Measuring quality of life in sarcopenia

Evaluation of the clinimetric properties of the SarQoL® questionnaire and its short form

Anton Geerinck

Dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in Public Health Sciences

2021

Division of Public Health, Epidemiology and Health Economics – University of Liège

Measuring quality of life in sarcopenia :

Evaluation of the clinimetric properties of the SarQoL® questionnaire and its short form

Anton Geerinck

Dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in Public Health Sciences

Promotor: Pr. Olivier Bruyère (ULiège)

Co-promotor: Dr. Charlotte Beaudart (ULiège)

President of the Jury: Pr. Jean-Yves Reginster (ULiège)

Members of the Jury:

- Pr. Jean-Marie Degryse (Université Catholique de Louvain)
- Pr. Olivier Ethgen (ULiège)
- Pr. Francis Guillemin (Université de Lorraine)
- Pr. Sophie Gillain (ULiège)
- Pr. Mickaël Hiligsmann (Universiteit Maastricht)
- Pr. Christian Monseur (ULiège)

Division of Public Health, Epidemiology and Health Economics – University of Liège

World Health Organization Collaborating Centre for Public Health aspects of musculoskeletal health and aging

2021

LIÈGE université Médecine

PUBLIC HEALTH EPIDEMIOLOGY AND HEALTH ECONOMICS

WHO Collaborating Centre for Public Health aspects of musculo-skeletal health and ageing

This page intentionally left blank

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of a host of people, who all contributed in their own way. I would like to particularly express my gratitude and respect to the following people:

My promotor Pr. Olivier Bruyère, who entrusted me with this project 4 years ago and whose advice and energetic support have helped me tremendously throughout this project.

My co-promotor Dr. Charlotte Beaudart, whose knowledge and responsiveness I deeply admire and appreciate.

Pr. Jean-Yves Reginster, for his kind support and warm welcome into the Division of Public Health.

The people who made up my thesis committee, and who donated their time and counsel to make this project a success: Pr. Reginster, Pr. Christian Monseur, Pr. Olivier Ethgen, Pr. Sophie Gillain and Pr. Mickaël Hiligsmann.

The external jury members, Pr. Jean-Marie Degryse and Pr. Francis Guillemin, who graciously accepted to be a part of the jury for this dissertation.

A special thank you to all the collaborators in the SarQoL project, too many to name, without whom this project would not have been possible.

Fabienne Damblon and Marie Zumbihl, for their unparalleled kindness and helpfulness.

My former and current colleagues: Germain Honvo, Alexia Charles, Médéa Locquet, Victoria Leclercq, Laetitia Lengelé, Justine Slomian and Fanny Buckinx.

My parents, André Geerinck and Hilde Wouters and my brothers Willem and Pieter, as well as my sister Johanna.

My partner, Constance Remacle, as well as her parents, Michelle and Christian Wilmotte-Remacle.

This page intentionally left blank

SUMMARY

Sarcopenia, the age-related loss of muscle function, has become a topic of great interest in ageing research, and considerable advances have been made in the last decade on its definition and diagnostic criteria. The evidence of its effects on health outcomes has greatly increased, including what we know about its influence on quality of life.

In 2015 the SarQoL questionnaire, a questionnaire specifically developed to measure quality of life in sarcopenia, was published. It measures health-related quality of life through 55 items categorized into 7 domains of dysfunction covering physical and mental health, locomotion, body composition, functionality, activities of daily living, leisure activities, and fears. It produces domain scores for each of the 7 domains, as well as an overall quality of life score over the entire questionnaire, all of which are scored from 20 (worst) to 100 (best). The first investigations of its measurement properties were carried out shortly after its release in a cohort of older, community-dwelling people in Belgium as well as a second cohort in England, and the questionnaire proved to be valid and reliable in both contexts.

This dissertation builds upon the work already performed on and with the SarQoL questionnaire and aimed to further improve the measurement of quality of life in sarcopenia by (1) investigating the measurement properties of the SarQoL questionnaire and (2) creating a shorter version with lower response burden.

Within the framework of the first objective of this dissertation we investigated the ability of the SarQoL questionnaire to detect changes in quality of life over time within a sample of 42 sarcopenic people followed over a 2-year period. We found that the questionnaire was responsive when looking at hypotheses on the strength of correlation between changes measured by the SarQoL and by 2 other quality of life questionnaires, the SF-36 and the EQ-5D. We also evaluated standardized response means, a measure of effect size, and found that the SarQoL questionnaire measured greater change than the SF-36 and EQ-5D/VAS. These first results show that the SarQoL questionnaire can be used to track the evolution of quality of life over time, and that, in older people affected by a reduction in muscle strength and/or function, it performs better than generic questionnaires.

We continued by calculating the standard error of measurement and the smallest detectable change of the SarQoL questionnaire in a sample of 278 sarcopenic participants recruited by 9 studies in multiple countries. When all participants were analyzed together, we found a standard error of 2.65 points and that an individual would need to change at least 7.35 points on the overall SarQoL score to be sure that a real change has occurred. This value can be interpreted as the trigger for further evaluation in clinical practice or to define patients as “responders” in interventional trial settings.

The third aspect of this dissertation was conceived in response to the publication of the revised consensus criteria for the diagnosis of sarcopenia by the European Working Group on Sarcopenia in Older People (EWGSOP2). We verified whether the SarQoL questionnaire retained its capacity to discriminate between sarcopenic and non-sarcopenic older people with these new diagnostic criteria and were able to confirm this when we found significantly lower scores for all 7 domains and the overall SarQoL score within a sample of 296 older people. Together with previous publications, it has become clear that the SarQoL questionnaire can discriminate between patients with musculoskeletal impairments affecting quality of life independent of the specific diagnostic criteria used, extending its usefulness outside of studies using the European consensus criteria.

Next, we looked beyond the application of the SarQoL questionnaire in sarcopenia and investigated the applicability and measurement properties of the questionnaire when used to measure health-related quality of life in physical frailty diagnosed with the Fried Criteria. We found that the overall SarQoL score could discriminate between frail, pre-frail, and non-frail groups in our sample of 382 participants. Internal consistency (alpha = 0.89) and test-retest reliability (ICC = 0.92) were good, and construct validity was confirmed. We found moderate responsiveness to change over time through hypothesis testing and a large effect size for the Overall QoL score of the SarQoL questionnaire. No existing questionnaire specifically designed to capture frailty-related quality of life is available, but this investigation showed that the SarQoL questionnaire is capable of providing a valid and reliable measure of musculoskeletal quality of life in this population.

For the final investigation within the scope of the first objective of this dissertation we explored the diagnostic performance of the SarQoL questionnaire as a screening instrument for sarcopenia diagnosed with the EWGSOP2 criteria. We analyzed a sample of 309 people and found an area-under-the-ROC-curve of 0.771 for the overall SarQoL score, indicating that there is a use for the questionnaire as a screening instrument. At the optimal threshold of ≤52.4 points for the overall SarQoL score, the sensitivity of the questionnaire was 64.7% and the specificity was 80.5%. This investigation showed that the SarQoL questionnaire could perform an additional function under certain circumstances, namely to identify people that are good candidates for further musculoskeletal tests to determine loss of muscle function.

Within the framework of the second objective of this dissertation we reduced the number of items in the SarQoL questionnaire from 55 to 14, selected from 6 out of the 7 domains present in the SarQoL. We used information collected though a Delphi method with experts and the calculation of item-impact scores for participants, as well as evidence on the questionnaire's measurement properties to allow an expert group to make decisions on which items to include in the short form. This new SF-SarQoL was then validated in a sample of 214 older people, and we found good discriminative power between sarcopenic and non-sarcopenic participants, high internal consistency, excellent test-retest reliability

both for the overall score as well as for the individual items (although a small systematic bias was present), high criterion construct validity as well as evidence for construct validity through hypotheses on the correlation between the SF-SarQoL and the EQ-5D/VAS. We were unable to confirm a one-factor model though confirmatory factor analysis but found instead two factors. Overall, this first evaluation of the SarQoL questionnaire indicates that it is a promising instrument, but its structural validity and test-retest validity should be further investigated. This short-form version of the SarQoL questionnaire could remove some of the obstacles that have hindered the adoption of the SarQoL questionnaire and still provide a valid and reliable measurement of overall quality of life, but sacrifices the detail provide by the domain scores of the full-length SarQoL questionnaire.

We finished this dissertation by carrying out a study on the relative importance of the 14 items in the SF-SarQoL questionnaire as judged by older people themselves. We used the best-worst scaling technique, a choice-experiment design, to solicit patient preferences from 163 participants, who indicated that they found the items “feeling a reduction of physical capacity” and “balance problems” to be the most important in light of their impact of quality of life, and the items “feeling a reduction in muscle mass” and “having difficulty carrying heavy objects” to be the least important. This is the first time a ranking of aspects of quality of life in sarcopenia was established, and clearly shows that not all aspects of quality of life are equally important.

In conclusion, this dissertation, together with validation studies performed in a number of countries, has demonstrated that the SarQoL questionnaire is a suitable tool for the measurement of quality of life in sarcopenia and has provided researchers with the information needed to argue for its inclusion in clinical trials. We have also shown that its usefulness is not limited to its primary function, measuring quality of life in sarcopenia, but that it can be used in physical frailty or as a screening tool. We created a shorter version of the SarQoL questionnaire and demonstrated that it possesses adequate measurement properties. Because of its shorter length, it is easier to administer and places a reduced cognitive burden upon the respondents. In trials where quality of life is not a primary outcome, it may possess the necessary balance between precision and burden that would make it the right tool for the job.

RESUME

La sarcopénie, la perte de la fonction musculaire liée à l'âge, est devenue un sujet de grand intérêt dans la recherche sur le vieillissement. De grandes avancées ont été réalisées au cours de la dernière décennie sur sa définition et ses critères de diagnostic, ainsi que sur son effet sur les résultats de santé, y compris sur la qualité de vie.

En 2015, le questionnaire SarQoL, développé spécifiquement pour mesurer la qualité de vie dans la sarcopénie, a été publié. Il mesure la qualité de vie liée à la santé à travers 55 items catégorisés en 7 domaines de dysfonctionnement couvrant la santé physique et mentale, la locomotion, la composition corporelle, la fonctionnalité, les activités de la vie quotidienne, les loisirs, et les peurs. Pour chacun des 7 domaines, le questionnaire fournit un score allant de 20 (le pire) à 100 (le meilleur), ainsi qu'un score global de qualité de vie sur l'ensemble du questionnaire. Les propriétés de mesure du questionnaire ont ensuite été étudiées sur une cohorte de personnes âgées vivant dans la communauté en Belgique et sur une deuxième cohorte en Angleterre. Le questionnaire s'est avéré valide et fiable dans les deux contextes.

Cette thèse s'appuie sur le travail déjà effectué dans le questionnaire SarQoL et vise à améliorer la mesure de la qualité de vie dans la sarcopénie, premièrement en étudiant ses différentes propriétés de mesure, et, deuxièmement, en créant une version plus courte avec un plus faible fardeau de réponse.

Dans le cadre du premier objectif de cette thèse, nous avons étudié la capacité du questionnaire SarQoL à détecter un changement de la qualité de vie à travers le temps, au sein d'un échantillon de 42 personnes sarcopéniques suivies sur une période de 2 ans. Nous avons constaté que le questionnaire était sensible lors de l'examen des hypothèses sur la force de la corrélation entre les changements mesurés par le SarQoL et par deux autres questionnaires de qualité de vie, le SF-36 et l'EQ-5D/VAS. Nous avons également évalué la sensibilité au changement en employant des réponses moyennes normalisées, un indicateur de l'ampleur de l'effet, et nous avons constaté que le SarQoL mesurait des changements plus importants que le SF-36 et l'EQ-5D/VAS. Ces premiers résultats montrent que le questionnaire SarQoL peut être utilisé pour suivre l'évolution de la qualité de vie à travers le temps, et qu'il est plus performant, chez les personnes âgées touchées par une réduction de la force et/ou la fonction musculaire, que les questionnaires génériques.

Nous avons poursuivi en calculant l'erreur type de mesure et le plus petit changement détectable du questionnaire SarQoL dans un échantillon de 278 participants sarcopéniques recrutés par 9 études dans plusieurs pays. Lorsque tous les participants ont été analysés ensemble, nous avons trouvé que l'erreur de mesure était de 2.65 points et qu'un individu devrait changer au moins 7.35 points sur le score global

du SarQoL pour être sûr qu'un réel changement ait eu lieu. Cette valeur peut être interprétée comme l'élément déclencheur pour une évaluation plus approfondie dans la pratique clinique ou pour définir les patients comme des « répondeurs » dans le cadre des essais interventionnels.

Le troisième aspect de cette thèse a été conçu en réponse à la publication des critères révisés pour le diagnostic de la sarcopénie par le Groupe de Travail Européen sur la Sarcopénie chez les Personnes Agées (EWGSOP2). Nous avons vérifié que le questionnaire SarQoL conservait sa capacité de discrimination entre les personnes sarcopéniques et non sarcopéniques avec ces nouveaux critères diagnostiques, et nous avons pu le confirmer après avoir trouvé des scores significativement plus faibles pour les 7 domaines et le score global du SarQoL au sein d'un échantillon de 296 personnes âgées. Avec les publications précédentes, il est devenu clair que le questionnaire SarQoL peut discriminer les patients présentant des troubles musculo-squelettiques affectant la qualité de vie indépendamment des critères de diagnostic spécifiques utilisés, étendant son utilité en dehors des études utilisant les critères de consensus européens.

Ensuite, nous sommes allés au-delà de l'application du questionnaire SarQoL dans la sarcopénie, et nous avons étudié l'applicabilité et les propriétés de mesure du questionnaire lorsqu'il est utilisé pour mesurer la qualité de vie liée à la santé chez les personnes physiquement fragiles selon les critères de Fried. Nous avons constaté que le score global SarQoL pouvait distinguer les groupes fragiles, pré-fragiles et non-fragiles dans notre échantillon de 382 participants. La cohérence interne (alpha = 0.89) et la fiabilité test-retest (ICC = 0.92) étaient bonnes, et la validité de construit a été confirmée. Nous avons constaté une sensibilité au changement dans le temps modérée par le biais de tests d'hypothèses et une grande taille d'effet pour le score global de qualité de vie du questionnaire SarQoL. Il n'existe aucun questionnaire spécifiquement conçu pour capturer la qualité de vie liée à la fragilité, mais cette enquête a montré que le questionnaire SarQoL est capable de fournir une mesure valide et fiable de la qualité de vie musculo-squelettique dans cette population.

Pour la dernière investigation dans le cadre du premier objectif de cette thèse, nous avons exploré la performance diagnostique du questionnaire SarQoL comme instrument de dépistage de la sarcopénie diagnostiquée avec les critères EWGSOP2. Nous avons analysé un échantillon de 309 personnes et nous avons trouvé une aire sous la courbe ROC de 0.771 pour le score global du SarQoL, ce qui indique que le questionnaire a une utilité en tant qu'instrument de dépistage. Au seuil optimal de $\leq$52.4 points pour le score global du SarQoL, la sensibilité du questionnaire était de 64.7% et la spécificité de 80.5%. Cette enquête a montré que le questionnaire SarQoL pouvait fournir des fonctions supplémentaires dans certaines circonstances, en particulier pour signaler les personnes qui sont de bons candidats pour d'autres tests musculo-squelettiques afin de déterminer la perte de la fonction musculaire.

Dans le cadre du deuxième objectif de cette thèse, nous avons réduit le nombre d'items du questionnaire SarQoL de 55 à 14, sélectionnés dans 6 des 7 domaines présents dans le SarQoL. Nous avons utilisé une méthode Delphi avec des experts, des scores d'impact d'item fournis par des personnes âgées, et de la documentation sur les propriétés de mesure pour permettre à un groupe d'experts de prendre une décision sur les items à inclure dans la version courte du questionnaire. Ce nouveau SF-SarQoL a ensuite été validé auprès d'un échantillon de 214 personnes âgées, et nous avons constaté une bonne discrimination entre les participants sarcopéniques et non-sarcopéniques, une cohérence interne élevée, une excellente fiabilité test-retest tant pour le score global que pour les items individuels (bien qu'un petit biais systématique soit présent), une validité de critère élevée ainsi que des preuves de validité de construit à travers des hypothèses sur la corrélation attendue entre le SF-SarQoL et l'EQ-5D/VAS. Nous n'avons pas pu confirmer un modèle à un facteur par une analyse factorielle confirmatoire, mais nous avons trouvé deux facteurs. Dans l'ensemble, cette première évaluation du questionnaire SF-SarQoL indique qu'il s'agit d'un instrument prometteur, mais sa validité structurelle et sa validité test-retest devraient être étudiées plus en profondeur. Cette version abrégée du questionnaire SarQoL pourrait supprimer certains des obstacles qui ont entravé l'adoption du questionnaire SarQoL, tout en continuant à fournir une mesure valide et fiable de la qualité de vie globale, toutefois, il sacrifierait les détails fournis par les scores de domaine de l'ensemble du long questionnaire SarQoL.

Nous avons terminé cette thèse en réalisant une étude sur l'importance relative des 14 items du questionnaire SF-SarQoL telle que jugée par les personnes âgées elles-mêmes. Nous avons utilisé la technique du *best-worst scaling*, une méthode d'expérimentation des choix, pour solliciter les préférences de 163 participants, qui ont indiqué qu'ils trouvaient les items « sentir une réduction de la capacité physique » et « avoir des problèmes d'équilibre » les plus importants au regard de leur impact sur la qualité de vie, et les items « sentir une réduction de la masse musculaire » et « avoir des difficultés à porter des objets lourds » les moins importants. C'est la première fois qu'un classement des aspects de la qualité de vie dans la sarcopénie est établi, et montre clairement que tous les aspects de la qualité de vie n'ont pas la même importance.

En conclusion, cette thèse, ainsi que des études de validation réalisées dans un certain nombre de pays, a démontré que le questionnaire SarQoL est un outil approprié pour la mesure de la qualité de vie dans la sarcopénie et a fourni aux chercheurs les informations nécessaires pour plaider en faveur de son inclusion dans la sarcopénie des essais cliniques. Nous avons également montré que son utilité ne se limite pas à sa fonction première, mesurer la qualité de vie dans la sarcopénie, mais qu'elle peut être utilisée dans la fragilité physique ou comme outil de dépistage. Nous avons créé une version plus courte du questionnaire SarQoL et démontré qu'il possède des propriétés de mesure adéquates. En raison de sa taille, il est plus facile à administrer et impose une moindre charge cognitive aux répondants. Dans les essais où la qualité de vie n'est pas un critère de jugement principal, il peut posséder l'équilibre nécessaire entre précision et charge qui en ferait le bon outil pour le travail.

TABLE OF CONTENTS

Introduction

Introduction

Throughout history, and certainly during the last few hundred years, life expectancy has steadily increased. Where a baby born in Belgium in 1900 was projected, on average, to live to the age of 46.5 years old, a baby born in 2015 is now projected to reach the age of 81 years. Even when child mortality is removed from the equation, a child that reached 10 years of age in 1900 was projected to have an average life expectancy of 62.5 years, compared to 81.7 years in 2015. Healthy life expectancy has also gone up, from 66.3 to 70.1 years over the same period, and years lived with disability increased by 12%, from 9.68 to 10.80 years [1]. It is estimated that the proportion of the global population aged over 60 years old will rise to 22% in 2050, representing 2 billion older people [2]. In reaction to these demographic trends, the promotion of healthy aging has become an important goal of many health systems, and the Assembly of the World Health Organization has sent a strong signal about the challenges created by the ageing of populations by declaring the period from 2021 to 2030 the "Decade of Healthy Ageing" [3].

As part of the increasing attention allocated to healthy ageing and age-related health conditions, there has been a steady increase in research activities on sarcopenia, loosely characterized as age-related loss of muscle mass and function, in the last 2 decades. Multiple initiatives have focused on its definition, diagnostic criteria, association with health outcomes, treatment, and prevention, culminating in the recognition of sarcopenia as a disease within the International Classification of Diseases and Related Health Problems with the code ICD-10-CM (M62.84) in 2016 [4].

Sarcopenia

Historical perspective

The history of sarcopenia as a research subject started in 1989, during a meeting in New Mexico on the epidemiology of ageing, where Irwin Rosenberg gave the name "sarcopenia" to the phenomenon of loss of muscle strength with ageing [5]. He was not the first to observe and ponder this phenomenon though: as early as 1931, MacDonald Critchley wrote that *"the entire musculature tends with advancing years to undergo involutional changes which are manifested as wasting"* [6]. What Rosenberg accomplished by labelling sarcopenia, and the reason why he is rightfully considered one of the founding fathers of this research domain, was to kickstart interest for this condition. Within a year of the publication of the summary notes of the aforementioned meeting, a call for research proposals on sarcopenia was launched by the National Institute of Health in the USA, with workshops and dedicated research programs following soon after [7].

The first articles using the word sarcopenia indexed in the MEDLINE database were published in 1993, and in the next 15 years the research output has steadily increased for a total of 610 publications

by the end of 2007. In total, 12,839 articles have been published on sarcopenia before the end of the year 2020. While this number is still relatively modest compared to other disorders associated with ageing (for example: there are 93,062 articles in the MEDLINE database on osteoporosis), it is clear that the last decade has seen an exponential growth in sarcopenia research.

Diagnostic criteria

From the start, one of the main issues researchers wrestled with was how to define and diagnose sarcopenia. The first propositions, formulated by Baumgartner, Janssens, and Delmonico, relied on a single indicator to draw a line between sarcopenic and non-sarcopenic persons, relying solely on muscle mass to distinguish between the two [8–10]. These definitions were tested in different cohorts of older people and served as the basis upon which an evolution of the conceptual model of sarcopenia was built. The next propositions characterized sarcopenia not only by muscle mass, but also through muscle strength and physical performance [11]. From 2010 until recently, a number of international organizations produced different sets of diagnostic criteria based on these 3 indicators of sarcopenia, notably the Special Interest Group on Sarcopenia of the European Society of Nutrition (ESPEN-SIG – 2010 [12]), The European Working Group on Sarcopenia in Older People (EWGSOP – 2010 & 2018 [13, 14]), the International Working Group on Sarcopenia (IWGS – 2011 [15]), the Society for Sarcopenia, Cachexia and Wasting Disorders (SSCWD – 2011 [16]), The Foundation for the National Institutes of Health (FNIH – 2014 [17]), the Asian Working Group on Sarcopenia (AWGS – 2014 & 2020 [18, 19]) and the Sarcopenia Definition and Outcomes Consortium (SDOC – 2020 [20]). Currently, the most widely supported definition in western populations is the one formulated by the second European Working Group on Sarcopenia in Older People (EWGSOP2), which is a revision of the definition put forward by the same working group in 2010 [13, 14]. The definition of sarcopenia proposed by the EWGSOP2 is that it is *"a progressive and generalised skeletal muscle disorder that is associated with increased likelihood of adverse outcomes including falls, fractures, physical disability and mortality"* [14]. It considers sarcopenia to be present if a person has low muscle strength, evaluated by measuring handgrip strength or with the chair stand test, as well as low muscle mass, which is evaluated by dual-energy x-ray absorptiometry (DXA) or bioelectrical impedance analysis (BIA). It uses a third criterium, physical performance, to evaluate the severity of sarcopenia present [14].

Aetiology

The EWGSOP2 makes the distinction between primary and secondary sarcopenia, where the first is considered to be related to the ageing process, while other causal factors (i.e., disease, inactivity, or malnutrition) are present for the second [14]. In primary or age-related sarcopenia, the decline of

muscle quantity (1 to 2% per year from the age of 50 years onward) and the loss of strength (1.5 to 5% per year) cross a critical threshold where adverse outcomes such as functional impairment and disability become much more likely [13, 14, 21]. These changes in skeletal muscle quantity and quality are thought to be a result of the loss of neurons with age, changes in the production of several hormones (growth hormone, insulin-like growth factor-I (IGF-1), corticosteroids, androgens, estrogens and insulin) affecting muscle protein metabolism, an increase in inflammatory factors leading to a semi-permanent state of low-grade inflammation, and the infiltration of adipose tissue into the muscle [13, 22]. However, the exact etiological process that leads to sarcopenia is still under investigation, and complete picture of the processes behind sarcopenia and their interactions is not yet available [23].

Prevalence

The long period in which multiple definitions were in use without a consensus on the most appropriate one have complicated attempts to estimate its prevalence. In 2014, prevalence of sarcopenia according to the EWGSOP1 criteria was estimated to be between 1-29% for older adults in the community and between 14 and 33% for those in long-term care institutions [24]. By 2017, in a meta-analysis of 35 studies diagnosing sarcopenia with the EWGSOP, AWGS or IWGS criteria, this range was reduced to 10% (95% CI: 8-12%) for men and 10% (95% CI: 8-13%) for women [25]. To date, no meta-analysis has quantified the prevalence of sarcopenia using the EWGSOP2 criteria, but several studies indicate that, in a paired-sample design, lower prevalence rates are found with the EWGSOP2 criteria compared to EWGSOP1 criteria [26].

Health and economic outcomes

Sarcopenia has been demonstrated to negatively affect several health outcomes, creating a burden on the patient. Veronese et al. assembled the available literature into an umbrella review in 2019 and found 6 meta-analyses reporting on 14 different outcomes. They found a significant association, backed up by evidence rated as highly suggestive, between sarcopenia and mortality in community-dwelling people, falls in community and nursing home settings, and disability in community-dwelling people. Less information was available for the other outcomes, leading the authors to consider the evidence as weak. Nonetheless, further significant associations were found for mortality in nursing home settings, fragility fractures in the community, hospitalization and length-of-stay for people living at home [27].

