538.2010.489130
4. Landi F, Liperoti R, Fusco D, Mastropaolo S et al. Sarcopenia and mortality among older nursing home residents. J Am Med Dir Assoc 2012;13:121-126. doi: 10.1016/j.jamda.2011.07.004
5. Wen X, An P, Chen WC, Lv Y, Fu Q. Comparisons of sarcopenia prevalence based on different diagnostic criteria in chinese older adults. J Nutr Health Aging 2015;19: 342-347. doi: 10.1007/s12603-014-0561-x
6. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, Boirie Y et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age Ageing 2010;39:412-23. doi: 10.1093/ageing/afq034
7. de Ayala RJ (2008) The Theory and Practice of Item Response Theory. Guilford Publications, New York
8. Hambleton RK, Swaminathan H. Item Response Theory: Principles and Applications (Evaluation in Education and Human Services). Springer, New York, 1985
9. Reckase MD. Multidimensional Item Response Theory. Springer-Verlag, New York
10. Ham, 2009bleton RK. Fundamentals of Item Response Theory. SAGE Publications, London, 1991
11. Reise SP, Haviland MG . Item response theory and the measurement of clinical change. J Pers Assess. 2005;84:228-238
12. Van der Linden WJ, Hambleton RK. Handbook of Modern Item Response Theory. Springer, New York, 2013
13. Mokken RJ. A Theory and Procedure of Scale Analysis: With Applications in Political Research. De Gruyter, Berlin, 1971
14. Molenaar IW. Mokken scaling revisited. Kwantitatieve Methoden. 1982;3:145 - 164
15. Spector WD, Fleishman JA. Combining activities of daily living with instrumental activities of daily living to measure functional disability. J Gerontol B Psychol Sci Soc Sci. 1998;53:S46-57
16. McGrory S, Shenkin SD, Austin EJ, Starr JM. Lawton IADL scale in dementia: can item response theory make it more informative? Age Ageing. 2014;43:491-495. doi: 10.1093/ageing/aft173
17. Jette AM, Haley SM, Coster WJ, Kooyoomjian JT et al. Late life function and disability instrument: I. Development and evaluation of the disability component. J Gerontol A Biol Sci Med Sci. 2002;57:M209-216
18. Rolland Y, Lauwers-Cances V, Cournot M, Nourhashemi F et al. Sarcopenia, calf circumference, and physical function of elderly women: a cross-sectional study. J Am Geriatr Soc. 2003;51:1120-1124
19. Janssen I, Baumgartner RN, Ross R, Rosenberg IH et al. Skeletal muscle cutpoints associated with elevated physical disability risk in older men and women. Am J Epidemiol. 2004;159:413-421
20. Lauretani F, Russo CR, Bandinelli S, Bartali B et al. Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia. J Appl Physiol. 2003;95:1851-1860
21. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ et al. A short physical performance battery assessing lower extremity function: association with self-

JNHA: GERIATRIC SCIENCE

reported disability and prediction of mortality and nursing home admission. J Gerontol. 1994;49:M85-94

22. Van der Ark LA. Mokken scale analysis in R. J Stat Softw. 2007;20:1 - 19
23. Sijtsma K, Debets P, Molenaar IW. Mokken scale analysis for polychotomous items: theory, a computer program and an empirical application. Quality and Quantity. 1990;24:173-188
24. Stochl J, Jones PB, Croudace TJ. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. BMC Med Res Methodol. 2012;12:74. doi: 10.1186/1471-2288-12-74
25. Ligtvoet R, Van der Ark LA, Te Marvelde JM, Sijtsma K. Investigating an Invariant Item Ordering for Polytomously Scored Items. Educ Psychol Meas. 2010;70:578-595. doi: 10.1177/0013164409355697
26. Loevinger J. A Systematic Approach to the Construction and Evaluation of Tests of Ability. Psychol Monogr. 1947;61: i - 49
27. Štochl J. Nonparametric extension of item response theory models and its usefullness for assessment of dimensionality of motor tests. Acta Univ Carol Kin. 2006;42:1 - 19
28. Sijtsma K, Meijer RR, Van der Ark AL. Mokken scale analysis as time goes by: An update for scaling practitioners. Pers Individ Dif. 2011;50:31-37
29. Sijtsma K, Molenaar IW. Introduction to Nonparametric Item Response Theory. SAGE Publications, London, 2002
30. Sijtsma K, Junker BW. A survey of theory and methods of invariant item ordering. Br J Math Stat Psychol. 1996;49:79-105
31. Vellas B, Fielding R, Miller R, Rolland Y et al. Designing drug trials for sarcopenia in older adults with hip fracture – a task force from the international conference on frailty and sarcopenia research (ICFSR). J Frailty Aging 2014;3:199-204
32. McLean RR, Shardell MD, Alley DE, Cawthon PM et al. Criteria for clinically relevant weakness and low lean mass and their longitudinal association with incident mobility impairment and mortality: the foundation for the National Institutes of Health (FNIH) sarcopenia project. J Gerontol A Biol Sci Med Sci 2014;69: 576-583. doi: 10.1093/gerona/glu012
33. Sallinen J, Stenholm S, Rantanen T, Heliovaara M et al. Hand-grip strength cut points to screen older persons at risk for mobility limitation. J Am Geriatr Soc 2010;58: 1721-1726. doi: 10.1111/j.1532-5415.2010.03035.x
34. Teresi JA. Mini-Mental State Examination (MMSE): scaling the MMSE using item response theory (IRT). J Clin Epidemiol. 2007;60:256-259
35. Steffl M, Masek M, Petr M, Bunc V et al. Appropriateness of five measures proposed by EWGSOP for diagnosing sarcopenia in clinical practice among the elderly living at the senior centre in Blansko, Czech republic - a case study. J Aging Res Clin Practice. 2013;2:221 - 225
36. Huo YR, Suriyaarachchi P, Gomez F, Curcio CL et al. Comprehensive nutritional status in sarco-osteoporotic older fallers. J Nutr Health Aging. 2015;19:474-80. doi: 10.1007/s12603-014-0543-z