Because of its impact on these health outcomes, sarcopenia also puts a burden on healthcare systems and society in general. A systematic review performed by Bruyère et al. in 2019 included 14 studies that had investigated healthcare costs in sarcopenia, and while they cautioned against over-

interpreting their results due to the heterogeneity between the included studies, they did report that for the majority of studies the healthcare expenditure was significantly higher for sarcopenic patients compared to non-sarcopenic patients [28]. No studies have specifically looked at healthcare cost associated with sarcopenia in Belgium, but a study conducted in the Netherlands, Belgium's neighbour to the north, provides the closest approximation. This study evaluated sarcopenia with the EWGSOP criteria and found that the healthcare costs for the 53 sarcopenic participants were about 3 times higher than the costs for the 174 non-sarcopenic participants (4,325 € versus 1,533 € per 3 months), which amounts to an annual extra cost of 11,168 € per sarcopenic person [29].

Treatment

Interventions to treat or prevent sarcopenia have focused on three areas: physical exercise, nutrition and pharmacological treatment, or a combination of these three approaches (for example: exercise and protein supplementation). There is high-quality evidence for the efficacy of high-volume and high-intensity resistance training in improving muscle mass, strength, and physical performance in older people thanks to an umbrella review on the subject [30]. The level of certainty concerning protein supplementation is lower, but because of the importance of adequate protein intake, clinicians are recommended to strongly consider this [31]. The effectiveness of protein supplementation depends on the baseline level, quantity and quality of the supplement, and the timing and duration of the administration [32]. The evidence for other interventions, such as vitamin D supplementation, or pharmacologic interventions (anabolic hormones, growth hormone, growth hormone-releasing hormone, combined testosterone-growth hormone, insulin-like growth factor-1, pioglitazone, testosterone, and angiotensin-converting enzyme inhibitors) is not yet sufficient to recommend their use although positive effects have been found in clinical trials [33–35].

Quality of life

Quality of life as a concept

Quality of life is an illusive term that molds itself to the context in which it is applied. Quality of life from the perspective of an architect or a city planner is different to a health perspective, which is why the definition of quality of life proposed by the World Health Organization Quality of Life Group is deliberately broad in stating that it is "*the individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals*" [36]. Because of the deliberate vagueness of this definition, the concept of health-related Quality of Life (HRQoL) was

formulated for health applications. Multiple authors have elaborated conceptual models for health-related quality of life, of which the Wilson & Cleary model, the Ferrans et al. model and the WHO International Classification of Functioning, Disability and Health model are most often used [37]. Of these three, the Ferrans et al. model (which is an evolution of the Wilson and Cleary model), is most interesting in relation to the SarQoL questionnaire. This model presents health-related quality of life as a function of five patient outcomes: biological function, symptoms, functional status, general health perception and overall quality of life. The authors indicate that causal relationships exist between the outcomes, where biological function acts on symptoms, which in turn acts on functional status and so on. The authors define two external actors, namely the characteristics of the individual and the characteristics of the environment that interact with each of the five patient outcomes [38]. A qualitative review from 2019 has also brought important information forward. The authors explored previous qualitative studies that looked specifically at what quality of life means to older community-dwelling people, and found 9 domains: health perception, autonomy, role and activity, relationships, attitude and adaptation, emotional comfort, spirituality, home and neighbourhood, and financial security. The authors also remark that the distinction into 9 domains is somewhat artificial because of the strong and dynamic connections and interactions between the different aspects of quality of life [39].

Quality of life in sarcopenia

While much of the research efforts in sarcopenia have focused on its diagnosis, reference values for muscle parameters and physical performance tests, and outcomes, there is a growing body of work on quality of life in older people with sarcopenia. Woo et al summarized the available evidence up to 2016 looking at the association between muscle strength, muscle mass and physical performance and their individual association with quality of life. They found a significant relationship between quality of life, muscle strength and physical performance, but not with muscle mass, in cross-sectional studies. The single study that defined sarcopenia with the EWGSOP criteria in this review article reported poorer quality of life for both genders linked to sarcopenia [40]. Since then, more studies have reported results: Silva Neto et al included 70 older people from the Quilombola ethnic group in Brazil and found a negative association between quality of life measured with the SF-36 instrument and sarcopenia (EWGSOP criteria) [41]. Also in Brazil, Marques et al. looked at 584 older, community-dwelling adults and found a significant association between quality of life measured with the CASP-16 instrument and sarcopenia in men, but not women (sarcopenia defined as muscle mass below 2 SD from young reference population) [42]. Yalcin et al focused on 241 nursing home residents in Turkey and found lower scores for all subscales of the SF-36 in the sarcopenic group (EWGSOP criteria) [43]. Lastly, Manrique-Espinoza used data from 543 community-dwelling Mexicans aged 70 years or older to demonstrate that those with severe sarcopenia (EWGSOP criteria) had significantly lower scores for

both the mental and physical component summary score of the SF-36 questionnaire [44]. While the majority of the available information points to lower quality of life associated with sarcopenia, it is important to highlight that the most widely used instruments to measure quality of life in sarcopenia are not specifically designed or validated for this target population and that the true impact of sarcopenia may be underestimated.

The Sarcopenia Quality of Life questionnaire

Conscious of the gap created by the absence of a quality-of-life questionnaire specifically designed for sarcopenia, the Sarcopenia Quality of Life (SarQoL®) questionnaire was developed within an international collaboration between researchers from the University of Liège, Geneva University Hospitals, CHU Toulouse and the Free University of Brussels and released in 2015 [45]. The authors generated an initial pool of 180 items from literature review, face-to-face interviews with sarcopenic individuals and experts' opinion, which was ultimately reduced to 55 items. These are categorized into 7 domains of health-related dysfunction: physical and mental health, locomotion, body composition, functionality, activities of daily living, leisure activities, and fears. The SarQoL® questionnaire is auto-administered and produces 7 domain scores and 1 overall quality of life score, which range between 20 and 100 points, with higher scores indicative of better quality of life [45]. Figure 1 on page 9 shows the 55 items and the domains into which they are categorized, while the SarQoL questionnaire itself can be found in annex.

After its development, a concerted effort was undertaken to examine its measurement properties in different populations, and to translate the questionnaire in as much languages as possible. To date, the questionnaire is available in 33 different languages from the website www.sarqol.org, and 13 articles have been published on its measurement properties (without the articles in this dissertation). Table 1 on page 11 lists the main results from these studies.

Measurement properties

Measurement properties of health measurement instruments

With regards to the measurement properties of health measurement instruments, and patient-reported outcome measures for health outcomes in particular, a Dutch initiative named COSMIN (which stands for COnsensus-based Standards for the selection of health Measurement INstruments) has been the most influential in setting standards, and they provided an international consensus on the terminology and definitions of different measurement properties [46]. In general, the measurement

properties of a questionnaire (sometimes called clinimetric properties) can be categorized into 3 large domains: its reliability, which describes the degree to which the questionnaire is free from measurement error; its validity, which looks at the degree to which the questionnaire measures what it is supposed to measure; and its responsiveness, which reflects on the capacity of the questionnaire to detect change over time [46]. Within the domains of reliability and validity, there are more specific properties. The reliability of a patient-reported outcome measure such as the SarQoL questionnaire can be judged on its internal consistency ("how interrelated are the items that make up the questionnaire?"), its reliability in a test-retest setting ("does the questionnaire provide the same score if the health status of the respondent remains stable?") or within or between raters if interviewer-administered, and its measurement error (that is, the systematic and random error not attributable to actual change in the construct being measured) [46]. The validity domain for patient-reported outcome measures is further divided into content validity ("is the questionnaire an adequate reflection of the construct being measured?"), construct validity (further subdivided into structural validity, hypotheses testing and cross-cultural validity) and criterion validity if there is a gold standard instrument with which you can compare the patient-reported outcome measure under investigation [46].

Measurement properties of the SarQoL questionnaire

A look at the measurement properties of the SarQoL questionnaire detailed in table 1, reveals a number of results for the domain reliability. Multiple studies reported on internal consistency through Cronbach's alpha values, the lowest of which is 0.87, indicating a high degree of interrelatedness between the items. There are also several studies that evaluated test-retest reliability with the intraclass correlation coefficient, all of which were larger than 0.9 for the overall QoL score for the SarQoL questionnaire, indicating a high degree of stability between the 2 administrations. The second domain, validity, was evaluated through known-groups validity (also know as discriminative power) where the hypothesis that sarcopenic participants should have lower scores on the SarQoL questionnaire is investigated, and by the testing of hypotheses on the strength of correlation between (domains of) the SarQoL questionnaire and (domains of) other quality-of-life-questionnaires that either measure similar/convergent constructs (where moderate to strong correlations are expected to be found) or different/divergent constructs (low or no correlation). The convergent hypotheses have mostly been confirmed, but the results for the divergent hypotheses have been mixed.

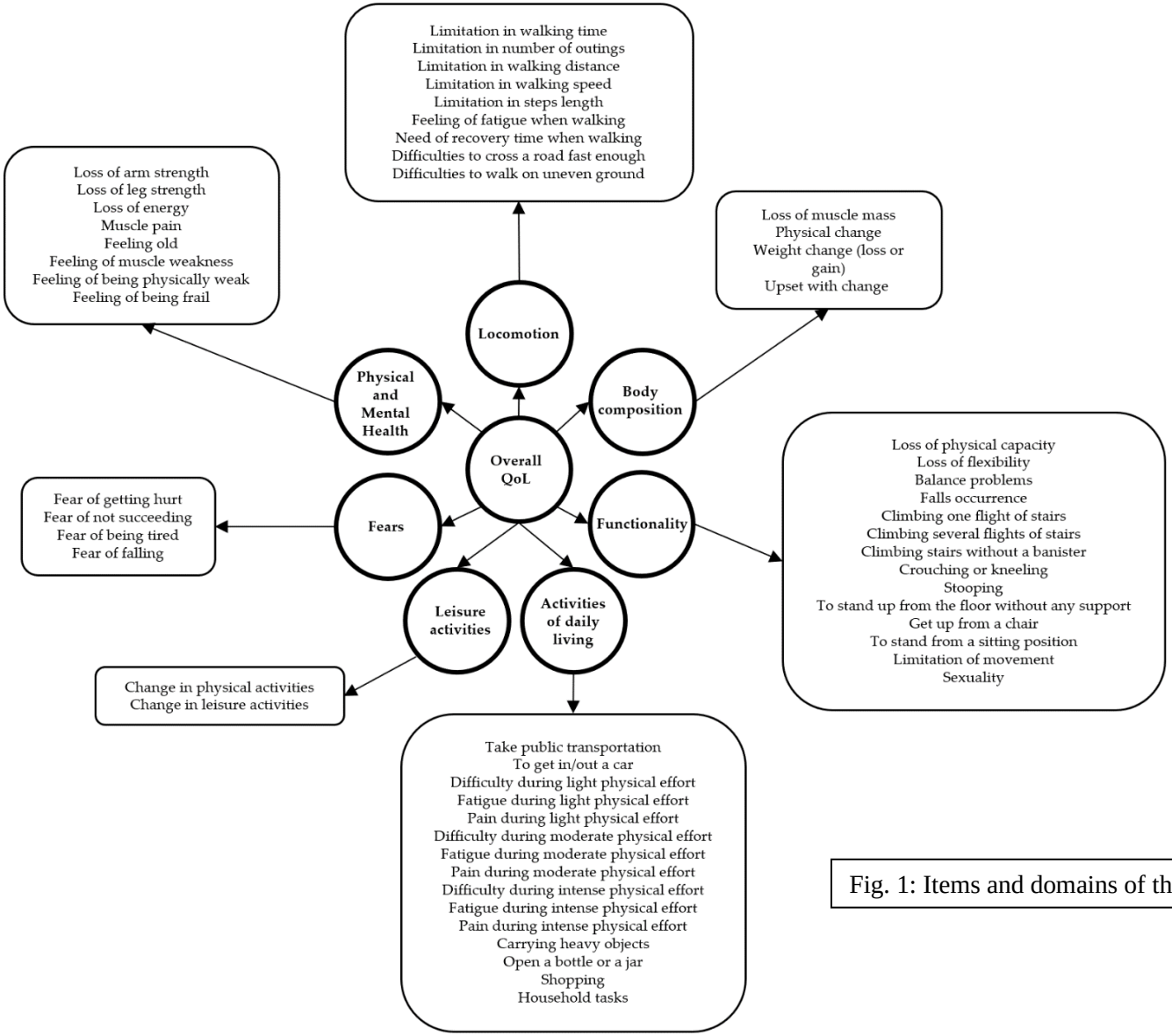


Fig. 1: Items and domains of the SarQoL questionnaire

There are a number of measurement properties not presented in table 1, because they have not yet been investigated. There is no information in the validation studies on the measurement error of the questionnaire, although there are a number of studies that evaluated test-retest reliability, so the necessary data to calculate the standard error of measurement is present. There is also no data on the responsiveness of the questionnaire, for which either a longitudinal study over several years or an interventional study would have to be designed. Lastly, there is no study reporting on the structural validity of the questionnaire through an exploratory or confirmatory factor analysis, leaving the current categorization of the items in their respective domains unconfirmed by a quantitative analysis.

Arguments for a shorter SarQoL questionnaire

The ideal patient-reported outcome measure would be a short, easy-to-complete, cheap to administer and clinimetrically sound questionnaire. While there is growing evidence for the measurement properties of the SarQoL questionnaire, and the cost per administration is low thanks to it being auto-administered and paper-based, the questionnaire cannot be called short. Because of its relative length, there is also an increased cognitive burden upon the respondents, which can translate into a higher rate of missing responses towards the end of the questionnaire, or other undesirable response patterns [47]. These considerations may hinder the inclusion of the SarQoL questionnaire under certain circumstances and provide an argument for the creation of a shorter version.

Table 1: Results from studies on the measurement properties of the SarQoL questionnaire

Version and year of publication	Reference	Sample size (n)	Sarcopenic subjects [n(%)]	Sarcopenia diagnosis	Discriminative power (overall SarQoL score)	Internal consistency (α)	Construct validity	Test-retest reliability (ICC)	Floor and ceiling effects
French (2016)	[48]	296	43 (14.5%)	EWGSOP	S: 54.7 (45.9-66.3) NS: 67.8 (57.3-79.0) p<0.001	0.87	Convergent and divergent hypotheses confirmed	0.91	absent
English (2016)	[49]	297	14 (4.7%)	EWGSOP	S: 61.9 ± 16.5 NS: 71.3 ± 12.8 p=0.01	0.88	Convergent and divergent validity confirmed	0.95 (0.92-0.97)	absent
Romanian (2017)	[50]	100	13 (13%)	EWGSOP	S: 57.3 (34.4-70.7) NS: 68.4 (55.7-85.2) p=0.018	0.946	Convergent hypotheses confirmed; divergent hypotheses not confirmed		absent
Dutch (2018)	[51]	92	30 (32.6%)	EWGSOP	S: 67.15 (54.75-81.52) NS: 79.72 (70.10-86.88) p=0.003	0.883	6/8 hypotheses confirmed	0.976 (0.947-0.989)	Absent
Polish (2018)	[52]	106	60 (56.6%)	EWGSOP	S: 54.9 ± 16.5 NS: 63.3 ± 17.1 p=0.013	0.92	Convergent validity confirmed	0.99 (0.995-0.999)	absent
Greek (2018)	[53]	176	50 (28.4%)	EWGSOP	S: 52.12 ± 11.04 NS: 68.23 ± 14.1 p<0.001	0.96	Mixed results for convergent and divergent hypotheses	0.96 (0.95-0.97)	absent
Lithuanian (2019)	[54]	176	58 (33.0%)	EWGSOP2	S: 50.32 ± 8.58 NS: 73.75 ± 13.51 p<0.001	0.95	8/8 hypotheses confirmed	0.976 (0.959-0.986)	absent
Russian (2019)	[55]	100	50 (50%)	EWGSOP	S: 50.65 ± 14.23 NS: 75.10 ± 14.46 p<0.001	0.924	Mixed results for convergent hypotheses	0.935 (0.91-0.96)	

Table 1: Results from studies on the measurement properties of the SarQoL questionnaire

Version and year of publication	Reference	Sample size (n)	Sarcopenic subjects [n(%)]	Sarcopenia diagnosis	Discriminative power (overall SarQoL score)	Internal consistency (α)	Construct validity	Test-retest reliability (ICC)	Floor and ceiling effects
Ukrainian (2020)	[56]	49	28 (57.1%)	Ishii test	S: 58.43 ± 17.13 NS: 69.89 ± 13.31 p=0.014	0.898	Convergent validity confirmed; divergent validity refuted	0.997 (0.994-0.998)	absent
Spanish (2020)	[57]	252	66 (26.2%)	EWGSOP2	S: 71.19 (57.51–78.89) NS: 76.04 (64.83–87.07) p=0.008	0.904	Convergent and divergent hypotheses confirmed	0.99 (0.98-0.99)	absent
Serbian (2020)	[58]	699	12 (1.7%)	EWGSOP2	S: 60.31 (44.48–68.85) NS: 64.60 (54.93–74.50) p=0.155	0.87	6/8 hypotheses confirmed		absent
Turkish (2021)	[59]	100	27 (27%) low muscle strength	EWGSOP2 – probable sarcopenia	S: 50 ± 16 NS: 68.9 ± 16.9 p< 0.001	0.88	Convergent validity confirmed, divergent validity not confirmed	0.97 (0.94-0.98)	absent
Korean (2021)	[60]	450	53 (11.8%)	EWGSOP2		0.886	Convergent and divergent validity confirmed	0.977 (0.975-0.979)	absent

α: Cronbach's alpha value; ICC: intraclass correlation coefficient; EWGSOP: European Working Group on Sarcopenia in Older People; S: sarcopenic; NS: non-sarcopenic.

Objectives

This dissertation project builds upon the work previously performed on the development and validation of the SarQoL questionnaire [61]. To keep the momentum around the questionnaire going, there was a need to provide more information to potential users about the measurement properties of the questionnaire, so as to inspire its inclusion in research studies involving sarcopenic participants. A considerable number of clinical trials have looked at potential treatments and interventions to prevent sarcopenia, and the importance of having a patient-reported outcome measure that is relevant to the outcome priorities of the patient themselves is widely recognized. Generic quality-of-life questionnaires provide information but having a specific quality of life questionnaire such as the SarQoL questionnaire available allows for a much more relevant measure of change over time. Without a thorough demonstration of the measurement properties of the SarQoL questionnaire, its adoption in clinical trials and other studies would be hindered and the quality-of-life measurements on which clinical decisions would be based would continue to be constrained by the generic nature of the available instruments. This objective constitutes the first axis of this dissertation, in which we investigated the measurement properties of the questionnaire, such as its measurement error and responsiveness, and its applicability in different situations. Parallel to this, discussions about the desirability of creating a shorter version of the SarQoL questionnaire took place. A shorter version of the questionnaire would help both patients, in that the cognitive burden associated with completing the questionnaire would be reduced, as well as researchers, by making it easier to integrate the questionnaire into clinical studies, and also clinical practitioners, for whom the long version was simply too time-consuming to be practical. A shorter version would also allow us to eliminate some of the weaknesses in the formulation of certain questions and response options. This became the second axis of this dissertation: the development and validation of a shorter version of the SarQoL questionnaire.

In practice, the two axes of this project were developed as follows:

- Axis 1: To investigate and document the measurement properties of the SarQoL questionnaire and its applicability in different situations.

During the development of the SarQoL questionnaire, one of the applications envisioned was the longitudinal evaluation of quality of life in sarcopenia. Therefore, we started off this dissertation project by evaluating the ability of the SarQoL questionnaire to detect change in quality of life over time, or its responsiveness for short, and have reported these results in chapter 1. We then investigated, in chapter 2, the measurement error of the questionnaire, in participants from 9 validation studies who completed the questionnaire twice in the span of a few weeks. We quantified the random error of the questionnaire

with the standard error of measurement, and calculated the smallest detectable change, which is the minimum amount of change that needs to be measured to be certain that the observed change is true, and not potentially due to measurement error. In chapter 3, we respond to the publication of the revised consensus criteria for the diagnosis of sarcopenia by re-evaluating the discriminative power of the questionnaire with these new criteria. In chapter 4, we have looked beyond the confines of the domain of sarcopenia and have investigated whether the SarQoL questionnaire could be used to evaluate QoL in physical frailty, a geriatric syndrome characterized by increased vulnerability to stressor events due to age-related declines in the physiological function and reserve capacity across multiple organ systems. Because of the conceptual overlap between the frailty phenotype and sarcopenia, we hypothesized that it could provide valid information on QoL and verified that it possessed adequate measurement properties to do so. In chapter 5, the last chapter of the first part, we investigated whether the questionnaire could serve to screen older people and single out those likely to be sarcopenic, and thus provide a secondary application of the SarQoL questionnaire.

- Axis 2: To develop a shorter version of the SarQoL questionnaire and evaluate its measurement properties.

In the second axis of this dissertation, we set out to reduce the number of items in the questionnaire and to create a short version that would decrease response burden and facilitate its integration in larger studies, all while safeguarding its measurement properties. In chapter 6, the development of a 14-item version of the questionnaire is reported, as well as an evaluation of its measurement properties in a sample of 214 older, community-dwelling people. Finally, in chapter 7, we investigated the relative importance of the 15 items in the short-form SarQoL through best-worst scaling, a choice experiment method, in the same sample. This allowed us to establish a ranking of the importance of the 14 items from the perspective of its target population.

On the next 2 pages, tables 2 and 3 detail the structure of this dissertation and the origin of the datasets used for each chapter.

Table 2: objectives and publications

Axis 1: Measurement properties of the SarQoL questionnaire and application for different purposes		
Chapter	**Objective**	**Publication**
1	To investigate the responsiveness of the SarQoL questionnaire.	Geerinck A, Bruyère O, Locquet M, Reginster J-Y & Beaudart C (2018). Evaluation of the Responsiveness of the SarQoL® Questionnaire, a Patient-Reported Outcome Measure Specific to Sarcopenia. Adv Ther 35, 1842–1858.
2	To investigate the measurement error of the SarQoL questionnaire.	Geerinck A, Alekna V, Beaudart C, Bautmans I, Cooper C, De Souza Orlandi F, Konstantynowicz J, Montero-Errasquín B, Topinková E, Tsekoura M, Reginster J-Y & Bruyère O (2019). Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: An analysis of subjects from 9 validation studies. PLOS ONE 14(4): e0216065.
3	To investigate the discriminative power of the SarQoL questionnaire with the revised EWGSOP2 consensus criteria for the diagnosis of sarcopenia.	Geerinck A, Locquet M, Reginster J-Y, Bruyère O & Beaudart C (2021). Letter to the Editor: Discriminative Power of the Sarcopenia Quality of Life (SarQoL®) Questionnaire with the EWGSOP2 Criteria. J Frailty Aging. 10(2):193-194.
4	To investigate the appropriateness and measurement properties of the SarQoL questionnaire for measuring quality of life in physical frailty.	Geerinck A, Locquet M, Bruyère O, Reginster J-Y & Beaudart C (2021). Evaluating quality of life in frailty: applicability and clinimetric properties of the SarQoL® questionnaire. J Cachexia Sarcopenia Muscle.
5	To investigate the diagnostic accuracy of the SarQoL questionnaire when used to screen older people for sarcopenia according to the EWGSOP2 criteria.	Geerinck A, Dawson-Hughes B, Beaudart, C, Locquet M, Reginster J-Y & Bruyère O (2021). Assessment of the performance of the SarQoL questionnaire in screening for sarcopenia in older people. Aging Clin Exp Res 33:2149-2155.
Axis 2: The short-form SarQoL questionnaire and the relative importance of its 14 items		
Chapter	**Objective**	**Publication**
6	To develop a shorter version of the SarQoL questionnaire and to investigate its measurement properties.	Geerinck A, Beaudart C, Reginster J-Y, Locquet M, Monseur C, Gillain S & Bruyère O (2021). Development and validation of a short version of the Sarcopenia Quality of Life questionnaire: the SF-SarQoL. Qual Life Res.
7	To investigate the relative importance, from a patient perspective, of the 14 aspects of quality of life in the SF-SarQoL.	Geerinck A, Locquet M, Hiligsmann M, Reginster J-Y, Bruyère O & Beaudart C. Patients' preferences for quality of life aspects in sarcopenia: a best-worst scaling study. Submitted to European Geriatric Medicine.

Table 3: Datasets used within this dissertation

			Ch. 1: Responsiveness	Ch. 2: SEM &SDC	Ch 3: EWGSOP2	Ch. 4: Frailty	Ch. 5: Screening	Ch. 6: SF-SarQoL	Ch. 7: BWS
Existing datasets used in this study	SarcoPhAge study	Baseline							
		T1	X	X	X	X			
		T2							
		T3	X				X		
		T4						X	X
		T5				X		X	X
	Validation studies	Flanders		X					
		Brazil		X					
		Czech		X					
		England		X					
		Greece		X					
		Lithuania		X					
		Poland		X					
		Spain		X					
Newly collected dataset	Previous participants from SarcoPhAge							X	X

Datasets used in the 7 chapters are indicated by "X"
SEM: standard error of measurement; SDC: smallest detectable change; EWGSOP2: 2nd European Working Group on Sarcopenia in Older People; BWS: best-worst scaling; T1: dataset collected 1 year after baseline measurements.

References

1. Roser M., Ortiz-Ospina E., Ritchie H. (2013) Life expectancy. In: OurWorldInData.org. https://ourworldindata.org/life-expectancy. Accessed 30 Apr 2021
2. World Health Organization. (2018) Ageing and health: key facts. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Accessed 30 Apr 2021
3. 73rd World Health Assembly Decisions. https://www.who.int/news/item/07-08-2020-73rd-world-health-assembly-decisions. Accessed 30 Apr 2021
4. Anker SD., Morley JE., von Haehling S. (2016) Welcome to the ICD-10 code for sarcopenia. J Cachexia Sarcopenia Muscle 7:(5):512–514
5. Rosenberg I. (1989) Summary comments: epidemiological and methodological problems in determining nutritional status of older persons. Am J Clin Nutr 50:1231–1233
6. Critchley M. (1931) the Neurology of Old Age. Lancet 217:(5623):1221–1231
7. Rosenberg I. (2011) Sarcopenia: Origins and clinical relevance. Clin Geriatr Med 27:(3):337–339
8. Baumgartner R., Koehler K., Gallagher D., Romero L., Heymsfield S., Ross R., et al. (1998) Epidemiology of sarcopenia among the elderly in New Mexico. Am J Epidemiol 147:(8):755–763
9. Delmonico MJ., Harris TB., Lee JS., Visser M., Nevitt M., Kritchevsky SB., et al. (2007) Alternative definitions of sarcopenia, lower extremity performance, and functional impairment with aging in older men and women. J Am Geriatr Soc 55:(5):769–774
10. Janssen I., Baumgartner RN., Ross R., Rosenberg IH., Roubenoff R. (2004) Skeletal Muscle Cutpoints Associated with Elevated Physical Disability Risk in Older Men and Women. Am J Epidemiol 159:(4):413–421
11. Bischoff-Ferrari HA., Orav JE., Kanis JA., Rizzoli R., Schlögl M., Staehelin HB., et al. (2015) Comparative performance of current definitions of sarcopenia against the prospective incidence of falls among community-dwelling seniors age 65 and older. Osteoporos Int 26:(12):2793–2802
12. Muscaritoli M., Anker SD., Argilés J., Aversa Z., Bauer JM., Biolo G., et al. (2010) Consensus definition of sarcopenia, cachexia and pre-cachexia: Joint document elaborated by Special Interest Groups (SIG) " cachexia-anorexia in chronic wasting diseases" and " nutrition in geriatrics." Clin Nutr 29:(2):154–159

13. Cruz-Jentoft AJ., Baeyens JP., Bauer JM., Boirie Y., Cederholm T., Landi F., et al. (2010) Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age Ageing 39:(4):412–423

14. Cruz-Jentoft AJ., Bahat G., Bauer J., Boirie Y., Bruyère O., Cederholm T., et al. (2018) Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing (1):1–16

15. Fielding RA., Vellas B., Evans WJ., Bhasin S., Morley JE., Newman AB., et al. (2011) Sarcopenia: An Undiagnosed Condition in Older Adults. Current Consensus Definition: Prevalence, Etiology, and Consequences. International Working Group on Sarcopenia. J Am Med Dir Assoc 12:(4):249–256

16. Morley JE., Abbatecola AM., Argiles JM., Baracos V., Bauer J., Bhasin S., et al. (2011) Sarcopenia With Limited Mobility: An International Consensus. J Am Med Dir Assoc 12:(6):403–409

17. Studenski S., Peters K., Alley D., Cawthon P., Mclean R., Harris T., et al. (2014) The FNIH Sarcopenia Project : Rationale , Study Description, Conference Recommendations and Final Estimates. J Gerontol A Biol Sci Med Sci 69:(5):547–558

18. Chen L-K., Liu L-K., Woo J., Assantachai P., Auyeung T-W., Bahyah KS., et al. (2014) Sarcopenia in Asia: consensus report of the Asian Working Group for Sarcopenia. J Am Med Dir Assoc 15:(2):95–101

19. Chen LK., Woo J., Assantachai P., Auyeung TW., Chou MY., Iijima K., et al. (2020) Asian Working Group for Sarcopenia: 2019 Consensus Update on Sarcopenia Diagnosis and Treatment. J Am Med Dir Assoc 21:(3):300-307.e2

20. Bhasin S., Travison TG., Manini TM., Patel S., Pencina KM., Fielding RA., et al. (2020) Sarcopenia Definition: The Position Statements of the Sarcopenia Definition and Outcomes Consortium. J Am Geriatr Soc 68:(7):1410-1418

21. Keller K., Engelhardt M. (2013) Strength and muscle mass loss with aging process . Age and strength loss. 3:(4):346–350

22. Kim TN., Choi KM. (2013) Sarcopenia : Definition , Epidemiology , and Pathophysiology. J Bone Metab 20:(1):1–10

23. Papadopoulou SK. (2020) Sarcopenia: A Contemporary Health Problem among Older Adult Populations. Nutrients 12:(5):1293

24. Cruz-Jentoft AJ., Landi F., Schneider SM., Zuniga C., Arai H., Boirie Y., et al. (2014) Prevalence of and interventions for sarcopenia in ageing adults: A systematic review. Report of the International Sarcopenia Initiative (EWGSOP and IWGS). Age Ageing 43:(6):48–759

25. Shafiee G., Keshtkar A., Soltani A., Ahadi Z., Larijani B., Heshmat R. (2017) Prevalence of sarcopenia in the world : a systematic review and meta- analysis of general population studies. J Diabetes Metab Disord 16:(21):1–10

26. Locquet M., Beaudart C., Petermans J., Reginster JY., Bruyère O. (2019) EWGSOP2 Versus EWGSOP1: Impact on the Prevalence of Sarcopenia and Its Major Health Consequences. J Am Med Dir Assoc 20:(3):384–385

27. Veronese N., Demurtas J., Soysal P., Smith L., Torbahn G., Schoene D. (2019) Sarcopenia and health-related outcomes : an umbrella review of observational studies. Eur Geriatr Med 10:853-862

28. Bruyère O., Beaudart C., Ethgen O., Reginster JY., Locquet M. (2019) The health economics burden of sarcopenia: a systematic review. Maturitas 119:61–69

29. Mijnarends D., Schols J., Halfens R., Meijers J., Luiking Y., Verlaan S., et al. (2016) Burden-of-illness of Dutch community-dwelling older adults with sarcopenia : Health related outcomes and costs. Eur Geriatr Med 7:(3):276–284

30. Beckwée D., Delaere A., Aelbrecht S., Baert V., Beaudart C., Bruyere O., et al. (2019) Exercise Interventions for the Prevention and Treatment of Sarcopenia. A Systematic Umbrella Review. J Nutr Heal Aging 23:(6):494–502

31. Dent E., Morley JE., Cruz-Jenthoft AJ., Arai H., Kritchevsky SB., Guralnik J., et al. (2018) International Clinical Practice Guidelines for Sarcopenia (ICFSR): Screening, Diagnosis and Management. J Nutr Heal Aging 22:(10):1148–1161

32. Gielen E., Beckwée D., Delaere A., De Breucker S., Vandewoude M., Bautmans I., et al. (2021) Nutritional interventions to improve muscle mass, muscle strength, and physical performance in older people: An umbrella review of systematic reviews and meta-analyses. Nutr. Rev. 79:121–147

33. Cruz-Jentoft AJ., Sayer AA. (2019) Sarcopenia. Lancet 393:(10191):2636–2646

34. De Spiegeleer A., Beckwée D., Bautmans I., Petrovic M., Beaudart C., Beyer I., et al. (2018) Pharmacological Interventions to Improve Muscle Mass, Muscle Strength and Physical Performance in Older People: An Umbrella Review of Systematic Reviews and Meta-analyses. Drugs and Aging 35:(8):719–734

35. Beaudart C., Buckinx F., Rabenda V., Gillain S., Cavalier E., Slomian J., et al. (2014) The effects of vitamin D on skeletal muscle strength, muscle mass, and muscle power: A systematic review and meta-analysis of randomized controlled trials. J Clin Endocrinol Metab 99:(11):4336–4345

36. The WHOQOL Group (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psych Med 28(3):551–558.

37. Bakas T., McLennon S., Carpenter J., Buelow J., Otte J., et al. (2012) Systematic review of health-related quality of life models. Health Qual Life Outcomes 2012;10:134.

38. Ferrans C., Zerwc J., Wilbur J., Larson, J. (2005). Conceptual model of health-related quality of life. J Nurs Scholarsh 37:(4):336-342

39. van Leeuwen K., van Loon M., van Nes F., Bosmans E., de Vet H., et al. (2019) What does quality of life mean to older adults? A thematic synthesis. PLoS ONE 14:(3):e0213263

40. Woo T., Yu S., Visvanathan R. (2016) Systematic literature review on the relationship between biomarkers of sarcopenia and quality of life in older people. J Frailty Aging 5:(2):88–99

41. Silva Neto LS., Karnikowski MGO., Osório NB., Pereira LC., Mendes MB., Galato D., et al. (2016) Association between sarcopenia and quality of life in quilombola elderly in Brazil. Int J Gen Med 9:89–97

42. Marques LP., Confortin SC., Ono LM., Barbosa AR., d'Orsi E. (2019) Quality of life associated with handgrip strength and sarcopenia: EpiFloripa Aging Study. Arch Gerontol Geriatr 81:234–239

43. Yalcin A., Silay K. (2017) Sarcopenia and health-related quality of life in Turkish nursing home residents: A cross-sectional study. Asian J Gerontol Geriatr 12:(2):42–46

44. Manrique-Espinoza B., Salinas-Rodríguez A., Rosas-Carrasco O., Gutiérrez-Robledo LM., Avila-Funes JA. (2017) Sarcopenia Is Associated With Physical and Mental Components of Health-Related Quality of Life in Older Adults. J Am Med Dir Assoc 18:(7):636.e1-635.e5

45. Beaudart C., Biver E., Reginster JY., Rizzoli R., Rolland Y., Bautmans I., et al. (2015) Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: The SarQoL. Age Ageing 44:(6):960–966

46. Mokkink LB., Terwee CB., Patrick DL., Alonso J., Stratford PW., Knol DL., et al. (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 63:(7):737–745

47. Galesic M., Bosnjak M. (2009) Effects of questionnaire length on participation and indicators of response quality in a web survey. Public Opin Q 73:(2):349–360

48. Beaudart C., Biver E., Reginster J-Y., Rizzoli R., Rolland Y., Bautmans I., et al. (2017) Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 8:(2):238-244

49. Beaudart C., Edwards M., Moss C., Reginster JY., Moon R., Parsons C., et al. (2017) English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. Age Ageing 46:(2):271–277

50. Ildiko GA., Gabriela M., Charlotte B., Olivier B., Raluca-Monica P., Jean-Yves R., et al. (2017) Psychometric performance of the Romanian version of the SarQoL®, a health-related quality of life questionnaire for sarcopenia. Arch Osteoporos 12:(1):103

51. Geerinck A., Scheppers A., Beaudart C., Bruyère O., Vandenbussche W., Bautmans R., et al. (2018) Translation and validation of the Dutch SarQoL®, a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact 18:(4):463–472

52. Konstantynowicz J., Abramowicz P., Glinkowski W., Taranta E., Marcinowicz L., Dymitrowicz M., et al. (2018) Polish Validation of the SarQoL®, a Quality of Life Questionnaire Specific to Sarcopenia. J Clin Med 7:(10):323

53. Tsekoura M., Billis E., Gliatis J., Tsepis E., Matzaroglou C., Sakkas GK., et al. (2018) Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. Disabil Rehabil 42:(7):1006-1012

54. Alekna V., Kilaite J., Tamulaitiene M., Geerinck A., Mastaviciute A., Bruyère O., et al. (2019) Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). Eur Geriatr Med 10:761-767

55. Safonova YA., Lesnyak OM., Baranova IA., Suleimanova AK., Zotkin EG. (2019) Russian translation and validation of SarQoL® - quality of life questionnaire for patients with sarcopenia. Nauchno-Prakticheskaya Revmatol 57:(1):38–45

56. Dzhus M., Dzhus M., Masnyi M., Kulyk M., Mostbauer H., Ivashkivsky O., et al. (2020) Cross-sectional evaluation of the sarcopenia quality of life (Sarqol) questionnaire: Translation and validation of its psychometric properties. Ann Geriatr Med Res 24:(2):139–147

57. Fábrega-Cuadros R., Martínez-Amat A., Cruz-Díaz D., Aibar-Almazán A., Hita-Contreras F. (2020) Psychometric Properties of the Spanish Version of the Sarcopenia and Quality of Life, a Quality of Life Questionnaire Specific for Sarcopenia. Calcif Tissue Int 106:(3):274–282

58. Matijević R., Hrnjaković O., Đurđević A., Geerinck A., Beaudart C., Bruyère O., et al. (2020) Translation and psychometric performance of the Serbian version of the sarcopenia quality of life (SarQoL®) questionnaire. Srp Arh Celok Lek 148:(11–12):742–748

59. Erdogan T., Eris S., Avci S., Oren MM., Kucukdagli P., Kilic C., et al. (2021) Sarcopenia quality-of-life questionnaire (SarQoL®): translation, cross-cultural adaptation and validation in Turkish. Aging Clin Exp Res (epub ahead of print)

60. Yoo J Il., Ha YC., Kim M., Seo SH., Kim MJ., Lee GY., et al. (2021) Translation and validation of the Korean version of the Sarcopenia Quality of Life (SarQoL-K®) questionnaire and applicability with the SARC-F screening tool. Qual Life Res 30:(2):603–611

61. Beaudart C. (2016) Contribution à l'étude de la sarcopénie : définition, diagnostic et conséquences. University of Liège (Doctoral dissertation)

PART 1

Investigation of the measurement properties of the SarQoL questionnaire and its applicability in different situations

Chapter 1: Responsiveness of the SarQoL questionnaire

Evaluation of the Responsiveness of the SarQoL® Questionnaire, a Patient-Reported Outcome Measure Specific to Sarcopenia

Published as:

Geerinck, A., Bruyère, O., Locquet, M., Reginster, J. Y., & Beaudart, C. (2018). Evaluation of the Responsiveness of the SarQoL® Questionnaire, a Patient-Reported Outcome Measure Specific to Sarcopenia. Advances in therapy, 35(11), 1842–1858. https://doi.org/10.1007/s12325-018-0820-z

1. Abstract

Introduction : The Sarcopenia Quality of Life (SarQoL®) questionnaire was developed to provide a patient-reported outcome measure specific to sarcopenia. Its psychometric properties indicate that it is a valid and reliable instrument. However, until now, its ability to detect change over time has not been examined. Therefore, the objective of this study is to evaluate the responsiveness (also known as sensitivity to change) of the SarQoL® questionnaire in a prospective, longitudinal cohort of community-dwelling, older, sarcopenic subjects.

Methods : Sarcopenic subjects from the SarcoPhAge (Sarcopenia and Physical impairment with advancing Age) study were included. Responsiveness was evaluated with nine pre-specified hypotheses on the correlation between the evolution of the SarQoL® scores after a 2-year interval and the evolution of the scores on the Short Form-36 (SF-36) and the Euroqol 5-dimension (EQ-5D) questionnaires. This technique considers responsiveness to be a form of longitudinal validity. Additionally, standardized response means were also calculated to compare the quantity of change measured by the different questionnaires.

Results : A total of 42 sarcopenic subjects were included. The median age of the sample was 72.9 (68.9–78.8) years, 59.5% were female, and the mean body mass index was 23.3 (20.4–25.7) kg/m^2. A good responsiveness was observed, as evidenced by the confirmation of eight out of nine hypotheses, well above the 75% confirmation threshold. The standardized response mean of the Overall SarQoL® score was significantly higher than those of the SF-36 Physical Component Summary ($p = 0.005$), the EQ-5D Utility Index ($p < 0.001$) and the Euroqol visual analogue scale ($p = 0.003$).

Conclusion : The first data available on the ability of the SarQoL® questionnaire to detect change over time indicates that the questionnaire has good responsiveness. This, together with the previously established psychometric properties, confirms that the SarQoL® questionnaire is a relevant instrument for the assessment of quality of life in sarcopenic populations.

2. Introduction

Sarcopenia, defined as "a syndrome characterised by progressive and generalised loss of skeletal muscle mass and strength and with a risk of adverse outcomes such as physical disability, poor quality of life and death" by the European Working Group on Sarcopenia in Older People (EWGSOP), is a growing public health problem [1]. It has recently been recognized as a geriatric condition with an ICD-10-CM code (M62.84) [2]. Sarcopenia has been shown to be associated with negative health outcomes, such as a higher rate of mortality and functional decline, a higher rate of falls, and a higher incidence of hospitalization [3]. Other research has shown an association between sarcopenia and depression [4]. Not much is yet known about the relationship between sarcopenia and quality of life. Although several studies have incorporated quality of life outcomes in their designs, the results are difficult to compare because of the different diagnostic criteria used to establish sarcopenia. Some studies that diagnosed sarcopenia with the EWGSOP criteria have found lower health-related quality of life (HRQoL) scores for sarcopenic subjects in select domains of the Short-Form 36-item (SF-36) questionnaire, but other studies (using other diagnostic criteria) have found no difference in SF-36 scores between sarcopenic and non-sarcopenic subjects [5].

Until recently, researchers only had generic questionnaires, such as the SF-36, available to assess quality of life in sarcopenic patients. These questionnaires are designed for use in broad populations and may thus not be sensitive enough to accurately measure quality of life in sarcopenic populations [6]. To address this problem, Beaudart et al. developed the Sarcopenia Quality of Life (SarQoL®) questionnaire in 2015 [7].

Until now, no study has evaluated the responsiveness, defined as "the ability of an instrument to detect change over time in the construct to be measured", of the SarQoL® questionnaire [8]. When an instrument is used for evaluative purposes, i.e. when the aim is to detect and measure longitudinal change in subjects or populations, responsiveness is a key psychometric property [9, 10]. This situation is often present in clinical studies aimed at testing the effect of an intervention, where an accurate assessment of HRQoL before and after the intervention is an important outcome. Researchers need to have valid data on the responsiveness of the instrument they wish to use to be certain of the results they obtain.

The psychometric properties of the SarQoL® questionnaire have been evaluated in several cross-sectional studies, but until now, its ability to detect change over time (responsiveness) had not yet been examined [11–14]. This study aimed to evaluate the responsiveness of the SarQoL® questionnaire in a sample of older, community-dwelling, sarcopenic subjects from the SarcoPhAge (Sarcopenia and Physical impairment with advancing Age) cohort.

3. Methods

3.1. Design

The current article describes an instrument validation study that examined data collected at the 2nd and 4th annual visit of the SarcoPhAge study, an ongoing 5-year prospective, longitudinal, observational cohort study being carried out in Liège (Belgium) [15, 16]. Participants in the SarcoPhAge study all provided written informed consent. The research protocol and its amendments were approved by the Ethics Committee of the University Teaching Hospital of Liège (no. 2012-277).

3.2. Participants

Participants from the SarcoPhAge study with valid data from the 2nd (T1) and 4th (T3) study visit (a 2-year interval) who were diagnosed as sarcopenic according to the EWGSOP criteria were included [1]. This 2-year interval was chosen because it covers the first and last available administrations of the questionnaire and because the SarcoPhAge study is an observational study; therefore, we relied on the natural progression of sarcopenia to cause a change in health status between the two measurements. The details of this study have been reported previously [11, 15, 17, 18].

Sarcopenia was diagnosed according to the EWGSOP algorithm, which demands the presence of low muscle mass in combination with low muscle strength and/or low physical performance [1]. Muscle mass was measured by dual-energy X-ray absorptiometry (DXA) (Hologic Discovery A, USA), which was calibrated daily by scanning a spine phantom. Male subjects with a skeletal muscle mass index (SMI = appendicular lean mass/height2) below 7.26 kg/m^2 and women with an SMI below 5.5 kg/m^2 were considered to have low muscle mass. Muscle strength was measured with a hydraulic hand dynamometer (Saehan Corporation, Korea), calibrated at the beginning of the study for 10, 40 and 90 kg. Men with a maximal handgrip strength below 30 kg and women below 20 kg were considered to have low muscle strength. Physical performance was examined with the help of the Short Physical Performance Battery (SPPB), with a value of 8 or less being considered low [15].

Participants were included in the current analysis when diagnosed as sarcopenic at T1 and/or T3 and when both SarQoL® questionnaires (T1 and T3) had less than 20% missing data for the calculation of the Overall score.

3.3. Measures

The SarQoL® Questionnaire

The SarQoL® questionnaire is a patient-reported outcome measure (PROM) specific to sarcopenia. The SarQoL® questionnaire consists of 22 questions incorporating 55 items, which fall into seven domains of HRQoL. These domains are "Physical and Mental Health", "Locomotion", "Body Composition", "Functionality", "Activities of Daily Living", "Leisure activities" and "Fears". Each domain is scored from 0 to 100, and an Overall score is calculated. The questionnaire is auto-administered and takes 10 min to complete [7]. The questionnaire is available in 16 languages and can be found on its webpage [19].

Several psychometric properties of the SarQoL® questionnaire have been examined previously. The questionnaire has demonstrated its ability to distinguish between sarcopenic and non-sarcopenic subjects (discriminative power). It has good internal consistency and construct validity, and its test–retest reliability is excellent. Furthermore, it has been demonstrated that there are no floor or ceiling effects for the Overall score [11–14].

The Short-Form 36-Item (SF-36) Questionnaire

The SF-36 is a multi-item generic health survey that uses 36 questions to measure functional health and wellbeing from the patient's perspective. It measures eight domains: "Physical Functioning", "Role limitation due to physical problems", "Bodily Pain", "General Health Perceptions", "Vitality", "Social Functioning", "Role limitations due to emotional problems" and "Mental Health", each of which provides a score between 0 and 100. Additionally, two composite scores can be calculated: the Physical Component Summary (PCS) and the Mental Component Summary (MCS) [20–22].

The EuroQol 5-Dimension 3-Level (EQ-5D-3L)

The EQ-5D-3L is a standardized measure of health status developed by the EuroQol Group in 1990. The instrument consists of two pages: the EQ-5D descriptive system, which is composed of five questions encompassing five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression); and the Visual Analogue Scale (EQ-VAS), which records the respondent's self-rated health on a vertical scale going from best (100) to worst imaginable health (0). The EQ-5D descriptive system is used to calculate an index score, which represents the utility value for current health [23, 24].

Physical Parameters

Parameters related to muscle mass, muscle strength and physical performance were collected. Apart from the SMI, we also determined appendicular lean mass (ALM) and ALM divided by body mass index (ALM/BMI) by DXA. As mentioned previously, muscle strength was determined with a hydraulic hand dynamometer. For physical performance, the patients performed the SPPB test, which also includes the usual gait speed on a 4-m track. The subjects also performed the timed-up-and-go (TUG) test, which uses the time that a subject takes to rise from a chair, walk three metres, turn around, walk back to the chair, and sit down to determine a subject's mobility. Lastly, the chair stand test (CST) was administered as part of the SPPB. In this test, the subjects are asked to stand up from a chair and sit back down five times as fast as they can.

3.4. Methodological Approach

Hypotheses Testing

It is recommended to treat responsiveness as the longitudinal form of construct validity and to evaluate it in much the same way as the construct validity of a questionnaire [25]. Thus, we formulated hypotheses between the changes in the scores of the SarQoL® questionnaire and the changes observed for the SF-36 and the EQ-5D. AG, CB and OB were responsible for the formulation of the hypotheses, on the basis of similarity in the construct of the different domains, and previously found results for the construct validity of the questionnaire. The data used in this analysis were collected before the formulation of the hypotheses, but no statistical manipulations in relation to the evaluation of responsiveness were carried out before the final set of hypotheses was agreed upon.

The hypotheses used for the evaluation of the responsiveness, the expected strength of the correlations and the rationale for their formulation are detailed in Table 1.

Table 1: Hypotheses for the evaluation of responsiveness

Hypotheses	Expected strength of correlation	Rationale
1. Δ SarQoL Overall score and Δ SF-36 General Health domain are correlated.	$r > 0.4$	The SarQoL Overall score and the SF-36 General Health score have been shown to be correlated in the French (r=0.67) and the English (r=0.49) validations. These domains are similar in that they both measure a subject's general view of either their HRQoL or health. Because of the strong interaction between general health status and HRQoL, we expect a correlation of at least 0.4, despite the difference in underlying construct.
2. Δ SarQoL Overall score and Δ SF-36 Vitality domain are correlated.	$r > 0.3$	Here also, two different constructs are measured, but they have been shown to be correlated (FR: r=0.72; ENG: r=0.74). Since the underlying constructs are less similar than in hypothesis 1, and we expect the influence of a change in vitality to be less impactful than one in General Health, the expected correlation was set to at least 0.3.
3. Δ SarQoL Overall score and Δ SF-36 Physical Functioning domain are correlated.	$r > 0.5$	The domain Physical functioning covers a significant portion of the content used to calculate the Overall score of the SarQoL®, although the Overall score also takes into account other aspects of HRQoL. The English validation confirmed this similarity with a correlation of 0.82, although the French validation found a smaller correlation of 0.49. Nevertheless, we expect changes on both measures to be correlated at a strength of at least 0.5.
4. Δ SarQoL Overall score and Δ EQ-VAS are correlated.	$r > 0.4$	The Overall score and the EQ-VAS both give a general view of the subjects' current health or HRQoL, and should thus, in theory, be correlated. We expect the difference in health as measured by the EQ-VAS to be reflected in changes in HRQoL (as evidenced by a cross-sectional correlation of r=0.597) but, since they measure two different but related constructs, it was decided to fix the expected strength of this association to at least 0.4.
5. Δ SarQoL domain 1 (Physical & Mental Health) and Δ SF-36 General Health domain are correlated.	$r > 0.3$	Domain 1 of the SarQoL® questionnaire carries significant weight in the calculation of the Overall score. Since we know a correlation exists for between the Overall score and the General Health domain for the construct validity (see hypothesis 1), we theorized that this same correlation should exist between Physical & Mental Health and General Health. We did expect this correlation to be weaker, although the cross-sectional correlation was r=0.655, since some aspects covered in the Overall score are not represented in Physical & Mental Health. It was decided to expect a correlation of at least 0.3.
6. Δ SarQoL domain 1 (Physical & Mental Health) and Δ EQ-VAS are correlated.	$r > 0.3$	In the same vein as hypothesis 5, we expected changes on Physical & Mental Health to be associated with changes on the EQ-VAS, as shown by a cross-sectional correlation of r=0.562. However, since a part of the content is lost when focusing on a single

Table 1: Hypotheses for the evaluation of responsiveness

Hypotheses	Expected strength of correlation	Rationale
		domain of the SarQoL®, it was decided to expect a weaker correlation than hypothesis 5, and to adopt at least 0.3 as the threshold.
7. Δ SarQoL domain 2 (Locomotion) and Δ SF-36 Physical Functioning domain are correlated.	$r > 0.4$	The ability to walk and the ease with which a person can walk are an important factor that influence the totality of how a person functions physically, demonstrated by a cross-sectional correlation of r=0.558. While the domain Locomotion is a much narrower construct than Physical Functioning, we expect both domains to be significantly correlated at a strength of at least 0.4.
8. Δ SarQoL domain 4 (Functionality) and Δ SF-36 Physical Functioning domain are correlated.	$r > 0.5$	The underlying constructs of the domains Functionality and Physical Functioning are, in theory, similar, and it was therefore felt that a relatively strong correlation of at least 0.5 was to be expected, even if the cross-sectional correlation was lower at r=0.420.
9. Δ SarQoL domain 5 (Activities of Daily Living) and Δ SF-36 Physical Functioning domain are correlated.	$r > 0.5$	While these two domains represent different underlying construct, we theorized that a change in physical functioning would be equally reflected in a change in a person's Activities of Daily Living, because one is a prerequisite for the other. It was felt that we should expect a relatively strong correlation of at least 0.5 since we expected these two domains to be interwoven even if the cross-sectional correlation was lower at r=0.460.

Δ = change in; r = correlation

We employed the criteria formulated by De Boer et al. to evaluate the results of the hypotheses testing. These state that a questionnaire has high responsiveness when less than 25% of hypotheses are refuted, moderate responsiveness when 25–50% are refuted and poor responsiveness when more than 50% are refuted [26].

Standardized Response Means (SRMs)

We also calculated SRMs for the different questionnaires, by dividing the mean difference between T1 and T3 by the standard deviation of the differences between the paired measurements [27]. The SRM reflects the magnitude of the change measured by the different questionnaires. Consequently, when greater SRMs are obtained, this is an indication of better responsiveness. To allow the use of the thresholds for responsiveness formulated by Cohen et al., which are designed for use with the effect size and which categorize an observed change, we applied the correction developed by Middel and Van Sonderen [28, 29]. After correcting the SRMs with the formula $[(SRM/\sqrt{2})/\sqrt{(1 - r)}$; with r = correlation between baseline and follow-up score], we categorized them as trivial when $SRM < 0.20$, small when $0.20 \leq SRM < 0.49$, moderate when $0.50 \leq SRM < 0.79$ and large when $SRM \geq 0.80$ [29].

A selection of SRMs were compared in pairs to evaluate whether they were significantly different. This was carried out using the modified jack-knife method, which uses linear regression to determine whether a significant difference exists between two SRMs [30]. For this measure, an individual SRM is first calculated for each subject by dividing their change score by the standard deviation of the change scores in the whole sample. Next, a "centred" SRM is calculated for each subject by subtracting the mean SRM score of the sample from the individual SRMs. With these variables, a linear regression is carried out with the individual SRMs of the two quality-of-life scores of interest as dependent variables and the "centred" SRM of one of the quality-of-life scores (either one will work) as the independent variable. A significant difference is demonstrated when the p value of the intercept is at most 0.05 [30, 31].

Correlations Between Physical Parameters and QoL

We investigated the relationship between the evolution of physical parameters linked to sarcopenia and the changes observed by the different questionnaires with the help of correlations. We selected the five summary/total scores available (SarQoL® Overall score, SF-36 PCS and MCS, EQ-5D Utility Index and EQ-VAS) to represent the HRQoL of the subjects and constructed correlations with usual gait speed, handgrip strength, SPPB score, ALM, ALM/BMI, SMI, TUG test and the chair stand test. The strength of the association was judged as excellent when larger than 0.81, very good when between 0.61 and 0.80, good when between 0.41 and 0.60, acceptable when between 0.21 and 0.40 and insufficient when less than 0.20 [32].

3.5. Statistical Analysis

Data were analysed using IBM SPSS Statistics, version 24.0.0.0 for Windows (Armonk, NY: IBM Corp).

The distribution of variables was determined by examining the histogram, the quantile–quantile plot, the Shapiro–Wilk test and the difference between mean and median. Gaussian variables are reported as the mean ± standard deviation and non-Gaussian variables as median (P25–P75). Nominal variables are reported as absolute (n) and relative frequencies (%). The presence of significant differences between T1 and T3 was examined with the paired samples t test for variables with normal distribution, the Wilcoxon matched-pair signed-rank test for non-Gaussian variables and the chi-squared test for nominal variables. Pearson correlations were calculated when both groups/variables had normal distributions. Spearman correlations were calculated when this was not the case.

Change scores were calculated by subtracting the scores from T1 from those obtained at T3. For quality of life, this means that a positive change score indicates an improvement and a negative change score a decline. The calculation of the SRMs, their correction with the technique from Middel and Van Sonderen and the modified jack-knife method used to detect significant differences between SRMs have been described in the preceding paragraphs.

A post hoc power analysis was conducted on the Pearson and Spearman correlations used in the primary outcome with the G*Power software, version 3.1.9.2 [33]. This analysis computes the achieved power for a bivariate normal model with an α-error of 0.05 and a sample size of 42 subjects.

Results were considered significant at $p \leq 0.05$.

4. Results

In total, 42 sarcopenic participants from the SarcoPhAge study fulfilled the inclusion criteria, which is a moderate sample size according to the COSMIN checklist [34]. The subjects had a median age of 73 (69–79) years at T1, and 25 out of 42 (59.5%) were women. The median number of drugs taken by the participants increased significantly ($p = 0.001$) from 6 (5–9) at T1 to 8 (6–10) at T3, as did the proportion of subjects who fell in the year before the study visits, from 8 (19.0%) at T1 to 16 (38.1%) at T3 ($p = 0.017$). The gait speed of the participants diminished significantly from a median of 1.02 (0.80–1.21) m/s at T1 to 0.89 (0.76–1.09) m/s at T3 ($p = 0.032$). In the sample as a whole, a slight but significant reduction in handgrip strength was observed, from a median of 19.75 (18.00–28.00) kg at T1 to 19.00 (16.75–22.50) kg at T3 ($p = 0.010$). This change was attributable to the female subjects ($p = 0.030$). No significant changes between T1 and T3 were found for BMI ($p = 0.393$), number of

comorbidities (p = 0.763), proportion of subjects who experienced a fracture in the year before the study visits (p = 0.268), independence in activities of daily living as measured by the Katz scale (0.942), SPPB score (p = 0.083), TUG test (p = 0.081), ALM/BMI (p = 0.197) and SMI (p = 0.451). The ALM of the whole sample diminished significantly (p = 0.035), but this effect was lost when the sample was divided into men (p = 0.287) and women (p = 0.072).

The three different questionnaires obtained different results for quality of life. The SarQoL® questionnaire measured a significant reduction for three domains (Body Composition, p = 0.023; Functionality, p = 0.002; Activities of Daily Living, p < 0.001) and the Overall score, which diminished from a median of 61.15 (51.15–71.76) at T1 to 54.56 (42.31–68.44) at T3 (p = 0.002). The SF-36 PCS and MCS, the EQ-5D Utility Index and the EQ-VAS, however, did not detect a significant change (respectively, p = 0.679, p = 0.062, p = 0.231 and p = 0.716). The complete clinical characteristics and the evolution of quality of life can be found in Table 2.

Table 2: Clinical characteristics and quality of life scores for sarcopenic sample (n=42)

	T1	T3	Change	P-value
Age (years)	72.90 (68.85;78.81)	NA	NA	NA
Gender				
Male	17 (40.5%)	NA	NA	NA
Female	25 (59.5%)	NA	NA	NA
BMI (kg/m²)	23.25 (20.35;25.68)	23.09 (20.06;25.84)	-0.03 (-0.67;0.58)	0.393[a]
Number of drugs	6.00 (5.00;9.00)	8.00 (6.00;10.00)	1.00 (0.00;3.00)	0.001[a]
Number of comorbidities	4.00 (3.00;6.25)	4.00 (2.75;7.00)	0.00 (0.00;0.00)	0.763[a]
Fall in last year				
Yes	8 (19.0%)	16 (38.1%)	NA	0.017[b]
No	34 (81.0%)	26 (61.9%)	NA	
Fracture in last year				
Yes	4 (9.5%)	4 (9.5%)	NA	0.268[b]
No	38 (90.5%)	38 (90.5%)	NA	
Katz score	8.00 (8.00;9.00)	8.00 (8.00;9.00)	0.00 (0.00;0.00)	0.942[a]
SPPB score	9.50 (8.00;11.00)	8.00 (6.75;11.00)	-0.50 (-2.00;0.25)	0.083[a]
Gait speed (m/s)	1.02 (0.80;1.21)	0.89 (0.76;1.09)	-0.10 (-0.26;0.14)	0.032[a]
Chair stand test (s)	14.57 (11.97;18.29)	16.07 (11.06;20.94)	1.06 (-0.86;3.34)	0.083[a]
Timed up-and-go (s)	10.67 (8.66;13.31)	12.23 (9.15;16.27)	0.88 (-1.18;3.14)	0.081[a]
Hand Grip strength (kg)	19.75 (18.00;28.00)	19.00 (16.75;22.50)	-1.50 (-5.25;1.00)	0.010[a]
HGS men (kg)	28.00 (21.00;37.00)	25.00 (20.50;31.50)	-1.00 (-7.50;1.00)	0.146[a]
HGS women (kg)	19.00 (14.50;20.25)	18.00 (12.00;19.25)	-2.00 (-4.25;1.00)	0.030[a]
ALM (kg)	14.31 (13.09;18.73)	14.02 (12.94;18.14)	-0.29 (-0.57;1.16)	0.035[a]
ALM men	18.92 (17.44;20.26)	19.10 (16.79;19.97)	-0.30 (-0.63;0.30)	0.287[a]

Table 2: Clinical characteristics and quality of life scores for sarcopenic sample (n=42)				
	T1	T3	Change	P-value
ALM women	13.40 (12.47;14.15)	13.21 (12.03;14.00)	-0.29 (-0.48;0.16)	0.072[c]
ALM/BMI	0.69 (0.57 (0.74)	0.67 (0.58-0.74)	-0.01 (-0.04;0.02)	0.197[a]
ALM/BMI men	0.74 (0.70;0.88)	0.74 (0.69;0.89)	-0.01 (-0.03;0.03)	0.940[a]
ALM/BMI women	0.60 (0.53;0.70)	0.60 (0.51;0.67)	-0.02 (-0.04;0.02)	0.141[a]
SMI (kg/m²)	5.57 (5.25-6.70)	5.51 (5.14-6.60)	-0.02 (-0.18;0.15)	0.451[a]
SMI men	6.86 (6.37-7.16)	6.84 (6.26;7.30)	-0.01 (-0.20;0.17)	0.454[a]
SMI women	5.26 (5.11;5.52)	5.32 (4.94;5.47)	-0.04 (-0.19;0.10)	0.440[c]
SarQoL D1 Physical and Mental Health	58.87 (45.53;69.15)	51.09 (41.37;67.19)	-5.00 (-12.51;4.72)	0.107[a]
SarQoL D2 Locomotion	55.56 (46.53;72.22)	55.56 (38.20;70.14)	-2.78 (-11.81;5.55)	0.331[c]
SarQoL D3 Body Composition	58.33 (45.83;67.71)	50.00 (41.67;60.63)	-4.16 (-12.92;4.17)	0.023[c]
SarQoL D4 Functionality	70.24 (59.49;82.85)	63.46 (47.60;75.89)	-4.55 (-10.70;1.78)	0.002[a]
SarQoL D5 Activities of Daily Living	61.61 (43.33;75.00)	48.22 (37.29;65.42)	-6.43 (-20.00;-3.12)	<0.001[a]
SarQoL D6 Leisure activities	33.25 (29.09;49.88)	33.25 (16.62;66.50)	0.00 (-16.62;16.62)	0.645[a]
SarQoL D7 Fears	87.50 (75.00;100.00)	87.50 (75.00;100.00)	0.00 (-12.50;0.00)	0.382[a]
SarQoL Overall score	61.15 (51.15;71.76)	54.56 (42.31;68.44)	-5.23 (-12.46;1.61)	0.002[a]
SF-36 PCS	42.08 (31.86;49.14)	37.65 (30.47;48.24)	1.40 (-5.36;4.78)	0.679[a]
SF-36 MCS	44.71 (33.86;53.31)	38.91 (30.55;50.40)	-2.18 (-10.13;3.77)	0.062[a]
EQ-5D Utility Index	0.800 (0.517-0.827)	0.800 (0.708-0.827)	0.00 (-0.193;0.1557)	0.231[a]
EQ-VAS	70.00 (60.00-75.00)	70.00 (60.00-75.00)	0.00 (-7.50;5.00)	0.716[a]
[a] Wilcoxon Matched-Pair Signed-Rank test [b] Chi-squared test [c] Paired Samples T-test NA= not applicable; PCS: Physical Component Summary; MCS: Mental Component Summary				

4.1. Responsiveness

Of the nine formulated hypotheses, 8 (89%) were confirmed. Hypothesis 9 was rejected when a correlation of $r = 0.467$ was found, just under the threshold of $r > 0.5$. In total, three very good correlations were found, five good correlations and two acceptable correlations. The results of this evaluation as well as of the power analysis are reported in Table 3.

Table 3: Evaluation of responsiveness with hypotheses					
Hypothesis	Expected strength of correlation	Observed correlation		Confirmation/ rejection	Power (1-β)
		r	p-value		
1. Δ SarQoL Overall score and Δ SF-36 General Health domain are correlated.	r > 0.4	0.442[a]	0.005	Confirmed	0.851
2. Δ SarQoL Overall score and Δ SF-36 Vitality domain are correlated.	r > 0.3	0.454[b]	0.004	Confirmed	0.872
3. Δ SarQoL Overall score and Δ SF-36 Physical Functioning domain are correlated.	r > 0.5	0.669[a]	<0.001	Confirmed	0.999
4. Δ SarQoL Overall score and Δ EQ-VAS are correlated.	r > 0.4	0.404[a]	0.009	Confirmed	0.773
5. Δ SarQoL domain 1 (Physical & Mental Health) and Δ SF-36 General Health domain are correlated.	r > 0.3	0.610[a]	<0.001	Confirmed	0.994
6. Δ SarQoL domain 1 (Physical & Mental Health) and Δ EQ-VAS are correlated.	r > 0.3	0.312[a]	0.047	Confirmed	0.531
7. Δ SarQoL domain 2 (Locomotion) and Δ SF-36 Physical Functioning domain are correlated.	r > 0.4	0.412[a]	0.010	Confirmed	0.791
8. Δ SarQoL domain 4 (Functionality) and Δ SF-36 Physical Functioning domain are correlated.	r > 0.5	0.680[a]	<0.001	Confirmed	0.999
9. Δ SarQoL domain 5 (Activities of Daily Living) and Δ SF-36 Physical Functioning domain are correlated.	r > 0.5	0.467[a]	0.003	Rejected	0.893
Δ = change in; *r* = correlation [a] Spearman correlation [b] Pearson correlation					

According to the criteria by De Boer et al., the SarQoL® questionnaire possesses high responsiveness because fewer than 25% of hypotheses are refuted [26].

4.2. Standardized Response Means

The magnitude of change observed in the sample was examined by calculating SRMs. The SarQoL® questionnaire had three domains with SRMs below 0.20, indicating that no change was observed, two domains with an SRM between 0.20 and 0.49 (small change) and three domains with an SRM between 0.50 and 0.79 (moderate change). In contrast, only one domain of the SF-36 had a moderate SRM (Physical Functioning; SRM = − 0.50), and six domains reported an SRM indicating small change. A further three domains of the SF-36 had SRMs indicating no change had occurred. For the EQ-5D, small

SRMs were observed for two domains, with the remaining five domains having SRMs indicating no change. All obtained SRMs can be found in Table 4.

Table 4: Standardized Response Means

Domains	SRM	Corrected SRM	Interpretation[a]
Δ SarQoL D1 Physical & Mental Health	-0.31	-0.34	small change
Δ SarQoL D2 Locomotion	-0.15	-0.19	no change
Δ SarQoL D3 Body Composition	-0.37	-0.47	small change
Δ SarQoL D4 Functionality	-0.50	-0.62	moderate change
Δ SarQoL D5 Activities of Daily Living	-0.57	-0.56	moderate change
Δ SarQoL D6 Leisure activities	0.04	-0.04	no change
Δ SarQoL D7 Fears	-0.01	-0.01	no change
Δ SarQoL Overall score	-0.54	-0.72	moderate change
Δ SF-36 Physical Functioning	-0.44	-0.50	moderate change
Δ SF-36 Social Functioning	-0.41	-0.48	small change
Δ SF-36 Role Limitations due to Physical Health	0.02	-0.02	no change
Δ SF-36 Role Limitations due to Emotional Problems	-0.28	-0.26	small change
Δ SF-36 Mental Health	-0.27	-0.35	small change
Δ SF-36 Vitality	-0.03	-0.03	no change
Δ SF-36 Bodily Pain	-0.17	-0.15	no change
Δ SF-36 General Health	-0.23	-0.28	small change
Δ SF-36 Physical Component Summary	-0.18	-0.20	small change
Δ SF-36 Mental Component Summary	-0.29	-0.34	small change
Δ EQ-5D Mobility	0.10	-0.08	no change
Δ EQ-5D Autonomy	-0.36	NA[b]	small change
Δ EQ-5D Usual activities	0.20	-0.33	small change
Δ EQ-5D Pain	-0.07	-0.06	no change
Δ EQ-5D Anxiety	-0.19	-0.17	no change
Δ EQ-5D Utility Index	0.19	0.18	no change
Δ EQ-VAS	-0.11	-0.09	no change

[a] Interpretation of corrected SRMs: $0.20 \leq SRM < 0.49$=small change; $0.50 \leq SRM < 0.79$= moderate change; $SRM \geq 0.80$= large change

[b] Correction for SRM of EQ-5D Autonomy cannot be computed because Δ EQ-5D Autonomy at T3 is constant (all subjects responded with the same answer)

The SRM of the SarQoL® Overall score was significantly larger than the SF-36 PCS ($p = 0.005$), the EQ-5D Utility Index ($p < 0.001$) and the EQ-VAS ($p = 0.003$). The SRMs of the SarQoL® Overall score and the SF-36 MCS were not significantly different ($p = 0.150$). The results of this analysis are reported in Table 5.

Table 5: Exploration of significant differences between SRMs				
Hypothesis	Intercept	p-value	Interpretation	Larger SRM
The SRMs of SarQoL Overall score and SF-36 PCS score are significantly different.	-0.326	0.005	different	SarQoL
The SRMs of SarQoL Overall score and SF-36 MCS score are significantly different.	-0.236	0.150	not different	none
The SRMs of SarQoL Overall score and EQ-5D Utility Index are significantly different.	-0.724	<0.001	different	SarQoL
The SRMs of SarQoL Overall score and EQ-VAS are significantly different.	-0.443	0.003	not different	SarQoL
Calculation of P-values carried out with modified jack-knife method (Bessette et al., 1998)				

4.3. Correlations Between Physical Parameters and QoL

Good correlations were found between change in the SarQoL® Overall score and change in gait speed ($r = 0.50$), SPPB score ($r = 0.47$) and the chair stand test ($r = -0.42$). Good correlations were also found between change in ALM/BMI and change on the EQ-VAS ($r = -0.48$) as well as between change on the timed up-and-go test and change on the SF-36 PCS ($r = -0.44$). Acceptable correlations were found between change in gait speed and change on the SF-36 PCS ($r = 0.39$), between change on the chair stand test and change on the SF-36 PCS ($r = -0.37$) and the SF-36 MCS ($r = -0.36$). No other correlations were statistically significant. The full analysis can be found in Table 6.

Table 6: Correlations between changes in physical parameters and evolution of quality of life			
Domains	*r*	p-value	interpretation
Δ Gait speed and Δ SarQoL Overall	0.50	0.001	good
Δ Gait speed and Δ SF-36 PCS	0.39	0.017	acceptable
Δ Gait speed and Δ SF-36 MCS	0.02	0.926	NS
Δ Gait speed and Δ EQ-5D Utility Index	-0.09	0.560	NS
Δ Gait speed and Δ EQ-VAS	0.16	0.324	NS
Δ Grip strength and Δ SarQoL Overall	0.08	0.592	NS
Δ Grip strength and Δ SF-36 PCS	0.27	0.104	NS
Δ Grip strength and Δ SF-36 MCS	-0.14	0.393	NS
Δ Grip strength and Δ EQ-5D Utility Index	0.22	0.165	NS
Δ Grip strength and Δ EQ-VAS	0.08	0.626	NS
Δ SPPB and Δ SarQoL Overall	0.47	0.002	good
Δ SPPB and Δ SF-36 PCS	0.30	0.068	NS
Δ SPPB and Δ SF-36 MCS	0.25	0.131	NS
Δ SPPB and Δ EQ-5D Utility Index	0.12	0.450	NS
Δ SPPB and Δ EQ-VAS	0.12	0.450	NS
Δ ALM and Δ SarQoL Overall	0.15	0.355	NS
Δ ALM and Δ SF-36 PCS	0.04	0.829	NS
Δ ALM and Δ SF-36 MCS	0.19	0.264	NS
Δ ALM and Δ EQ-5D Utility Index	0.03	0.832	NS
Δ ALM and Δ EQ-VAS	-0.01	0.986	NS
Δ ALM/BMI and Δ SarQoL Overall	-0.02	0.901	NS

Table 6: Correlations between changes in physical parameters and evolution of quality of life			
Domains	*r*	p-value	interpretation
Δ ALM/BMI and Δ SF-36 PCS	-0.14	0.807	NS
Δ ALM/BMI and Δ SF-36 MCS	-0.11	0.537	NS
Δ ALM/BMI and Δ EQ-5D Utility Index	-0.06	0.726	NS
Δ ALM/BMI and Δ EQ-VAS	-0.48	0.002	good
Δ ALM/Ht² and Δ SarQoL Overall	0.11	0.477	NS
Δ ALM/Ht² and Δ SF-36 PCS	0.10	0.570	NS
Δ ALM/Ht² and Δ SF-36 MCS	0.22	0.192	NS
Δ ALM/Ht² and Δ EQ-5D Utility Index	0.01	0.964	NS
Δ ALM/Ht² and Δ EQ-VAS	<0.01	0.989	NS
Δ TUG and Δ SarQoL Overall	-0.17	0.279	NS
Δ TUG and Δ SF-36 PCS	-0.44	0.007	good
Δ TUG and Δ SF-36 MCS	-0.23	0.174	NS
Δ TUG and Δ EQ-5D Utility Index	-0.02	0.923	NS
Δ TUG and Δ EQ-VAS	-0.02	0.882	NS
Δ CST and Δ SarQoL Overall	-0.42	0.013	good
Δ CST and Δ SF-36 PCS	-0.37	0.032	acceptable
Δ CST and Δ SF-36 MCS	-0.36	0.040	acceptable
Δ CST and Δ EQ-5D Utility Index	-0.13	0.470	NS
Δ CST and Δ EQ-VAS	-0.11	0.546	NS
Δ = change in; *r* = correlation NS: Not significant; SPPB: Short Physical Performance Battery; ALM: Appendicular Lean Mass; ALM/BMI: ALM divided by Body Mass Index; ALM/Ht²: ALM divided by height squared; TUG: Timed Up-and-Go test; CST: Chair Stand Test			

5. Discussion

The aim of this study was to evaluate the responsiveness of the SarQoL® questionnaire in a population of older, community-dwelling, sarcopenic subjects by formulating hypotheses on the correlations between change scores, and by calculating the standardized response means. Additionally, we examined the correlations between changes in physical parameters and the evolution of the quality-of-life scores.

The results from the hypotheses reveal that the SarQoL® questionnaire has high responsiveness according to the criteria of De Boer et al., with only one hypothesis out of nine (11%) refuted [26]. The most notable results are the strong correlations found for the Overall score and domain 4 (Functionality) of the SarQoL® questionnaire, and the Physical Functioning domain of the SF-36. These correlations, respectively $r = 0.669$ and $r = 0.680$, were larger than the expected correlation of $r = 0.5$ but make sense in light of the similarity of their content and the relatively important weight of domain 4 in the calculation of the Overall score of the SarQoL® questionnaire.

The SRMs show that the change measured by the Overall score of the SarQoL® questionnaire was significantly larger than that measured by the SF-36 PCS, the EQ-5D utility index and the EQ-VAS, but not the SF-36 MCS. The absence of a significant difference between the SRM of the Overall score and the SF-36 MCS indicates a very large 95% confidence interval of the latter. The SRM obtained for the SarQoL® Overall score is in accordance with the change in physical parameters of the subjects. Participants lost approximately 10% of their original gait speed (from a median of 1.02 m/s to 0.89 m/s), and the female participants lost a median of 2 kg of grip strength in the 2-year interval. It is also interesting to note that the number of falls experienced in the year preceding the administration of the test doubled from 8 (19.0%) to 16 (38.1%). The SarQoL® Overall score more accurately reflects these changes, more so than the SF-36 and the EQ-5D.

The SarQoL® questionnaire measured an SRM indicating moderate change for domain 4 (Functionality) and domain 5 (Activities of Daily Living), highlighting that the effects of diminished muscle strength and physical performance manifest themselves most in all the physical tasks performed on a regular basis. SRMs indicating small change were reported for domain 1 (Physical and Mental Health) and domain 3 (Body Composition). The smaller SRM for domain 1 may result from the way the questions are formulated, with many more abstract concepts (energy, physical capacity, muscle mass, etc.) instead of the very relatable examples from domains 4 and 5 (climbing a flight of stairs, opening a bottle or jar, etc.). Subjects may have more difficulty finding the right answers for them because these changes are much less perceptible in absolute terms. The SRM for domain 3 (Body Composition) covers an area where drastic change is not necessarily expected given that the median age in the sample is 73 years old and that many of the age-related changes to the way one looks have already manifested themselves. Finally, three domains reported SRMs that indicate no change has occurred. Domain 6 (Leisure Activities) and domain 7 (Fears) are represented by, respectively, two and four items in the questionnaire and may be much less sensitive than domains with more items. For domain 2 (Locomotion), this reasoning does not apply. This domain asks pointed questions connected to walking (length, frequency, difficulties, tiredness, etc.), and given that the usual gait speed has significantly diminished, one would expect to see an effect in this domain. However, the questions in this domain may be affected by the phenomenon of response shift, whereby the internal standards of measurement of the subject are recalibrated.

The SF-36 reported moderate change for the domain Physical Functioning, and small change for the domains Social Functioning, Role limitations due to emotional problems, Mental Health, and General Health, and reported no change for the other domains. These results are in line with our hypothesis that the SarQoL® questionnaire, being specific to sarcopenia, should detect a greater change than generic questionnaires such as the SF-36. The EQ-5D reported a small change for the domains Autonomy and Usual Activities and no change for all other scores. This should not be surprising given the distance

between the response options for the EQ-5D, which means a significant change needs to occur in real life for it to be registered in the change scores.

Lastly, the correlations between changes in physical parameters and the changes on the different overall/composite scores revealed three good correlations for the SarQoL® Overall score, one good and two acceptable correlations for the SF-36 PCS, one acceptable correlation for the SF-36 MCS, no correlations for the EQ-5D Utility Score and one good correlation for the EQ-VAS. In general, the SarQoL® Overall score correlates well with physical performance, with good correlations for change in gait speed, SPPB and CST. However, these results should be interpreted with caution given the multidimensional nature of sarcopenia, which is unlikely to be covered in a single test.

This study has several strengths. The methodology we adopted supplied us with evidence from different sources and allowed us to show both the quality and quantity of responsiveness. We were able to draw upon the data collected within the SarcoPhAge study, which allowed us to have a moderate sample size ($n = 42$) despite the relatively low prevalence of sarcopenia. Furthermore, the SarcoPhAge study collected muscle mass data with DXA, which is, in practice, the most reliable method, and collected data on a number of tests for physical performance, which allowed us to compare the changes on several physical parameters [35].

There are, however, several limitations in this study. The SarcoPhAge study was not specifically designed to allow the evaluation of the responsiveness of the SarQoL® questionnaire, lacking both a known intervention and a transition question. A second limitation is that the primary methodology used in this study, the testing of hypotheses, has only been introduced a few years ago and that several questions about this process have not yet found a consensus, such as how many hypotheses should be tested, what percentage should be confirmed for good responsiveness and how to set the strength of the expected correlations. We have tried to address these issues by using pre-defined, specific and challenging hypotheses but recognize that this methodology should be considered an ongoing process and hope that other studies can re-evaluate our hypotheses and add their own. Lastly, the SF-36 PCS and MCS scores were used in the evaluation of the SRMs but not in the hypotheses. We acknowledge that the PCS and MCS would have made good targets for the formulation of hypotheses, but unfortunately, the choice to calculate these scores was made after the hypotheses were formulated and after the statistical manipulations had started. It was therefore impossible for us to include the PCS and MCS scores in the hypotheses. It is our hope that future responsiveness studies will include the PCS and MCS in their hypotheses.

6. Conclusions

This study contributed data on the last major psychometric property of the SarQoL® questionnaire not yet studied. The questionnaire has good responsiveness, measured both in an evaluation with hypotheses (8/9 confirmed) and by the strength of its standardized response means. The SarQoL® questionnaire appears to be the optimal tool for the assessment of quality of life in sarcopenic populations. Its use in clinical trials assessing biochemical entities for the management of sarcopenia should be recommended, as patient-related outcomes are encouraged to be included as co-primary endpoints in such studies [36]

7. References

1. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, et al. Sarcopenia: European consensus on definition and diagnosis. *Age Ageing.* 2010;39:412–423. doi: 10.1093/ageing/afq034.

2. Cao L, Morley JE. Sarcopenia is recognized as an independent condition by an International Classification of Disease, tenth revision, clinical modification (ICD-10-CM) code. *J Am Med Dir Assoc.* 2016;17:675–677. doi: 10.1016/j.jamda.2016.06.001.

3. Beaudart C, Zaaria M, Pasleau F, Reginster J-Y, Bruyère O. Health outcomes of sarcopenia: a systematic review and meta-analysis. *PLoS One.* 2017;12:e0169548. doi: 10.1371/journal.pone.0169548.

4. Chang K-V, Hsu T-H, Wu W-T, Huang K-C, Han D-S. Is sarcopenia associated with depression? A systematic review and meta-analysis of observational studies. *Age Ageing.* 2017;46:1–9. doi: 10.1093/ageing/afx094.

5. Beaudart C, Reginster J-Y, Geerinck A, Locquet M, Bruyère O. Current review of the SarQoL®: a health-related quality of life questionnaire specific to sarcopenia. *Expert Rev Pharmacoecon Outcomes Res.* 2017;17:335–341. doi: 10.1080/14737167.2017.1360768.

6. Rizzoli R, Reginster JY, Arnal JF, et al. Quality of life in sarcopenia and frailty. *Calcif Tissue Int.* 2013;93:101–120. doi: 10.1007/s00223-013-9758-y.

7. Beaudart C, Biver E, Reginster J-Y, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. *Age Ageing.* 2015;44:960–966. doi: 10.1093/ageing/afv133.

8. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63:737–745. doi: 10.1016/j.jclinepi.2010.02.006.

9. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol.* 1992;45:1341–1345. doi: 10.1016/0895-4356(92)90194-R.

10. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;40:171–178. doi: 10.1016/0021-9681(87)90069-5.

11. Beaudart C, Biver E, Reginster J-Y, et al. Validation of the SarQoL, a specific health-related quality of life questionnaire for sarcopenia. *J Cachexia Sarcopenia Muscle.* 2017;8:238–244. doi: 10.1002/jcsm.12149.

12. Beaudart C, Edwards M, Moss C, et al. English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. *Age Ageing.* 2017;46:271–277.

13. Ildiko GA, Mihai G, Beaudart C, et al. Psychometric performance of the Romanian version of the SarQoL®, a health-related quality of life questionnaire for sarcopenia. *Arch Osteoporos.* 2017;12:103. doi: 10.1007/s11657-017-0397-1.

14. Geerinck A, Scheppers A, Beaudart C, et al. Translation and validation of the Dutch SarQoL®, a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact. 2018. http://www.ismni.org/jmni/accepted/jmni_aa_BAUTMANS.pdf.

15. Beaudart C, Reginster J-Y, Petermans J, et al. Quality of life and physical components linked to sarcopenia: the SarcoPhAge study. *Exp Gerontol.* 2015;69:103–110. doi: 10.1016/j.exger.2015.05.003.

16. Locquet M., Beaudart C., Bruyère O., Kanis J. A., Delandsheere L., Reginster J.-Y. Bone health assessment in older people with or without muscle health impairment. *Osteoporosis International.* 2018;29(5):1057–1067. doi: 10.1007/s00198-018-4384-1.

17. Beaudart C, Reginster J-Y, Petermans J, Bruyère O. Qualité de vie du patient sarcopénique: apport de l'étude liégeoise SarcoPhAge. *Geriatr Psychol Neuropsychiatr Vieil.* 2015;13:391–395.

18. Locquet M, Beaudart C, Reginster J-Y, et al. Prevalence of concomitant bone and muscle wasting in elderly women from the SarcoPhAge cohort: preliminary results. *J Frailty Aging.* 2017;6:18–23.

19. The SarQoL questionnaire [Internet]. http://www.sarqol.org. Accessed 7 Sep 2018.

20. Ware JJ, Sherbourne C. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30:473–483. doi: 10.1097/00005650-199206000-00002.

21. Ware JJ, Konsinski M. *SF-36 physical and mental health summary scales: a manual for users of version 1.* Lincoln: QualityMetric; 2001.

22. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in orthopaedics: a brief guide. *J Bone Jt Surg.* 2015;97:1628–1634. doi: 10.2106/JBJS.O.00030.

23. Van Reenen M, Oppe M. EQ-5D-3L User guide. Basic information on how to use the EQ-5D-3L instrument. 2015. EuroQol Group. https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-3L_UserGuide_2015.pdf.

24. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33:337–343. doi: 10.3109/07853890109002087.

25. De Vet HCW, Knol DL, Terwee CB, Mokkink LB. *Measurements in medicine.* Cambridge: Cambridge University Press; 2011.

26. De Boer M, Moll A, de Vet H, Terwee CB, Völker-Dieben HJM, van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt.* 2004;24:257–273. doi: 10.1111/j.1475-1313.2004.00187.x.

27. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol.* 2000;53:459–468. doi: 10.1016/S0895-4356(99)00206-1.

28. Middel B, Van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *Int J Integr Care.* 2002;2:e15. doi: 10.5334/ijic.65.

29. Cohen J. Quantitative methods in psychology. *Psychol Bull.* 1992;112:155–159. doi: 10.1037/0033-2909.112.1.155.

30. Bessette L, Sangha O, Kuntz K, et al. Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. *Med Care.* 1998;36:491–502. doi: 10.1097/00005650-199804000-00005.

31. Angst F, Verra ML, Lehmann S, Aeschlimann A. Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. *BMC Med Res Methodol.* 2008;8:26. doi: 10.1186/1471-2288-8-26.

32. Evans JD. *Straightforward statistics for the behavioral sciences.* Pacific Grove: Brooks/Cole; 1996.

33. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39:175–191. doi: 10.3758/BF03193146.

34. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21:651–657. doi: 10.1007/s11136-011-9960-1.

35. Buckinx Fanny, Landi Francesco, Cesari Matteo, et al. Pitfalls in the measurement of muscle mass: a need for a reference standard. *Journal of Cachexia, Sarcopenia and Muscle.* 2018;9(2):269–278. doi: 10.1002/jcsm.12268.

36. Reginster J-Y, Cooper C, Rizzoli R, et al. Recommendations for the conduct of clinical trials for drugs to treat or prevent sarcopenia. *Aging Clin Exp Res.* 2016;28:47–58. doi: 10.1007/s40520-015-0517-y.

Chapter 2: Measurement error of the SarQoL questionnaire

Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: an analysis of subjects from 9 validation studies

Published as:

Geerinck A, Alekna V, Beaudart C, Bautmans I, Cooper C, et al. (2019) Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: An analysis of subjects from 9 validation studies. PLOS ONE 14(4): e0216065. https://doi.org/10.1371/journal.pone.0216065

1. Abstract

Objectives : The Sarcopenia Quality of Life (SarQoL) questionnaire, a sarcopenia-specific patient-reported outcome measure, evaluates quality of life with 55 items. It produces 7 domain scores and 1 overall quality of life score, all between 0 and 100 points. This study aims to contribute to the interpretation of the SarQoL scores by calculating the standard error of measurement (SEM) and smallest detectable change (SDC) in a sample of subjects from 9 studies.

Methods : Subjects from 9 studies (conducted in Belgium, Brazil, Czech Republic, England, Greece, Lithuania, Poland and Spain) were included. The SEM, a measure of the error in the scores that is not due to true changes, was calculated by dividing the standard deviation of the difference between test and retest scores (SD_{diff}) by $\sqrt{2}$. The SDC, defined as change beyond measurement error, was calculated by multiplying SD_{diff} by 1.96. Bland-Altman plots were assessed for the presence of systematic errors.

Results : A total of 278 sarcopenic subjects, aged 77.67 ± 7.64 years and 61.5% women, were included. The SEM for the overall SarQoL score ranged from 0.18 to 4.20 points for the individual studies, and was 2.65 points when all subjects were analyzed together. The SDC for the overall score ranged from 0.49 to 11.65 points for the individual studies, and was 7.35 points for all subjects. The Bland-Altman plots revealed no systematic errors in the questionnaire.

Conclusion : This study shows that, for individual subjects, a change in overall quality of life of at least 7.35 points (on a scale from 0 to 100) would have to be observed to confirm that a true change, beyond measurement error, has occurred. It also demonstrated that the SarQoL questionnaire is a precise instrument, with the observed scores within less than 3 points of the theoretical “true score”.

2. Introduction

Sarcopenia, often described as the age-related loss of muscle mass and strength, and defined by the European Working Group on Sarcopenia in Older People (EWGSOP2) as "a progressive and generalised skeletal muscle disorder that is associated with increased likelihood of adverse outcomes including falls, fractures, physical disability and mortality", has been the subject of increased scientific attention as its prevalence and consequences have become more known [1]. Sarcopenia is confirmed to be present when a patient is diagnosed with low muscle strength and low muscle mass. When low physical performance is also established, that person is diagnosed with severe sarcopenia [1]

A systematic review conducted in 2014 which estimated the prevalence of sarcopenia diagnosed with the EWGSOP-algorithm in older community-dwelling adults found a range of 1 to 29% (up to 30% in women), while a recent meta-analysis which included 35 articles and a total of 58404 healthy subjects aged 60 years and older found an overall prevalence of sarcopenia of 10% (95% CI: 8–12%) in men and 10% (95% CI: 8–13%) in women diagnosed with the EWGSOP, the International Working Group on Sarcopenia (IWGS) or the Asian Working Group for Sarcopenia (AWGS) definitions [2,3]. It should be mentioned that the prevalence of sarcopenia varies greatly depending on the definition used, as demonstrated by Beaudart et al., who applied 6 different diagnostic criteria for sarcopenia to a single cohort of subjects and found a prevalence rate from 4.39% to 32.8% [4].

Projections about the future prevalence of sarcopenia (as diagnosed by the EWGSOP-criteria) in the European Union (EU28) predict a rise from 10.9 million people in 2016 to 18.7 million in 2045 on the low end and from 19.7 million to 32.3 million people on the high end [5]. Sarcopenia is a major public health problem and its impact will continue to grow, which should incite policy makers to act.

The available evidence concerning the impact and association of sarcopenia with several health outcomes has been steadily growing during the last decade. A systematic review and meta-analysis published in 2017 provided a comprehensive summary of what is currently known on the subject. This review included 17 prospective studies in which sarcopenia was diagnosed according to the EWGSOP guidelines. The authors found a higher risk for mortality (OR = 3.596; 95% CI = 2.96–4.37) and functional decline (OR = 3.03; 95% CI = 1.80–5.12) as well as a higher rate of falls and a higher incidence of hospitalization. The evidence on the incidence of fractures and the length of hospital stay was inconclusive [6].

The subject of quality of life in sarcopenia has mostly been examined using generic questionnaires such as the Short-Form 36-Item (SF-36) and the EuroQoL 5-Dimension (EQ-5D) [7]. Recently, a new instrument, the Sarcopenia Quality of Life (SarQoL) questionnaire has become available. It is specifically designed to measure quality of life in sarcopenic, community-dwelling individuals aged 65

years or older and was developed in 2013–2015 by Beaudart et al. [8]. It has, to date, been translated into more than 20 languages [8].

The psychometric properties of the SarQoL questionnaire have been evaluated and published for 6 language-versions: the original questionnaire in French, and the English, Dutch, Polish, Romanian and Greek translations [9–14]. These examined the discriminative power, internal consistency, construct validity, test-retest reliability and the presence of floor or ceiling effects. These 6 studies found that the questionnaire can discriminate between sarcopenic and non-sarcopenic participants, with the former having significantly lower scores for the 7 domains and the overall score, and that the questionnaire possesses good internal consistency (Cronbach's alpha of 0.87, 0.88, 0.95, 0.92, 0.88 and 0.96). These studies also confirmed the construct validity of the SarQoL questionnaire with the help of hypotheses on correlations between the questionnaire and the SF-36 and EQ-5D, and demonstrated that the SarQoL questionnaire has an excellent test-retest reliability (intraclass correlation coefficient/ICC = 0.91, 0.95, 0.99, 0.98 and 0.96) [9–14]. Lastly, floor and ceiling effects were absent from all 6 published validation studies [9–14]. These results provide convincing evidence for the validity and reliability of the SarQoL questionnaire for the evaluation of quality of life in sarcopenic, community-dwelling older people.

However, until now, the standard error of measurement (SEM) and the smallest detectable change (SDC) of the SarQoL questionnaire have not yet been calculated. These parameters supply important information on the reliability of the instrument in question by indicating the range in which the theoretical "true" score lies; and supply context when interpreting data from longitudinal measurements by indicating by how much the score needs to change before one can be reasonably certain that a true change has occurred. Clinicians and researchers could use the values for SEM and SDC as a yardstick in the interpretation of the SarQoL scores, whether obtained in clinical practice or as part of a research project. The results of this study should prove particularly valuable in the interpretation of data from interventional clinical trials, and will hopefully expedite the adoption of this PROM in clinical trials [15].

The primary objective of this study is to determine the SEM and SDC of the SarQoL questionnaire in a sample of subjects from 9 international validation studies. The secondary objectives are to examine the measurement error of the questionnaire with the help of a Bland-Altman analysis, and to update the results previously obtained for the test-retest reliability of the SarQoL questionnaire in the complete sample.

3. Material and methods

This study combined data from 9 cohorts in 8 different countries that were established to test the psychometric properties of the SarQoL questionnaire after translation into the local language. The team behind the SarQoL questionnaire have made a concerted effort to widen the reach of the questionnaire by having it translated into a multitude of languages. To accomplish this, they have partnered with researchers from a host of countries and language groups, who were able and willing to undertake a translation of the questionnaire. The local teams responsible for the translations were also encouraged to carry out a validation study of the translation they produced, if feasible. A considerable number of them undertook this effort, although not all validations have been published. The researchers from 9 validation studies that had the necessary data for the current analysis were contacted and agreed to share their data. All the included studies obtained approval from their local ethics committees, and written informed consent from their participants.

Population

Subjects were included in the 9 validation studies if they were 60 years of age or older and community-dwelling. For this analysis, we included all subjects who were diagnosed as being sarcopenic, who completed the SarQoL questionnaire twice and reported that their health had been stable in the interval between the two administrations.

The SarQoL questionnaire

The analyses in this article center around the test-retest data for the SarQoL questionnaire collected by the 9 included studies. The SarQoL questionnaire is a patient-reported outcome measure (PROM) designed specifically for use with sarcopenic, community-dwelling subjects 65 years of age or older. The questionnaire consists of 55 items distributed over 22 questions, with the items categorized into 7 domains of health-related quality of life (HRQoL). These domains are: “Physical and Mental Health” (D1), “Locomotion” (D2), “Body Composition” (D3), “Functionality” (D4), “Activities of Daily Living” (D5), “Leisure activities” (D6), and “Fears” (D7). Apart from the domain scores, an Overall score for quality of life is also calculated. All scores are situated on a scale from 0 to 100, with 0 being the worst possible quality of life, and 100 the best possible. The questionnaire is auto-administered and takes about 10 minutes to complete [9]. More information on the SarQoL questionnaire and the different language-specific versions can be found on www.sarqol.org.

Test-retest reliability

The test-retest reliability of a questionnaire quantifies the extent to which a questionnaire produces the same scores during repeated measurements, provided that the participants’ health remains stable. It is measured by the intraclass correlation coefficient (ICC) under a 2-way mixed model with absolute

agreement specified, and its associated 95% confidence interval. A questionnaire is considered reliable if the obtained ICC values are greater than 0.70 [16].

Standard error of measurement

The standard error of measurement has been defined as "the determination of the amount of variation or spread in the measurement errors for a test" [17]. The SEM is considered to be a parameter for the amount of measurement error present in an instrument, and is subsequently an indicator of the reliability of said instrument. Much like the interpretation of the standard deviation around the mean value, the SEM can be used to provide a range around the observed value within which the theoretical "true" value lies. The interval between plus and minus 1 SEM provides a probability of 68% of containing the true value. For ± 2 SEM the probability becomes 95% and for ± 3 SEM we end up with 99% probability.

Smallest detectable change

The smallest detectable change is defined as the change in the instrument's score beyond measurement error [18]. This means that the SDC provides a value for the minimum change that needs to be observed in order to be confident that the observed change is real and not, potentially, a product of measurement error in the instrument. The SDC can be calculated for individual subjects (SDC_{ind}) as well as for comparisons of mean scores between groups (SDC_{group}) [18]. Both provide utility: The SDC_{ind} can be used in clinical practice or to label individual subjects in a study sample as either changed or unchanged. The SDC_{group} provides an aid to the interpretation of mean scores of groups. This can lend greater credibility to the results of interventional trials that use the SarQoL questionnaires, and that want to know whether quality of life has changed in the intervention and control group as a whole.

Bland-Altman analysis

The Bland-Altman plot provides a visual representation of the presence of systematic errors in an instrument. The Bland-Altman plot is based around three variables: the mean systematic difference between test and retest scores ($\bar{d}$), and the upper and lower limit of agreement, which span 95% of observations, assuming that the values for the difference between test and retest scores are distributed normally [18,19]. These variables are integrated into a scatter plot where the difference between test and retest values is put on the Y-axis and the average of the test and retest values is put on the X-axis.

Statistical analysis

Data were analyzed using IBM SPSS Statistics, version 24.0.0.0 for Windows (Armonk, NY: IBM Corp). The distribution of the variables was determined by examining the histogram, the quantile-quantile-plot, the Shapiro-Wilk test and the difference between mean and median. Variables that are normally distributed are reported as mean ± standard deviation and non-normal variables as median (25th percentile– 75th percentile). Nominal variables are reported as absolute (n) and relative frequencies (%). Differences between groups with regards to clinical characteristics were examined with one-way anova analysis for continuous variables and chi-squared test for nominal variables.

The SEM was calculated by first creating a variable for the difference between the score obtained during the first and the second administration (test score—retest score = Difference). Next, we calculated the standard deviation of Difference in our sample ($SD_{difference}$) and divided the obtained value by the square root of 2 (SEM = $SD_{diff} / \sqrt{2}$) [18,20]. The SDC_{ind} was calculated with the formula [$SDC_{ind} = 1.96 * \sqrt{2} * SEM$], and the SDC_{group} was calculated by dividing the SDC_{ind} by the square root of the number of subjects in the sample ($SDC_{ind} / \sqrt{n}$) [18]. The ICC was calculated with a 2-way mixed model and absolute agreement specified. The mean difference score ($\bar{d}$) was calculated by calculating the mean of the differences between test and retest scores for all subjects [Mean(test score—retest score)]. The 95% limits of agreement were calculated with the formula [± (1.96 * SD_{diff})] [18,21]. Bland-Altman plots were created in SPSS following the instructions given in IBM tech-note n° 19420 [22]. Results were considered significant at $p \leq 0.05$.

4. Results

4.1. Characteristics of included studies

Information on the diagnosis of sarcopenia and the characteristics of the test-retest administration are given in Table 1.

4.2. Clinical characteristics

The 278 participants included in the analysis had a mean age of 77.67 ± 7.64 years, ranging from 60 to 98 years old. The majority of subjects were women, namely 171 participants or 61.5% of the complete sample. The participants had a mean body mass index of 25.57 ± 4.40 kg/m2, spanning the whole gambit from underweight to morbidly obese with a minimum value of 17.42 kg/m2 and a maximum value of 46.10 kg/m2. In terms of prescription drug use, the subjects took on average 4.78 ± 2.71 drugs (range: 0–13), linked to the number of comorbidities which was 3.59 ± 2.01 (range: 0–11). Clinical characteristics are reported in Table 2.

As expected, one-way anova analyses and chi-squared test revealed that the 9 studies differed significantly in terms of clinical characteristics. The results from these post-hoc analyses can be found in S1–S5 Tables in annex.

The test-retest reliability of the SarQoL questionnaire in the complete sample resulted in an ICC of 0.969 (95% CI = 0.961–0.975) for the Overall score. Of the individual domains, 4 obtained an ICC higher than 0.9, namely domain 1, 2, 4 and 5, and all obtained ICC's higher than 0.7. The detailed results for the test-retest reliability can be found in Table 3.

Table 1. Characteristics of included studies.

	Sarcopenia diagnosis				Time between test and retest administration	Mode of administration	
	Sarcopenia definition	Muscle mass assessment	Muscle strength assessment	Physical performance assessment		Test	Retest
Belgium (Dutch)	EWGSOP	BIA	Martin-Vigorimeter	Gait speed	2 weeks	At study center	At home
Belgium (French) [9]	EWGSOP	DXA	Hand dynamometer	SPPB	2 weeks	At study center	At home
Brazil	EWGSOP	DXA	Hand dynamometer	Gait speed	2 weeks	At home	At home
Czech Republic [23]	FNIH	DXA	Hand dynamometer	SPPB	2 weeks	At home or at study center without staff present	At home or at study center without staff present
England [10]	EWGSOP	DXA	Hand dynamometer	Gait speed	2 weeks	At home	At home
Greece [14]	EWGSOP	BIA	Hand dynamometer	Gait speed	2 weeks	At study center	At study center
Lithuania	EWGSOP	DXA	Hand dynamometer	SPPB	2 weeks	At study center	At study center
Poland [13]	EWGSOP	Lee equation [24]	Hand Dynamometer	Not performed	2 weeks	At study center	At study center
Spain	FNIH	DXA	Hand dynamometer	SPPB	2 weeks	At study center	At home

EWGSOP: European Working Group on Sarcopenia in Older People; BIA: bioelectrical impedance analysis; DXA: dual-energy x-ray absorptiometry; FNIH: Foundation for the National Institutes of Health

Table 2. Clinical characteristics for individual studies – mean ± SD or n(%).										
	All	Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain
n	278	26	29	12	48	10	58	50	30	15
Age (years)	77.67 ± 7.64	81.00 ± 5.88	77.03 ± 6.58	70.75 ± 6.57	82.96 ± 6.05	78.90 ± 2.56	80.18 ± 6.42	72.10 ± 7.71	73.82 ± 7.06	77.60 ± 6.27
Gender										
Female	171 (61.5)	12 (46.2)	19 (65.5)	6 (50.0)	37 (77.1)	3 (30.0)	28 (48.3)	37 (74.0)	19 (63.3)	10 (66.7)
Body mass index (kg/m^2)	25.57 ± 4.40	26.71 ± 4.75	23.16 ± 3.19	24.84 ± 4.32	29.16 ± 5.78	24.00 ± 2.73	24.62 ± 2.54	24.05 ± 3.39	27.01 ± 4.46	24.17 ± 1.99
Drugs (n)	4.78 ± 2.71	3.81 ± 2.62	6.72 ± 2.76	7.25 ± 1.55	6.27 ± 3.30	6.00 ± 2.45	4.36 ± 1.25	3.50 ± 1.28	2.70 ± 2.84	5.13 ± 2.75
Concomitant illnesses (n)	3.59 ± 2.01	2.48 ± 1.64	4.93 ± 2.36	4.17 ± 1.59	5.79 ± 1.47	NA	2.98 ± 0.78	2.96 ± 1.01	1.60 ± 1.85	3.80 ± 2.04

Table 3. Results for complete analysis (n=278).

	Test scores	Retest scores	ICC (95% CI)	$\bar{d}$ (95% CI)	SD_{diff}	SEM	SDC_{ind}	SDC_{group}	95% LoA
D1: Physical & mental health	56.56 ± 17.00	57.42 ± 17.12	0.915 (0.894 ; 0.933)	0.86 (0.04 ; 1.68)	6.98	4.94	13.68	0.82	-12.82 ; 14.54
D2: Locomotion	54.95 ± 21.40	54.88 ± 21.54	0.944 (0.929 ; 0.955)	-0.07 (-0.93 ; 0.78)	7.23	5.11	14.17	0.85	-14.24 ; 14.1
D3: Body composition	55.36 ± 16.91	56.10 ± 17.18	0.836 (0.797 ; 0.869)	0.74 (-0.41 ; 1.89)	9.74	6.89	19.09	1.14	-18.35 ; 19.83
D4: Functionality	62.31 ± 17.08	62.70 ± 16.61	(0.952 (0.939 ; 0.962)	0.39 (-0.23 ; 1.01)	5.24	3.71	10.27	0.62	-9.88 ; 10.66
D5: Activities of daily living	55.55 ± 17.33	55.40 ± 17.73	0.915 (0.894 ; 0.933)	-0.15 (-1.00 ; 0.70)	7.23	5.11	14.17	0.85	-14.32 ; 14.02
D6: Leisure activities	37.61 ± 17.83	37.00 ± 19.23	0.754 (0.698 ; 0.800)	-0.59 (-2.13 ; 0.94)	13.04	9.22	25.56	1.53	-26.15 ; 24.97
D7: Fears	78.98 ± 17.47	78.96 ± 17.13	0.783 (0.733 ; 0.825)	-0.02 (-1.37 ; 1.33)	11.42	8.08	22.38	1.34	-22.4 ; 22.36
Overall score	57.71 ± 14.97	57.89 ± 15.03	0.969 (0.961 ; 0.975)	0.18 (-0.26 ; 0.63)	3.75	2.65	7.35	0.44	-7.17 ; 7.53

ICC= intraclass correlation coefficient; $\bar{d}$= mean difference score; CI= confidence interval; SD_{diff}= standard deviation of difference scores; SEM= standard error of measurement; SDC_{ind}= smallest detectable change for individual subject; SDC_{group}= smallest detectable change for group LoA= limits of agreement

4.3. Standard error of measurement

The SEM for the Overall score of the SarQoL questionnaire in the complete sample is 2.65 points. This means that one can be 68% confident (± 1 SEM) that the 'true' score of a subject can be found between -2.65 and +2.65 points from the observed score, and 95% confident (± 2 SEM) that the 'true' score is situated between -5.3 and +5.3 points of the observed score. The SEM for the different domains of the SarQoL questionnaire in the complete sample varied between 3.71 for domain 4 and 9.22 points for domain 6. The SEM-values for the complete sample can be found in Table 3, while the SEM-values for the individual included studies are available in Table 4.

4.4. Smallest detectable change

The SDCind for the Overall score of the SarQoL questionnaire in the complete sample is 7.35 points. This means that the Overall quality of life score of an individual would have to change with at least 7.35 points (on a scale of 0 to 100) before the observed change can be considered to be a true change in the quality of life of a subject, and not potentially a result of measurement error. The SDCind for the 7 domains of the SarQoL questionnaire goes from a minimum value of 10.27 points for domain 4 to a maximum value of 25.56 points for domain 6. The SDCgroup for the Overall score in the complete sample is 0.44 points. The SDC-values for the complete sample can be found in Table 3. The SDC-values for the individual included studies are available in Table 4.

4.5. Bland-Altman analysis

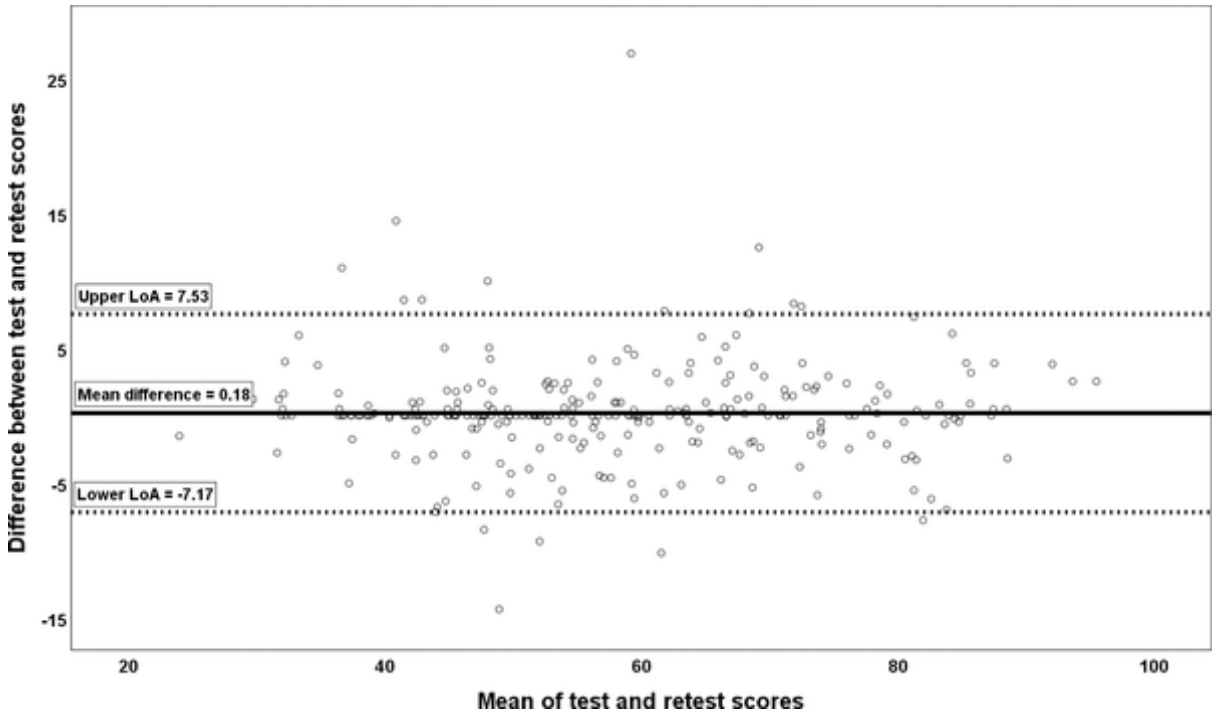
The mean difference score in the complete sample for the Overall score of the SarQoL questionnaire is 0.18 points (95% CI = -0.26; 0.63) which shows that there is no systematic bias between the two administrations of the questionnaire because the confidence interval contains zero. The mean difference scores in the complete sample for the 7 domains are not significant (95% CI contains zero) for domains 2, 3, 4, 5, 6 and 7, once again indicating the absence of systematic bias. One domain in the complete sample does have a small but significant mean difference score, namely domain 1 [0.86 points (0.04; 1.68)], indicating the presence of a very slight systematic error. The full results of the Bland-Altman analysis are detailed in Table 3. A Bland-Altman plot for the Overall score in the complete sample is provided as Fig 1.

Table 4. SEM and SDC for individual studies.

		Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain
SEM	D1	6.57	6.50	3.69	7.02	9.48	0.54	3.04	2.61	3.08
	D2	6.13	8.26	3.63	6.91	2.89	0.68	4.41	1.19	5.14
	D3	7.81	10.59	1.70	10.05	6.37	1.57	7.09	1.69	4.14
	D4	3.75	5.75	3.16	4.65	4.77	0.53	3.82	1.97	3.28
	D5	7.38	8.07	2.45	4.65	6.30	0.54	6.36	2.92	2.51
	D6	14.70	12.98	7.09	10.68	12.14	0	7.29	0.00	7.52
	D7	16.26	20.85	0.00	5.72	7.74	2.50	10.12	3.23	4.05
	Overall	2.54	4.06	2.17	2.86	4.20	0.18	3.34	1.07	1.73
SDC_{ind}	D1	18.21	18.30	10.24	19.45	26.28	1.49	8.41	7.23	8.54
	D2	16.99	22.79	10.07	19.15	8.00	1.89	12.22	4.67	14.26
	D3	21.71	29.21	4.71	27.86	17.65	4.35	19.65	7.43	11.46
	D4	16.15	10.40	8.75	12.89	13.23	1.47	10.60	5.46	9.10
	D5	20.46	22.27	6.79	12.89	17.47	1.51	17.62	8.11	6.96
	D6	40.76	35.98	19.64	29.61	33.65	0	20.22	0.00	20.85
	D7	45.07	29.43	0.00	15.85	21.45	6.94	28.05	8.95	11.21
	Overall	7.05	11.34	6.00	7.92	11.65	0.49	9.24	2.96	4.81
SDC_{group}	D1	3.57	3.40	2.95	2.81	8.31	0.20	1.19	1.32	2.21
	D2	3.33	4.23	2.91	2.76	2.53	0.25	1.73	0.85	3.68
	D3	4.26	5.42	1.36	4.02	5.58	0.57	2.78	1.36	2.96
	D4	3.17	1.93	2.53	1.86	4.18	0.19	1.50	1.00	2.35
	D5	4.01	4.14	1.96	1.86	5.53	0.20	2.49	1.48	1.80
	D6	7.99	6.68	5.67	4.27	10.64	0	2.86	0.00	5.38
	D7	8.84	5.47	0.00	2.29	6.78	0.91	3.97	1.63	2.90
	Overall	1.38	2.11	1.73	1.14	3.68	0.06	1.31	0.54	1.24

SEM: standard error of measurement; SDC_{ind}: smallest detectable change for individual subjects; SDC_{group}: smallest detectable change for groups

Fig 1. Bland Altman plot for the SarQoL overall score in the complete sample.



5. Discussion

In this study, values were obtained for the standard error of measurement and the smallest detectable change of the SarQoL questionnaire in a sample of 278 sarcopenic subjects hailing from 8 different countries and 9 different language-groups. The measurement error inherent to the questionnaire was found to be 2.65 points, and the minimum change needed to be confident that a real change in overall quality of life has occurred for an individual patient was 7.35 points. Systematic bias was further investigated with the method of Bland & Altman, and showed that there is no systematic bias for almost all domains (with domain 1 as the exception) and the overall score of the SarQoL questionnaire.

The SEM for the Overall score of the SarQoL questionnaire of 2.65 points represents 2.65% of the possible range of the Overall score (0–100) and 3.81% of the observed range of the SarQoL scores in the complete sample (min = 24.74; max = 94.22; range = 69.48).

This value for the standard error of measurement compares favorably with SEMs for the SF-36, the most frequently used quality of life questionnaire in sarcopenic populations. Hart found a SEM of 4 points for the Physical Component Summary (PCS–range: 0–100 points) and the Mental Component Summary (MCS–range: 0–100 points) of the SF-36 in a population of 68 subjects with a variety of

orthopedic impairments [25] and Palmer calculated a SEM of 3.09 points for the PCS and 5.57 points for the MCS in a population of 233 subjects with joint hypermobility [26]. Other studies looked at the SEM for the 8 domains of the SF-36 (all range between 0–100 points), and found SEMs between 8.82 and 34.52 points in 106 women undergoing surgery for breast cancer [27], between 13.2 and 44.7 points in 92 subjects with neck pain [28], between 6.82 and 11.22 points for 628 subjects undergoing foot or ankle surgery [29], and between 11 and 32 points for 515 subjects undergoing orthopedic surgery [30]. While these have been calculated in populations that differ from ours, they show a trend for higher standard errors of measurement compared to the SarQoL questionnaire.

The SDC of the Overall score (7.35 points) of the SarQoL questionnaire is similar to the SDC found for the PCS and MCS of the SF-36. Palmer obtained SDCs of 8.56 points for the PCS and 15.44 points for the MCS, while Hart found SDCs of 9 points both for the PCS and MCS [25,26].

The results for the 7 domains of the SarQoL questionnaire in the complete sample show considerably higher SEM and SDC values compared to the Overall score. These values seem to correspond roughly to the number of items in each domain. When looking at the 3 domains with the least number of items (D6: 2 items; D3: 3 items; D7: 4 items), the largest SEM and SDC values are found, between 6.89 and 9.22 points for the SEM and between 19.09 and 25.51 points for the SDC. This contrasts with the 4 domains with larger numbers of items (D1: 8 items, D2: 9 items; D4: 14 items; D5: 15 items) which have SEM-values between 3.71 and 5.11 points and SDC-values between 10.27 and 14.17 points. It is not surprising that a domain score based on a larger number of items has greater precision and lower variability, represented by the standard deviation of the difference between test and retest scores.

The detailed breakdown of the SEM and SDC values obtained for the individual studies included in the analysis demonstrates the fact that the SEM and SDC depend on the population in which they are calculated. There is considerable variability between the studies, but not within the studies (i.e. studies with lower or higher SEM and SDC values are so for all the domains and the Overall score, and do not report low values for one domain and high for another). On the lower end are found the studies carried out in Lithuania, Poland and Spain, in the middle those carried out in Belgium (Dutch), Brazil and the Czech Republic and on the higher end those carried out in Greece, England and Belgium (French). We were unable to formulate convincing hypotheses that could begin to explain why certain studies reported lower or higher values for SEM and SDC based on the clinical or study characteristics. It is likely that the observed variation is just the manifestation of the fact that the SEM and SDC are specific to the population in which they have been measured.

The Bland-Altman analysis, detailed in Table 3 and visually represented for the Overall score in Fig 1, shows that a very small systematic bias exists in only one domain. It is unlikely that this systematic bias is clinically relevant because of its small confidence interval and the fact that the lower end of the interval is extremely close to zero (95% CI = 0.04; 1.68). These results mean that clinicians and

researchers can have confidence when administering the questionnaire that the results will not be distorted by systematic bias.

The analysis of the test-retest reliability in the complete sample confirmed the results from previous validation studies. The significantly larger sample in the combined analysis means that the confidence intervals found are much narrower than has been obtained previously. These results should inspire confidence that the SarQoL questionnaire is a reliable instrument.

The main strength of this study is the fact that we were able to assemble a relatively large and heterogeneous sample (n = 278) of sarcopenic participants. This has the important advantage that the values calculated for the SEM and SDC are not dependent on a particular population, and could thus be more confidently used as a benchmark in future studies. The studies included in the analysis used different diagnostic criteria and instruments to establish sarcopenia. This is an advantage in this particular situation because the SEM and SDC values found in this study are not specific to a single definition of sarcopenia, but should be valid for different diagnostic criteria for sarcopenia, measured with different instruments. By combining multiple samples that differ with regards to clinical characteristics, we were able to find a middle ground and values for the SEM and SDC that are not highly specific to a single population. The sample size, which would be very difficult to gather in a single study, increased the accuracy of the standard deviation of the difference between test and retest score. Given that this parameter is key in the calculation of the SEM and SDC, the accuracy of these two parameters was enhanced by the large sample size. Because the SarQoL questionnaire has undergone validation in multiple languages, we were able to use test-retest data to calculate the SEM and the SDC, which is the preferred method because it takes into account biological variation, change of mood or concentration and other circumstances [18]. Since the data on which this study was based incorporates these elements and their subsequent influence on the SarQoL score, they have greater credibility than if other methods for calculating the SEM and SDC were to have been used.

There are, however, also limitations to this study. Although the researchers who carried out the individual translation and validation studies received the same guidance on the preferred design and conduct of these studies, local circumstances sometimes led them to deviate with regards to measurement of sarcopenia components (muscle mass, muscle strength and physical performance). Therefore, the methods for establishing the presence of sarcopenia are not standardized. This could, however, also be regarded as an opportunity in that we have a mix of subjects in the combined sample that represent a spectrum of methods and instruments. Secondly, because of the original purpose of the included studies, only the SarQoL questionnaire was administered twice, to calculate the test-retest reliability. It would have been preferable to compare the SEM and SDC of the SarQoL questionnaire to values for the SF-36 and the EQ-5D measured in the same populations. But, since this data does not

exist, we feel that a comparison to data from the literature was the second-best option and does provide a valid frame of reference.

6. Conclusion

The current study, which analyzed a sample of 278 subjects from 9 validation studies, obtained a standard error of measurement of 2.65 points and a smallest detectable change of 7.35 points for the Overall score of the SarQoL questionnaire. These values can be applied in future longitudinal research to evaluate the veracity of measured changes.

7. Annexes

Table S1: One-way Anova (Tukey) for Age									
	Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain
Belgium (Dutch)	1								
Belgium (French)	0.382	1							
Brazil	<0.001	0.124	1						
Czech Republic	0.950	0.005	<0.001	1					
England	0.995	0.997	0.093	0.696	1				
Lithuania	1.000	0.469	<0.001	0.426	1.000	1			
Greece	<0.001	0.038	0.999	<0.001	0.073	<0.001	1		
Poland	0.002	0.629	0.909	<0.001	0.461	0.001	0.969	1	
Spain	0.805	1.000	0.154	0.132	1.000	0.913	0.107	0.667	1

Table S2: One-way Anova (Tukey) for BMI									
	Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain
Belgium (Dutch)	1								
Belgium (French)	0.030	1							
Brazil	0.917	0.949	1						
Czech Republic	0.223	<0.001	0.025	1					
England	0.663	1.000	1.000	0.007	1				
Lithuania	0.395	0.796	1.000	<0.001	1.000	1			
Greece	0.132	0.989	1.000	<0.001	1.000	0.998	1		
Poland	1.000	0.008	0.806	0.334	0.496	0.165	0.038	1	
Spain	0.565	0.997	1.000	0.001	1.000	1.000	1.000	0.370	1

Table S3: One-way Anova (Tukey) for number of drugs									
	Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain
Belgium (Dutch)	1								
Belgium (French)	0.000	1							
Brazil	0.001	0.999	1						
Czech Republic	0.001	0.996	0.934	1					
England	0.234	0.996	0.946	1.000	1				
Lithuania	0.986	0.000	0.004	0.001	0.521	1			
Greece	1.000	0.000	0.000	0.000	0.059	0.615	1		
Poland	0.710	0.000	0.000	0.000	0.005	0.048	0.867	1	
Spain	0.722	0.457	0.331	0.785	0.993	0.969	0.311	0.032	1

µ

Table S4: One-way Anova (Tukey) for number of concomitant illnesses									
	Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain
Belgium (Dutch)	1								
Belgium (French)	<0.001	1							
Brazil	0.035	0.821	1						
Czech Republic	<0.001	0.236	0.022	1					
England	NA	NA	NA	NA	1				
Lithuania	0.861	<0.001	0.214	<0.001	NA	1			
Greece	0.900	<0.001	0.208	<0.001	NA	1.000	1		
Poland	0.386	<0.001	<0.001	<0.001	NA	0.002	0.003	1	
Spain	0.136	0.270	0.998	<0.001	NA	0.576	0.561	<0.001	1

Table S5: Chi-squared test for gender [n(%)]										
	Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain	p-value Chi²
n	26	29	12	48	10	58	50	30	15	
Gender										
Male	14 (53.8)	10 (34.5)	6 (50.0)	11 (22.9)	7 (70.0)	30 (51.7)	13 (26.0)	5 (33.3)	5 (33.3)	0.009
Female	12 (46.2)	19 (65.5)	6 (50.0)	37 (77.1)	3 (30.0)	28 (48.3)	37 (74.0)	19 (63.3)	10 (66.7)	

8. References

1. Cruz-Jentoft A, Bahat G, Bauer J, Boire Y, Bruyère O, Cederholm T, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing. 2018;0: 1–16.
2. Cruz-Jentoft AJ, Landi F, Schneider SM, Zuniga C, Arai H, Boirie Y, et al. Prevalence of and interventions for sarcopenia in ageing adults: A systematic review. Report of the International Sarcopenia Initiative (EWGSOP and IWGS). Age Ageing. 2014;43: 48–759.
3. Shafiee G, Keshtkar A, Soltani A, Ahadi Z, Larijani B, Heshmat R. Prevalence of sarcopenia in the world: a systematic review and meta- analysis of general population studies. J Diabetes Metab Disord. Journal of Diabetes & Metabolic Disorders; 2017;16: 21.
4. Beaudart C, Locquet M, Reginster JY, Delandsheere L, Petermans J, Bruyère O. Quality of life in sarcopenia measured with the SarQoL: impact of the use of different diagnosis definitions. Aging Clin Exp Res. Springer International Publishing; 2018;30: 307–313.
5. Ethgen O, Beaudart C, Buckinx F, Bruyère O, Reginster JY. The Future Prevalence of Sarcopenia in Europe: A Claim for Public Health Action. Calcif Tissue Int. 2017; 229–234.
6. Beaudart C, Zaaria M, Pasleau F, Reginster J-Y, Bruyère O. Health outcomes of sarcopenia: a systematic seview and meta-analysis. PLoS One. 2017;12: e0169548.
7. Beaudart C, Biver E, Bruyère O, Cooper C, Al-Daghri N, Reginster JY, et al. Quality of life assessment in musculo-skeletal health. Aging Clin Exp Res. Springer International Publishing; 2017;0: 1–6.
8. Beaudart C, Biver E, Reginster J-Y, Rizzoli R, Rolland Y, Bautmans I, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. Age Ageing. England; 2015;44: 960–966.
9. Beaudart C, Biver E, Reginster J-Y, Rizzoli R, Rolland Y, Bautmans I, et al. Validation of the SarQoL, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle. 2017;8: 238–244.
10. Beaudart C, Edwards M, Moss C, Reginster JY, Moon R, Parsons C, et al. English translation and validation of the SarQoL, a quality of life questionnaire specific for sarcopenia. Age Ageing. 2017;46: 271–277.
11. Ildiko GA, Mihai G, Beaudart C, Bruyère O, Pop R-M, Reginster J-Y, et al. Psychometric performance of the Romanian version of the SarQoL, a health-related quality of life questionnaire for sarcopenia. Arch Osteoporos. Archives of Osteoporosis; 2017;12: 103.

12. Geerinck A, Scheppers A, Beaudart C, Bruyère O, Vandenbussche W, Bautmans R, et al. Translation and validation of the Dutch SarQoL, a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact. 2018;18: 463–472.

13. Konstantynowicz J, Abramowicz P, Glinkowski W, Taranta E, Marcinowicz L, Dymitrowicz M, et al. Polish Validation of the SarQoL, a Quality of Life Questionnaire Specific to Sarcopenia. J Clin Med. 2018;7: 323.

14. Tsekoura M, Billis E, Gliatis J, Tsepis E, Matzaroglou C, Sakkas GK, et al. Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. Disabil Rehabil. Taylor & Francis; 2018;0: 1–7.

15. Reginster JY, Cooper C, Rizzoli R, Kanis JA, Appelboom G, Bautmans I, et al. Recommendations for the conduct of clinical trials for drugs to treat or prevent sarcopenia. Aging Clin Exp Res. 2016;28: 47–58.

16. Terwee C, Bot S, de Boer M, van der Windt D, Knol D, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60: 34–42.

17. Harvill LM, Tennessee E, Uniuersit S. Standard Error of Measurement. Educ Meas Issues Pract. 1991;10: 33–41.

18. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurements in Medicine. Cambridge University Press; 2011. https://doi.org/10.1186/1741-7015-9-29

19. Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. Lancet. 1986;327: 307–310.

20. Polit D. Getting serious about test-retest reliability: A critique of retest research and some recommendations. Qual Life Res. 2014;23: 1713–1720.

21. Bland J.M. & Altman DG. Measuring agreement in method comparison studies. Stat. Meth. 1999;2802: 135–160.

22. IBM Support. The Bland-Altman Plot [Internet]. [cited 15 Mar 2019]. Available: https://www-01.ibm.com/support/docview.wss?uid=swg21476730

23. Michalkova H, Ryznarova I, Seget J, Topinková E. Health-related quality of life in sarcopenia: translation and validation of the Czech version of the SarQoL instrument. European Geriatric Medicine. 2017. p. s104.

24. Lee RC, Wang Z, Heo M, Ross R, Jannsen I, Heymsfield SB. Total-body skeletal muscle mass: Development and cross-validation of anthropometric prediction models. Am J Clin Nutr. 2000;72: 796–803.

25. Hart DL. Test-Retest Reliability of an Abbreviated Self-Report Overall Health Status Measure. J Orthop Sport Phys Ther. 2003;33: 734–744.

26. Palmer S, Manns S, Cramp F, Lewis R, Clark EM. Test-retest reliability and smallest detectable change of the Bristol Impact of Hypermobility (BIoH) questionnaire. Musculoskelet Sci Pract. Elsevier Ltd; 2017;32: 64–69.

27. Oliveira IS, Costa LCM, Manzoni ACT, Cabral CMN. Assessment of the measurement properties of quality of life questionnaires in Brazilian women with breast cancer. Brazilian J Phys Ther. 2014;18: 372–383.

28. Juul T, Søgaard K, Davis AM, Roos EM. Psychometric properties of the Neck OutcOme Score, Neck Disability Index, and Short Form–36 were evaluated in patients with neck pain. J Clin Epidemiol. Elsevier Inc; 2016;79: 31–40.

29. Dawson J, Boller I, Doll H, Lavis G, Sharp R, Cooke P, et al. Minimally important change was estimated for the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. J Clin Epidemiol. Elsevier Inc; 2014;67: 697–705.

30. Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. Health Qual Life Outcomes. 2008;6: 1–12.

Chapter 3: Discriminative power of the SarQoL with the EWGSOP2 sarcopenia criteria

Discriminative Power of the Sarcopenia Quality of Life (SarQoL®) Questionnaire with the EWGSOP2 Criteria

Published as a letter to the editor:

Geerinck, A., Locquet, M., Reginster, JY. et al. Discriminative Power of the Sarcopenia Quality of Life (SarQoL®) Questionnaire with the EWGSOP2 Criteria. J Frailty Aging 10, 193–194 (2021). https://doi.org/10.14283/jfa.2020.47

Dear editor,

The Sarcopenia Quality of Life (SarQoL®) questionnaire was developed in 2015 to fill the need for a specific instrument to measure quality of life in sarcopenia. Since then, its validity and reliability have been evaluated in multiple languages, and it is now available in 30 language-specific versions. In multiple validation studies, the SarQoL® has demonstrated its ability to discriminate between sarcopenic and non-sarcopenic subjects when diagnosed according to the EWGSOP criteria [1]. However, these criteria have now been updated, and the discriminative power of the SarQoL® questionnaire should be reaffirmed using the EWGSOP2 criteria [2]. The analysis presented below aims to establish whether the SarQoL® questionnaire can discriminate between sarcopenic, probably sarcopenic (low grip strength in the EWGSOP2 algorithm) and non-sarcopenic participants.

This study used data gathered from older, community-dwelling volunteers recruited within the framework of the Sarcopenia and Physical Impairment with advancing Age (SarcoPhAge) cohort [3]. The same data was used in the original validation study of the SarQoL® questionnaire [4]. The sarcopenia components of muscle mass and muscle strength were measured with, respectively, dual-energy x-ray absorptiometry and a hydraulic hand dynamometer. We applied the thresholds specified by the EWGSOP2 for appendicular lean mass divided by height squared (ALM/Ht²: less than 5.5 kg/m² for women and 7 kg/m² for men) and handgrip strength (less than 16 kg for women and 27 kg for men) [2]. Quality of life was measured with the SarQoL® questionnaire, which provides an overall QoL score and 7 domain scores for specific aspects of QoL, all between zero (worst QoL) and 100 (best QoL). In line with the case-finding algorithm elaborated by the EWGSOP2, we considered participants to have "probable sarcopenia" when they demonstrated low grip strength, and sarcopenia when both low grip strength and low muscle mass were present [2].

In total, 296 participants, with a median age of 73.3 (68.9-78.6) years, were included in this analysis. In a previous analysis, 43 subjects were diagnosed as sarcopenic with the EWGSOP criteria [4]. As

expected, we found a lower prevalence of sarcopenia when applying the EWGSOP2 criteria, with 38 participants displaying low grip strength, of which 13 were ultimately considered sarcopenic.

Sarcopenic participants, as diagnosed with EWGSOP2 criteria, had significantly lower scores for all 7 SarQoL® QoL domains (all $p<0.05$) and the overall QoL score of the SarQoL® questionnaire [45.83 (38.62-60.26) versus 66.43 (56.10-78.26); $p<0.001$], indicating that the SarQoL® questionnaire can discriminate between sarcopenic and non-sarcopenic individuals. When the sample was categorized in probably sarcopenic (n=38) and non-sarcopenic (n= 258), similar results were obtained. All 7 domain scores of the SarQoL® questionnaire were significantly lower (all $p<0.05$) for probably sarcopenic participants, as well as the Overall QoL score [53.24 (41.18-63.24) versus 67.74 (57.35-79.02); $p<0.001$]. Detailed results are presented in table 1.

We investigated the robustness of these results by carrying out binary logistic regression analyses including age, gender, body mass index, n° of comorbidities and n° of medications as covariates. We found that for every one-unit increase in Overall QoL, we expect to see a 10% decrease in the odds of belonging to the EWGSOP2 sarcopenic group (OR: 0.90; 95% CI: 0.85-0.95), and a 6% decrease in the odds of belonging to the EWGSOP2 probable sarcopenia group (OR: 0.94; 95% CI: 0.90-0.97). The current analysis shows that the SarQoL® questionnaire retains its capacity to discriminate between sarcopenic and non-sarcopenic persons when using the EWGSOP2 criteria for sarcopenia, despite the reduced prevalence of sarcopenic individuals. These results reinforce the results found during the validation of the Lithuanian version of the SarQoL® questionnaire, which also found significantly lower QoL scores for all 7 domains and the overall QoL score between non-sarcopenic and EWGSOP2 sarcopenic participants. The odds ratio found for the overall QoL score in this study is nearly identical to our own, at 0.913 (95% CI 0.876-0.951) (5).

We also found that participants with low grip strength, categorized as probably sarcopenic in the EWGSOP2 algorithm, had significantly lower QoL scores for all 7 domains and the overall QoL score. This is an important finding because it shows that, when an older person is found to have low muscle strength, his or her quality of life is likely to already have been impacted. This adds strength to the EWGSOP argument that the observation of low grip strength in clinical practice could be a sufficient indication to put in place interventions to mitigate and improve a patient's musculoskeletal health.

The SarQoL® questionnaire is currently the only sarcopenia-specific QoL questionnaire, and has demonstrated to be able to discriminate between sarcopenic, probably sarcopenic and non-sarcopenic groups. Its use, in combination with the EWGSOP2 criteria, could provide greater detail and precision on the impact of sarcopenia on QoL.

Table 1: Discriminative power of the SarQoL® questionnaire using the EWGSOP2 criteria for sarcopenia

	EWGSOP2 sarcopenia				EWGSOP2 probable sarcopenia			
	Not sarcopenic (n=283)	Sarcopenic (n=13)	p [a]	OR (95% CI) [b]	Not sarcopenic (n= 258)	Probably sarcopenic (n=38)	p [a]	OR (95% CI) [b]
1. Physical and mental health	63.33 (54.43 - 76.67)	55.57 (38.33 - 60.55)	0.006	0.94 (0.89-0.98)	63.33 (55.57 - 76.67)	55.57 (45.57 - 63.33)	<0.001	0.96 (0.93-0.99)
2. Locomotion	61.11 (50.00 - 83.33)	30.56 (25.00 - 62.50)	0.004	0.96 (0.93-0.99)	61.11 (50.00 - 86.11)	50.00 (27.78 - 61.81)	<0.001	0.97 (0.95-0.99)
3. Body composition	60.00 (50.00 – 70.00)	50.00 (29.58 - 54.17)	0.001	0.92 (0.88-0.97)	60.00 (50.00 - 70.83)	50.00 (40.00 – 60.00)	0.003	0.98 (0.95-1.00)
4. Functionality	73.21 (60.71 - 84.62)	53.85 (42.58 - 75.21)	0.002	0.95 (0.91-0.98)	75.00 (62.50 - 85.71)	57.69 (45.67 - 72.32)	<0.001	0.96 (0.93-0.99)
5. Activities of daily living	63.33 (51.67 – 80.00)	40.00 (30.51 – 50.00)	0.001	0.92 (0.86-0.96)	65.00 (52.98 – 80.00)	48.03 (34.96 - 58.75)	<0.001	0.95 (0.93-0.98)
6. Leisure activities	58.31 (33.25 - 66.75)	33.25 (33.25 – 50.00)	0.008	0.95 (0.91-0.99)	66.69 (33.25 - 66.75)	33.25 (33.25 – 50.00)	<0.001	0.97 (0.94-0.99)
7. Fears	87.50 (87.50 - 100.00)	75.00 (75.00 – 100.0)	0.045	0.94 (0.89-0.99)	87.50 (87.50 – 100.00)	87.50 (75.00 – 100.00)	0.044	0.98 (0.95-1.01)
Overall score	66.43 (56.10 - 78.26)	45.83 (38.62 - 60.26)	<0.001	0.90 (0.85-0.95)	67.74 (57.35 - 79.02)	53.24 (41.18 - 63.24)	<0.001	0.94 (0.90-0.97)

[a] p-value calculated with Mann-Whitney U-test
[b] Binary logistic regression adjusted for age, gender, BMI, n° of comorbidities and n° of medications

References

1. SarQoL publications. www.sarqol.org/en/publications.
2. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019;48:16–31
3. Beaudart C, Locquet M, Reginster JY, Delandsheere L, Petermans J, Bruyère O. Quality of life in sarcopenia measured with the SarQoL®: impact of the use of different diagnosis definitions. Aging Clin Exp Res 2018;30:307–313
4. Beaudart C, Biver E, Reginster J-Y, et al. Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 2017;8
5. Alekna V, Kilaite J, Tamulaitiene M, et al. Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). Eur Geriatr Med 2019

Chapter 4: SarQoL and physical frailty

Evaluating quality of life in frailty: applicability and clinimetric properties of the SarQoL® questionnaire

Published as:

Geerinck A, Locquet M, Bruyère O, Reginster JY, Beaudart C. Evaluating quality of life in frailty: applicability and clinimetric properties of the SarQoL® questionnaire. J Cachexia Sarcopenia Muscle. 2021 Apr;12(2):319-330. doi: 10.1002/jcsm.12687

1. Abstract

Background: The SarQoL® questionnaire was specifically designed to measure quality of life (QoL) in sarcopenia. Frailty and sarcopenia have areas of overlap, notably weak muscle strength and slow gait speed, which may mean that the SarQoL could provide a measure of QoL in frailty. This study aimed to evaluate the clinimetric properties of the SarQoL questionnaire in physical frailty using the Fried criteria.

Methods: Analyses were carried out on data from the Sarcopenia and Physical impairment with advancing Age study. Frailty was assessed with the Fried criteria and QoL with the SarQoL, the Short-Form 36-Item, and the EuroQoL 5-Dimension (EQ-5D) questionnaires. We evaluated discriminative power (with the Kruskal-Wallis analysis of variance test), internal consistency (with Cronbach's alpha), construct validity (through hypotheses testing), test-retest reliability (with the intraclass correlation coefficient), measurement error (calculating standard error of measurement and smallest detectable change), and responsiveness (through hypotheses testing and standardized response mean).

Results: In total, 382 participants were included for the validation and 117 for the responsiveness evaluation. They had a median age of 73 (69-79) years, took 5 (3-8) drugs, and had 4 (3-5) co-morbidities. There were more women (n = 223; 58.4%) than men and, in total, 172 (45%) robust, 167 (44%) pre-frail, and 43 (11%) frail participants. Discriminative power was confirmed when significantly lower ($P < 0.001$) overall SarQoL scores, and thus also worse QoL, were observed between robust [77.1 (64.35-85.90)], pre-frail [62.54 (53.33-69.57)], and frail [49.99 (40.45-56.06)] participants. Six of the SarQoL domains performed likewise, with significantly lower scores according to frailty status with Domain 7 (fears) being the exception. Internal consistency was good ($\alpha = 0.866$). Convergent (using Short-Form 36-Item and EQ-5D) and divergent construct validity (using EQ-5D) was confirmed. Test-retest reliability was excellent [intraclass correlation coefficient = 0.918 (0.834-0.961)], with a standard error of measurement of 3.88 and a smallest detectable change of 10.76 points.

We found moderate responsiveness when five of the nine hypotheses were confirmed, coupled with a large effect size for the overall SarQoL score (corrected standardized response mean of -1.44).

Conclusions: The SarQoL questionnaire has adequate clinimetric properties for use with frail patients in clinical practice and trials and could provide data that are more appropriate and detailed than the generic questionnaires currently used.

2. Introduction

The World Health Organization declared the period from 2020 to 2030 to be the decade of healthy ageing, which they define as 'the process of developing and maintaining the functional ability that enables wellbeing in older age'[1]. This concept is closely linked to the syndrome of frailty, a clinically recognizable state of increased vulnerability in older people, caused by age-related losses in physiological reserves and function across multiple organ systems, such that the ability to cope with everyday or acute stressors is compromised [2].

This state of increased vulnerability is associated with negative health outcomes, as evidenced by a recent meta-analysis, which found an increased likelihood of premature mortality, hospitalization, and institutionalization [3]. Frailty was also associated with an increased risk for developing disability in both basic and instrumental activities of daily living, an increased risk for physical limitations, dependency, falling, fractures, cognitive decline, decline in lean body mass, and lower life satisfaction.3 These outcomes, in combination with an estimated prevalence of 10.7–18%, mean that frailty represents an important burden on public health [4-6].

While hard outcomes such as mortality and hospitalizations remain the primary indicators in research settings, outcomes measuring the subjective experience of patients are becoming as essential part of the arsenal. Health-related quality of life is one of the main patient-reported outcome measures used in research, and several studies have already focused on quality of life (QoL) in frailty in the last decade. A 2019 systematic review listed 22 studies that assessed QoL in frailty and which demonstrated that frail participants had worse QoL than robust participants. However, these differences between frail and robust people were only clear for the sub-concepts of physical functioning and satisfaction with life. For social and environment scales, results were inconsistent between the different questionnaires used, limiting their usefulness in assessing the psychosocial well-being pre-frail and frail individuals. In this systematic review, the Short-Form 36-Item (SF-36) was the most frequently used instrument out of the 14 instruments included, followed by the WHOQOL-BREF, the CASP-19, and the EUROHIS-QOL [7]. Several observations can be made from the results of this systematic review. First, the SF-36, which was the most frequently used instrument to measure QoL in frailty, is a generic instrument and not

adapted to specific populations or diseases [8]. While generic instruments allow QoL to be compared between a range of populations, specific instruments often possess better construct validity and are more sensitive to changes in QoL over time [9]. Secondly, the concept of QoL and the components needed to provide a holistic assessment were interpreted differently between each of the QoL questionnaires. While some concepts from the generic QoL questionnaires mentioned previously are shared with the sarcopenia quality of life (SarQoL®) questionnaire (i.e. physical and mental health and activities of daily living), others such as 'body composition', 'leisure activities', and 'fears' are unique.

The systematic review did not include frailty-specific QoL instruments. A QoL instrument specific to the frailty syndrome might improve sensitivity to change in disease-specific QoL over time in this group [10].

One such specific questionnaire is the SarQoL questionnaire, developed in 2015 with the aim of measuring health-related QoL in sarcopenic persons [11]. The questionnaire was constructed using input from experts, literature review, and crucially, interviews with older, sarcopenic individuals. It has been validated for use with sarcopenic, older, community-dwelling participants in multiple languages and has consistently been shown to be a valid and reliable instrument, as well as responsive to changes in QoL [12-21].

Multiple authors have argued that the conceptual frameworks of frailty and age-related sarcopenia overlap substantially, notably on the similar clinical manifestations used to diagnose the two conditions. The slowness indicator in the Fried criteria for frailty and the low gait speed indicator used to characterize muscle function in sarcopenia are one area of overlap between the two conditions. Partial overlap exists between weight loss in frailty and muscle loss in sarcopenia, and fatigue/exhaustion in frailty and grip strength in sarcopenia. Some have argued that sarcopenia is equivalent to the physical component of frailty, separate from the cognitive, psychological, sociological, and spiritual components of frailty [22-24].

Because of the overlap between sarcopenia and physical frailty, we considered it worthwhile to explore whether the SarQoL questionnaire could be used in the assessment of QoL in frail and pre-frail individuals, as diagnosed with the Fried criteria. This study aims to examine the clinimetric properties of the SarQoL questionnaire in robust, pre-frail, and frail participants of the Sarcopenia and Physical impairment with advancing Age (SarcoPhAge) study.

3. Methods

3.1. Population

The analyses described in this manuscript have been carried out using the data collected during the SarcoPhage study. This cohort study followed a sample of community-dwelling older people for 5 years and has been described in multiple publications [25-29]. In brief, the SarcoPhAge study recruited a convenience sample of volunteers aged 65 years or older living in the Liège province of Belgium. Participants were recruited from different departments of an outpatient clinic in Liège, as well as through advertisement in the local press. Candidates were not eligible for inclusion in the cohort if they presented with a body mass index (BMI) >50 kg/m2 or if they had one or more amputated limbs. No other exclusion criteria were applied. Participants were invited to the research centre once yearly, where they performed physical tests and completed questionnaires [25]. For the analyses presented here, we used data from Year 1 of follow-up, except for the evaluation of the responsiveness of the questionnaire, where we used data from the visits carried out at 1 and 5 years into the study.

3.2. Frailty evaluation

In the SarcoPhAge sample, physical frailty was evaluated with the criteria described by Fried et al [30]. The Fried diagnostic criteria evaluate five items to determine whether a person is considered to be robust, pre-frail, or frail. In this study, the five criteria were measured with the following instruments: weakness was present if handgrip strength measured with hydraulic dynamometer was below the cut-offs based on gender and BMI, low gait speed was detected by evaluating usual walking speed on a 4 m track (results corrected to 4.5 m track) with cut-offs based on gender and height, low physical activity was measured with the Minnesota Leisure Time Activity Questionnaire [31] using gender-specific cut-offs for kilocalories used in physical activity in the preceding week, exhaustion was established using two items from the Center for Epidemiological Studies Depression scale [32] and weight loss was detected through a self-reported question on unintentional weight loss of more than 4.5 kg in the past year [25, 30]. For each item, participants were given 1 point if below the cut-off, and 0 if not, and these item scores were summed for a frailty score between 0 and 5. Participants with zero points were considered robust, a score of 1 or 2 points indicated a pre-frail state, and subjects with a score of 3 or more points were considered to be frail. A detailed description of the criteria, instruments, and cut-off values is provided in Table S1 in annex.

3.3. Quality of life measurement

The SarQoL questionnaire, the focus of this validation study, is a patient-reported outcome measure specifically designed to evaluate QoL in older, sarcopenic, community-dwelling people. There are 55 items in the questionnaire, categorized into seven domains of health-related dysfunction: (i) physical and mental health, (ii) locomotion, (iii) body composition, (iv) functionality, (v) activities of daily living, (vi) leisure activities, and (vii) fears. A score between 0 (worst QoL) and 100 (best QoL) is provided for each domain, and an overall QoL score (range: 0–100 points) is calculated on the entirety of the questionnaire [11]. The scoring algorithm is not publicly available, but tools to calculate the scores are available upon request via info@sarqol.org or via the website www.sarqol.org and free for non-sponsored research. The questionnaire is self-reported and takes about 15 min to complete. The SarQoL questionnaire has been validated in multiple languages and has been shown to be a valid and reliable instrument [13, 15-19, 33] The questionnaire was shown to be responsive to changes in QoL in a sample of 42 sarcopenic subjects followed over 3 years, and its standard error of measurement (SEM) and smallest detectable change have been calculated in different European populations as well as pooled [20, 21].

Complementary to the SarQoL questionnaire, two generic QoL questionnaires were also completed by each participant to allow the evaluation of the construct validity of the SarQoL questionnaire. The first of these, the SF-36 questionnaire, measures functional health and well-being from the patient's perspective, providing eight domain scores (physical functioning, social functioning, role functioning physical, role functioning emotional, vitality, bodily pain, mental health, and general health) and two summary scores (physical and mental), all scored from 0 (worst QoL) to 100 (best QoL) points.34 Secondly, the EuroQoL 5-Dimension 3-Level (EQ-5D-3L) and the associated visual analogue scale (EQ-VAS) were administered. The EQ-5D is a generic measure of health status, which records self-reported problems (none, some, and extreme) on five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) [35]. Results are reported as an index score (between 0 and 1, with 0 indicating death and 1 indicating perfect health) and a self-rated health evaluation (between 0, worst imaginable health, and 100, best imaginable health) [36].

3.4. Clinimetric properties

The measurement properties to be included in this validation were selected based on the COSMIN taxonomy and its related documentation [37, 38]. These include known-groups validity (also known as discriminative power), internal consistency, construct validity (through hypotheses testing), reliability (test–retest), measurement error, and responsiveness. We also looked at the presence of floor and/or ceiling effects and provided the smallest detectable change to aid in the interpretation of the evolution of the SarQoL scores over time.

i. Known-groups validity is based on the hypothesis that two or more groups with distinctive characteristics should logically differ in the construct that is measured [39]. In the context of this study, the hypothesis is that robust participants should have higher QoL scores than pre-frail and frail participants, which would mean that the SarQoL questionnaire can discriminate between the three frailty profiles.

ii. Internal consistency quantifies the degree of interrelatedness between the items in the questionnaire, that is, whether all items in the SarQoL measure the same underlying construct (QoL) [37, 38]

iii. Construct validity is used to assess whether the questionnaire under investigation actually measures what it theoretically aims to measure. This is performed by comparing the questionnaire with other questionnaires (or subscales of) that should, in theory, measure the same construct (convergent validity) or a different construct (divergent validity) [37, 38]. In this study, we utilized the same eight hypotheses on the strength of association between the overall QoL score of the SarQoL questionnaire and domains of the SF-36 and EQ-5D that were used in previous validations [13-19, 33].

iv. The test–retest reliability of the questionnaire shows whether the scores measured by the SarQoL questionnaire remain stable between multiple administrations, on the condition that the participants' health state also remains stable [37, 38]. To measure this, the SarQoL questionnaire was administered twice, with an approximate interval of 2 weeks in between, and participants provided information on the stability of their health. Because of the different objectives of the study that collected the data analysed in this article, only the 43 participants who were diagnosed as sarcopenic with the European Working Group on Sarcopenia in Older People criteria were invited, at the time of the original validation study, to participate in the retest part, with 30 providing usable data [13]. Reliability is also demonstrated by the SEM, which provides a measure of the dispersion of observed scores around the ‘true’ score from repeated measurements. The smallest detectable change provides the value for the minimum change in QoL scores that needs to be observed to be certain that the measured change in QoL is real and not possibly due to measurement error [38].

v. Floor and ceiling effects indicate that the range of the scale is too narrow and that extreme profiles cannot be accurately measured. They are present when >15% of the participants obtain either the highest or lowest score.

vi. The last clinimetric property investigated was the responsiveness of the questionnaire, that is, its capacity to detect change over time, between the first and fifth years of the SarcoPhAge study [21] We used the same methodology as in a previous evaluation of the responsiveness of the SarQoL, which was combination of hypothesis testing and effect size

evaluation [21]. In short, we evaluated nine hypotheses (see Table 5) on the theorized strength of correlation between the changes observed with the SarQoL questionnaire between Year 1 and Year 5 and the changes observed with (domains of) the SF-36, EQ-5D, and EQ-VAS. The results were interpreted with the criteria from de Boer et al., which indicate that a questionnaire has high responsiveness if at least 75% of hypotheses are confirmed, moderate when at least 50–75% are confirmed, and poor responsiveness when less than 50% are confirmed [40]. In this analysis, we included all participants for whom we had valid data at Year 1 and Year 5. For the second method, we calculated standardized response means (SRMs) (a measure of effect size), which reflect the magnitude of change measured by the SarQoL and by the other questionnaires used in this study. Larger effect sizes indicate that the questionnaire possesses better responsiveness [41]. Because this method is based on the assumption that a change in health status has occurred, we could only include those participants for whom we had valid data at Year 1 and Year 5 and whose frailty status changed in the years between evaluations. The change in frailty status is used here as a proxy measure of change in health status, and we hypothesize that a change in frailty status will be reflected in the observed change in QoL.

3.5. Statistical analysis

Normality of distribution for quantitative variables was tested with the Shapiro–Wilk test, by comparing mean and median and by evaluating the histogram and Q–Q plot. Continuous variables following a Gaussian distribution are reported as mean ± standard deviation, while those who do not are reported as median (25th–75th percentile). Nominal variables are reported as absolute (n) and relative (%) frequencies. The evaluation of differences between groups for nominal variables was carried out using Pearson's χ^2 test. All results were considered significant at 5% level ($P \leq 0.05$), except for pairwise comparisons between the robust, pre-frail, and frail groups, which were considered significant at $P \leq 0.017$ (P-value adjusted for the number of comparisons: $\alpha = 0.05/3$). IBM SPSS Statistics Version 25 for Windows (IBM Corp., Armonk, NY) was used for all statistical manipulations.

i. Analysis of continuous variables to determine the known-groups validity between the three frailty categories was carried out with the analysis of variance test if distributions in all groups were Gaussian and with the Kruskal–Wallis test if they were not. Paired comparisons were carried out with multinomial regression analysis, so as to obtain P-values for the differences between the robust, pre-frail, and frail groups.
ii. Internal consistency was determined with Cronbach's alpha test, where a value between 0.70 and 0.95 indicates good internal consistency [42].

iii. Associations between two continuous variables (such as used in the hypotheses for evaluating construct validity and responsiveness) were examined with Pearson's or Spearman's correlations, depending on normality of distribution.

iv. The test–retest reliability was quantified by calculating the intraclass correlation coefficients (ICCs) (two-way mixed model − absolute agreement type) between the scores from the first and the second administrations. ICCs greater than 0.7 indicate acceptable reliability [38]. The SEM was calculated by dividing the standard deviation of the difference between the scores from the first administration and those of the second administration by the square root of 2. This gives the following formula: SEM = (SD(test score − retest score) / √2). The smallest detectable change is derived from the SEM value, by the following formula: 1.96 * √2 * SEM.

v. Floor and ceiling effects were evaluated following inspection of the frequency tables.

vi. Finally, SRMs, a measure of effect size and used to evaluate responsiveness, were calculated by dividing the mean difference between the SarQoL scores from the first year and the fifth year of the SarcoPhAge study by the standard deviation of the differences between these paired values. The SRM values were subsequently transformed with the formula SRM/ √2√(1 − r), where 'r' signifies the correlation between Year 1 and Year 5 scores [43]. The corrected SRM values can now be interpreted with the thresholds formulated by Cohen et al., where SRM < 0.20 is trivial effect, 0.20 ≤ SRM < 0.50 is a small effect, 0.50 ≤ SRM < 0.80 is a moderate effect, and SRM ≥ 0.80 is considered a large effect [44].

4. Results

4.1. Clinical characteristics

In total, 382 subjects were eligible for inclusion at the first follow-up visit of the SarcoPhAge study. These subjects had a median age of 73 (69–78) years old, were slightly overweight at a median BMI of 27 (24–30) kg/m^2, took a median of 5 (3–8) drugs, and had a median of 4 (3–5) co-morbidities. There were slightly more women (n = 223; 58.4%) than men in the sample. The median grip strength was 39 (33–45) kg for men and 21 (17.5–25) kg for women. Lastly, the median gait speed in the complete sample was 1.09 (0.91–1.28) m/s.

All 382 participants were evaluated for frailty with the Fried criteria, and we found 172 (45%) robust, 167 (44%) pre-frail, and 43 (11%) frail individuals. Clinical characteristics were significantly different between these three groups. Frail participants were older than pre-frail participants, who, in turn, were older than robust individuals (P < 0.001). The same dynamic was present for BMI (with frail participants

having the highest BMI; $P < 0.001$), drug consumption (with frail participants taking the most drugs; $P < 0.001$), and co-morbidities (with frail participants having the highest number of co-morbidities; $P < 0.001$). As expected, frail participants had lower grip strength than pre-frail participants, who, in turn, had lower grip strength than robust people ($P < 0.001$ for men and women). The same was observed for gait speed (with frail participants having the lowest gait speed; $P < 0.001$). Detailed results and pairwise comparisons are available in Table 1.

4.2. Known-groups validity

The overall QoL score measured by the SarQoL questionnaire was significantly different ($P < 0.001$) between the three categories of frailty, following a downward trend with robust participants having the best QoL [77.10 (64.35–85.90)], followed by the pre-frail participants [62.54 (53.33–69.57)] and with the frail participants presenting with the worst QoL [49.99 (40.45–56.06)]. The differences between the overall QoL scores in these three groups were revealed to be significant in the paired comparisons (all $P < 0.001$).

The QoL scores for the seven domains of health-related quality of life in the SarQoL questionnaire were also shown to be highly significantly different ($P < 0.001$). An examination of the paired differences showed that only Domain 7 (fears) was not significantly different in the comparison between pre-frail and frail groups ($P = 0.119$).

The complete results of the known-groups validity are presented in Table 2.

4.3. Internal consistency

The homogeneity of the questionnaire was found to be excellent, with Cronbach's alpha of 0.866, at the upper end of the 0.70 to 0.95 range considered good. This shows that the questionnaire is consistent without showing the increased likelihood for redundancy associated with alpha values greater than 0.95. The influence of individual domains was tested by deleting a single domain at a time. The resulting alpha values ranged from 0.854 to 0.894, indicating that no domain unduly influences the internal consistency.

Table 1: Clinical characteristics

	All (n=382)	Robust (n=172)	Pre-frail (n=167)	Frail (n=43)	P R-PF[a]	P R-F[b]	P PF-F[c]	P Overall[d]
Age (years)	73.18 (68.77 – 78.47)	70.76 (68.15 - 76.05)	73.79 (69.45 - 79.59)	76.52 (71.87 - 81.82)	<0.001	<0.001	0.031	<0.001
BMI (kg/m²)	26.86 (23.74 – 30.12)	26.24 (23.4 - 28.68)	27 (23.82 - 30.43)	28.74 (24.67 – 34.00)	0.077	0.001	0.037	0.004
Drug consumption (n)	5 (3-8)	5 (3 - 6)	6 (4 - 8)	7 (5 - 10)	<0.001	<0.001	0.079	<0.001
Comorbidities	4 (3-5)	3 (2 - 5)	4 (3 - 6)	4 (3 - 7)	<0.001	<0.001	0.249	<0.001
Female [n(%)]	223 (58.4%)	97 (56.4%)	98 (58.7%)	28 (65.1%)				0.580
Grip strength (kg)								
Men	39 (33 – 45)	40 (37 - 45)	38 (31 - 45)	26 (18 - 32)	0.012	<0.001	<0.001	<0.001
Women	21 (17.5 – 25)	24 (21 - 27.5)	19 (16 - 22)	13 (11.63 - 17.75)	<0.001	<0.001	<0.001	<0.001
Gait speed (m/sec)	1.09 (0.91 – 1.28)	1.2 (1.05 - 1.35)	1.06 (0.86 - 1.23)	0.72 (0.53 - 0.87)	<0.001	<0.001	<0.001	<0.001

[a] p-value for pairwise comparison between robust and pre-frail group (significant at p≤0.017)
[b] p-value for pairwise comparison between robust and frail group (significant at p≤0.017)
[c] p-value for pairwise comparison between pre-frail and frail group (significant at p≤0.017)
[d] p-values obtained through Kruskal-Wallis Anova test with pairwise comparisons, except for gender (Chi-squared test)
All results reported as median (25th percentile – 75the percentile) except for gender, reported as absolute and relative frequency.

Table 2: Discriminative power of the SarQoL® questionnaire in frailty

	Robust (n=172)	Pre-frail (n=167)	Frail (n=43)	P R-PF[a]	P R-F[b]	P PF-F[c]	P Overall[d]
Domain 1: Physical and mental health	72.2 (59.5-86.63)	59.97 (52.2-68.87)	45.53 (38.87-55.53)	<0.001	<0.001	<0.001	<0.001
Domain 2: Locomotion	77.78 (58.33-91.67)	55.56 (47.22-69.44)	41.67 (30.56-52.78)	<0.001	<0.001	0.001	<0.001
Domain 3: Body composition	70.83 (55.21-80)	58.33 (50-66.67)	50 (41.67-58.33)	<0.001	<0.001	0.014	<0.001
Domain 4: Functionality	82.69 (73.11-91.07)	69.23 (59.62-78.85)	55.36 (48.08-62.50)	<0.001	<0.001	<0.001	<0.001
Domain 5: Activities of daily living	76.67 (64.47-85.32)	60.71 (48.33-72.50)	50 (33.33-58.33)	<0.001	<0.001	<0.001	<0.001
Domain 6: Leisure activities	66.50 (49.88-66.50)	33.25 (33.25-49.88)	33.25 (16.62-33.25)	<0.001	<0.001	0.011	<0.001
Domain 7: Fears	100.00 (87.50-100)	87.50 (87.50-100.00)	87.5 (75.00-87.50)	0.032	0.004	0.119	<0.001
Overall QoL score	77.10 (64.35-85.90)	62.54 (53.33-69.57)	49.99 (40.45-56.06)	<0.001	<0.001	<0.001	<0.001

[a] p-value for comparison between robust and pre-frail group
[b] p-value for comparison between robust and frail group
[c] p-value for comparison between pre-frail and frail group
[a, b, c] p-values obtained through multinomial regression analysis adjusted on age, BMI, n° of drugs, n° of comorbidities and gender
[d] P-value obtained through Kruskal-Wallis Anova test
All results reported as median (25th percentile – 75the percentile) except for gender, reported as absolute and relative frequency.

4.4. Construct validity

Two sets of hypotheses were examined: for the convergent construct validity, we theorized that the overall QoL score of the SarQoL questionnaire measures a construct related to the SF-36 physical functioning, role limitation due to physical problems, and vitality domains as well as to the EQ-5D utility score. We therefore expect to find moderate to strong correlations between the SarQoL and these domains. For the divergent construct validity, we theorized that the overall QoL score of the SarQoL questionnaire measures a different construct than the SF-36 role limitation due to emotional problems and mental health domains, as well as the self-care and anxiety/depression items of the EQ-5D. If this is correct, we expect to find weak or non-existent correlations between the SarQoL and these domains.

The convergent validity of the SarQoL questionnaire in the entire sample was excellent, as evidenced by the confirmation of the four pre-specified hypotheses and the strong correlations between the overall QoL score of the SarQoL questionnaire and the four domains theorized to measure similar constructs (correlation coefficients between 0.447 and 0.798). When isolating the three frailty groups, the results were largely similar, with the exception of the correlation between the SarQoL and the SF-36 role limitation due to physical problems domain in the frail group, which dropped from $r = 0.628$ ($P < 0.001$) to $r = 0.246$ ($P = 0.199$).

The results of the divergent construct validity were less straightforward: both in the complete sample and in the three frailty categories, we found moderate to strong correlations between the SarQoL questionnaire and the two domains of the SF-36 theorized to be measuring a different construct. The hypotheses with the EQ-5D self-care and anxiety/depression items were confirmed by weak correlations (respectively, $r = -0.273$; $P < 0.001$ and $r = -0.257$; $P < 0.001$).

The full results, shown in Table 3, demonstrate that six out of the eight pre-specified hypotheses were confirmed, fulfilling the criteria of 75%, which indicates acceptable construct validity.

4.5. Reliability

One of the 30 participants was not evaluated for physical frailty, which means that test–retest data were available for 29 participants, of which 4 (13.8%) were robust, 18 (62.1%) were pre-frail, and 7 (24.1%) were frail. An ICC of 0.918 [95% confidence interval (CI) = 0.834–0.961] was found for the overall QoL score of the SarQoL questionnaire, demonstrating excellent test–retest reliability. The ICCs for the individual domains showed acceptable (ICC > 0.7) reliability for all but two domains: Domain 6, leisure activities [ICC = 0.391 (95% CI = 0.029–0.660)], and Domain 7, fears [ICC = 0.318 (95% CI = −0.055 to 0.612)]. Detailed results for the test–retest reliability are reported in Table 4.

Table 3: Construct validity of the SarQoL® questionnaire in frailty								
	Robust (n=172)		Pre-frail (n=167)		Frail (n=43)		All (n=382)	
	r	p	*r*	p	*r*	p	*r*	p
Convergent validity								
SF-36 Physical functioning	0.761	<0.001	0.693	<0.001	0.608	<0.001	0.798	<0.001
SF-36 Role limitation physical	0.408	<0.001	0.611	<0.001	0.246	0.199	0.628	<0.001
SF-36 Vitality	0.595	<0.001	0.499	<0.001	0.695	<0.001	0.678	<0.001
EQ-5D Utility score	0.305	<0.001	0.311	<0.001	0.564	0.001	0.447	<0.001
Divergent validity								
SF-36 Role limitation emotional	0.400	<0.001	0.473	<0.001	0.015	0.936	0.503	<0.001
SF-36 Mental health	0.588	<0.001	0.489	<0.001	0.392	0.035	0.554	<0.001
EQ-5D Self-care	/ [a]	/ [a]	-0.207	0.011	-0.278	0.145	-0.273	<0.001
EQ-5D Anxiety/depression	-0.161	0.070	-0.222	0.010	-0.489	0.007	-0.257	<0.001
[a]: all 172 robust subjects responded identically on the EQ-5D self-care question								

Table 4: Test-retest reliability of the SarQoL® questionnaire in frailty				
	ICC	95% CI	SEM	SDC
Domain 1: Physical and mental health	0.764	0.558 - 0.881	7.06	19.57
Domain 2: Locomotion	0.850	0.706 - 0.926	7.92	21.94
Domain 3: Body composition	0.700	0.454 - 0.847	8.81	24.41
Domain 4: Functionality	0.879	0.759 - 0.941	5.50	15.25
Domain 5: Activities of daily living	0.812	0.638 - 0.907	6.78	18.80
Domain 6: Leisure activities	0.391	0.029 - 0.660	13.82	38.30
Domain 7: Fears	0.318	-0.055 - 0.612	14.32	39.68
Overall QoL score	0.918	0.834 - 0.961	3.88	10.76

The SEM in this sample was calculated to be 3.88 points, leading to a smallest detectable change of 10.76 points. In practical terms, the overall QoL score of an individual participant would have to change by 10.76 points to be able to be sure that the observed change in QoL is due to a real change in QoL in the patient. SEM and smallest detectable change for the individual domains are reported in Table 4.

4.6. Floor and ceiling effects

None of the 382 participants obtained the lowest (0) or the highest (100) score possible for the overall QoL score of the SarQoL® questionnaire, showing the absence of floor and ceiling effects in the summary score.

4.7. Responsiveness

Out of the 382 participants who provided usable data at the first year of the SarcoPhAge study, 235 remained in the study at the fifth year of follow-up and were included in the responsiveness evaluation. Of these 235, a further 117 changed in terms of their frailty status between the first and fifth years of the study, and these were included in the analysis of responsiveness through the evaluation of effect size (SRMs).

We examined nine hypotheses used in an earlier study of the responsiveness of the SarQoL questionnaire, on the theorized correlation between changes measured by the SarQoL questionnaire and by other questionnaires. We were able to confirm five out of nine hypotheses but had to reject Hypothesis 1 (ΔSarQoL overall score and ΔSF-36 general health), Hypothesis 4 (ΔSarQoL overall score and ΔEQ-VAS), Hypothesis 5 (ΔSarQoL Domain 1 and ΔSF-36 general health), and Hypothesis 6 (ΔSarQoL Domain 1 and ΔEQ-VAS). According to the criteria formulated by De Boer et al., this indicated that the SarQoL questionnaire possesses moderate responsiveness because 45% of the hypotheses have been refuted. The details of the hypotheses and the observed correlations can be found in Table 5.

We also evaluated responsiveness with the metric of effect size. We calculated SRMs for all domains and summary scores of the SarQoL, SF-36, and EQ-5D questionnaires. The complete results are reported in Table 6. We can observe that the SRM of the SarQoL overall score (corrected SRM = −1.14) is much larger than the SF-36 PCS (corrected SRM = −0.634), the EQ-5D index (corrected SRM = 0.064), and the EQ-VAS (corrected SRM = −0.267). Globally, the SarQoL questionnaire had small effect sizes for three domain scores, moderate for 2 and large for 1. The SF-36 obtained small effect sizes for five domains and the MCS, and moderate effect sizes for three domains and the PCS.

Table 5: Evaluation of responsiveness with hypotheses

Hypothesis	Expected strength of correlation	Observed correlation		Confirmation/ rejection
		r	p-value	
1. Δ SarQoL® Overall score and Δ SF-36 General Health domain are correlated.	r > 0.4	0.389[a]	<0.001	Rejected
2. Δ SarQoL® Overall score and Δ SF-36 Vitality domain are correlated.	r > 0.3	0.460[b]	<0.001	Confirmed
3. Δ SarQoL® Overall score and Δ SF-36 Physical Functioning domain are correlated.	r > 0.5	0.690[a]	<0.001	Confirmed
4. Δ SarQoL® Overall score and Δ EQ-VAS are correlated.	r > 0.4	0.226[a]	<0.027	Rejected
5. Δ SarQoL® domain 1 (Physical & Mental Health) and Δ SF-36 General Health domain are correlated.	r > 0.3	0.139[a]	0.176	Rejected
6. Δ SarQoL® domain 1 (Physical & Mental Health) and Δ EQ-VAS are correlated.	r > 0.3	0.142[a]	0.166	Rejected
7. Δ SarQoL® domain 2 (Locomotion) and Δ SF-36 Physical Functioning domain are correlated.	r > 0.4	0.539[a]	<0.001	Confirmed
8. Δ SarQoL® domain 4 (Functionality) and Δ SF-36 Physical Functioning domain are correlated.	r > 0.5	0.601[a]	<0.001	Confirmed
9. Δ SarQoL® domain 5 (Activities of Daily Living) and Δ SF-36 Physical Functioning domain are correlated.	r > 0.5	0.617[a]	<0.001	Confirmed

Δ = change over 4 years; *r* = correlation
[a] Spearman correlation
[b] Pearson correlation

Table 6: Standardized Response Means		
Domains	Corrected SRM	Interpretation[a]
Δ SarQoL® D1 Physical & Mental Health	-0.383	small
Δ SarQoL® D2 Locomotion	-0.755	moderate
Δ SarQoL® D3 Body Composition	-0.315	small
Δ SarQoL® D4 Functionality	-0.940	large
Δ SarQoL® D5 Activities of Daily Living	-0.883	large
Δ SarQoL® D6 Leisure activities	-0.255	small
Δ SarQoL® D7 Fears	-0.070	trivial
Δ SarQoL® Overall score	-1.144	large
Δ SF-36 Physical Functioning	-0.749	moderate
Δ SF-36 Social Functioning	-0.204	small
Δ SF-36 Role Limitations due to Physical Health	-0.301	small
Δ SF-36 Role Limitations due to Emotional Problems	-0.251	small
Δ SF-36 Mental Health	-0.274	small
Δ SF-36 Vitality	-0.577	moderate
Δ SF-36 Bodily Pain	-0.490	small
Δ SF-36 General Health	-0.693	moderate
Δ SF-36 Physical Component Summary	-0.634	moderate
Δ SF-36 Mental Component Summary	-0.224	small
Δ EQ-5D Utility Index	0.064	trivial
Δ EQ-VAS	-0.267	small

[a] Interpretation of corrected SRMs: $0.20 \leq SRM < 0.49$=small change; $0.50 \leq SRM < 0.79$= moderate change; $SRM \geq 0.80$= large change
SRM are calculated by dividing the mean difference between scores from the first year and the 5th year by the standard deviation of the differences between these paired values. The SRM values were subsequently corrected with the formula $SRM / \sqrt{2} / \sqrt{(1-r)}$, where 'r' signifies the correlation between the year 1 and year 5 scores.
Δ = change over 4 years

5. Discussion

This study examined whether the SarQoL questionnaire could be used as a disease-specific instrument to measure health-related QoL in frailty. The psychometric results presented in this article indicate that it has adequate measurement properties when used with the Fried frailty criteria. This means that the SarQoL could be a new option for researchers seeking to evaluate QoL in populations characterized by the presence of pre-frailty and/or frailty.

This study demonstrated that the SarQoL questionnaire can discriminate between robust, pre-frail, and frail subjects, with declining QoL scores according to the category of frailty, and that it can do so over a wide range of concepts. The systematic review by Crocker et al. highlighted that (sub)scales measuring physical aspects of QoL were broadly able to discriminate between robust and frail people but reported inconsistent results for other aspects of QoL [7]. Therefore, it is encouraging to see that the SarQoL questionnaire is able to discriminate on more than just the physical aspects of QoL and that it brings extra precision in being able to discriminate between robust, frail, and pre-frail individuals. A note of caution is warranted with regard to Domain 7, where only the comparison between robust and frail participants yielded significantly different QoL scores. This domain should not be interpreted in a vacuum but taking into account the other domain scores and the overall QoL score.

The internal consistency was shown to be high ($\alpha = 0.866$), indicating that the domains in the questionnaire are highly interrelated and measure the same construct, QoL. Mixed results were obtained in the evaluation of the construct validity of the questionnaire. All four hypotheses on the convergent validity were confirmed, but two out of the four hypotheses for divergent validity were rejected. The two rejected hypotheses, where we found stronger correlations than expected, were between the overall QoL score of the SarQoL questionnaire and the mental health and role limitations due to emotional problems domains of the SF-36. It may be that our hypotheses are erroneous and that these two domains are conceptually closer to the SarQoL questionnaire than we theorized. One correlation of particular interest is between the SarQoL overall QoL score and the SF-36 role limitation due to physical limitations in the frail group ($r = 0.246$), because it is significantly lower than the correlation coefficients in the robust ($r = 0.408$) and pre-frail groups ($r = 0.611$). Upon further investigation, this discrepancy is linked to the significant floor effect in this SF-36 domain, where 16 of the 30 participants obtain the lowest score possible.

The test–retest reliability of the questionnaire was excellent, with an ICC of 0.918 (95% CI = 0.834–0.961) for the overall score. However, because the original study only contacted the participants diagnosed as sarcopenic according to the diagnostic criteria of the European Working Group on Sarcopenia in Older People to enter the evaluation of the test–retest reliability, there were only data available for 29 participants. So, while this is a result that indicates good test–retest reliability, with an

elevated ICC and a relatively small CI, these results should be confirmed in a larger sample and in particular samples with sufficient pre-frail and frail participants to calculate ICCs for these particular groups. Because the SEM and the smallest detectable change are based on the test–retest data, this same remark also applies to these two indicators. It should also be noted that, in this study, Domain 6 (leisure activities) and Domain 7 (fears) did not demonstrate adequate reliability. We hypothesize that this may because of the low sample size in combination with the low number of items for these two domains (two items for Domain 6 and four items for Domain 7), which causes any difference between the responses between the first and second administration of the questionnaire to be exaggerated in the scores.

We examined the ability of the SarQoL questionnaire to detect a change in QoL. We found moderate responsiveness through the confirmation of five out of nine hypotheses on the correlation between changes in QoL observed by the SarQoL questionnaire and by other questionnaires. It is possible that the rejection of several hypotheses is linked to the lower SRMs found between the different questionnaires. In fact, the SRM of the overall QoL score of the SarQoL questionnaire is markedly stronger at SRM = −1.144 compared with the strongest effect size of the SF-36, which was the physical functioning subscale at SRM = −0.749. It may be that the rejection of some hypotheses was thus caused not by poor responsiveness of the SarQoL questionnaire but by smaller effect sizes found by the SF-36. Similarly, for the EQ-VAS, we found a small effect at SRM = −0.267 and the rejection of two hypotheses associated with this instrument. Here also, this may be more linked to the performance of the EQ-VAS in combination with the 4-year interval between the assessments. It is highly likely that an instrument such as the EQ-VAS would be influenced by response shift, which is defined as a change in the self-evaluation of the meaning of a target construct caused by reconceptualization of the construct, a reprioritization of the participants' values, or a recalibration of the respondents' internal standards of measurement [38, 45]. Overall assessments, such as the EQ-VAS, which asks the respondent to indicate on a scale from 0 to 100 ‘how good or bad your health state is today’, are more vulnerable to response shift because they require careful consideration and interpretation of the question. The participants had to evaluate for themselves the meaning of the concept ‘health state’ and what is considered ‘good’ and ‘bad’ and assign a numerical value to this, leaving open the possibility of reconceptualization, reprioritization, or recalibration [46]. Researchers investigating changes in QoL over time or pre-intervention/post-intervention should make the overall QoL score of the SarQoL questionnaire their main outcome, given that it has the highest SRM and the smallest detectable change. If a significant change in overall QoL is found, further analyses of the individual domains could be useful in indicating on what domains a participant's QoL has changed.

Because this study used data collected during a previous study, we were unable to investigate and quantify the content validity of the SarQoL questionnaire in a population of frail, older, community-dwelling individuals. In the development of the questionnaire, content validity had been put at the heart

of the process by soliciting, at each step of the item generation and selection process, input from multiple sarcopenic persons [11]. In this study, we were unable to provide this information from frail individuals. However, some authors have theorized that sarcopenia, the target condition for which the SarQoL questionnaire was developed, constitutes one of the main components of the clinical frailty syndrome, all the while recognizing that frailty should not be limited to physical manifestations but should also incorporate psychological, cognitive, emotional, social, and spiritual factors [24, 47] Currently, to our knowledge, the only questionnaire that measures QoL and that is specifically designed with and for older frail persons is the Geriatric Quality of Life Questionnaire [48]. However, the developers left the definition of what constitutes the 'frail elderly' up to the appreciation of the clinicians responsible for recruitment, instead of a recognized diagnostic tool. While the SarQoL questionnaire was not specifically developed for frailty, the shared characteristics between sarcopenia and frailty mean that it should be able to provide a precise measurement of the physical weakness aspect of frailty. Apart from the physical domains, the SarQoL has also incorporated items on mental health, body image, sexuality, activities of daily living, leisure activities, and fears, making for a multidimensional framework of QoL.

Healthy ageing is already high on the agenda for most health systems in both Western and Asian countries and will only gain in importance as the number of older people increases [49]. Concepts such as frailty, sarcopenia, or the construct recently proposed by the World Health Organization called Intrinsic Capacity, which is a composite of all the physical and mental capacities of an individual, may play an important role in any future medical approach [50]. Whatever approach is adopted, it must take in the perspective and priorities of the target population, and QoL can be an important metric for this. Having valid, reliable, and precise instruments to measure QoL that can pick up on the impact of a specific target condition is a prerequisite to be able to rely on QoL instruments to provide information on the patients' lived experience.

There are some limitations to this study. First off, we adopted the frailty criteria developed by Fried et al., but other diagnostic approaches are available, such as the Rockwood Clinical Frailty Scale or the IF-VIG, among others [51-53]. Although all these approaches purportedly measure the same concept, frailty, we cannot be sure that the results on the validity and reliability of the SarQoL would have been the same if we had applied other diagnostic approaches. Secondly, our sample of robust, pre-frail, and frail participants is not necessarily representative of frailty in the wider community. Because these data were collected within a study that recruited volunteers, and which asked those volunteers to make several trips to the research centre, it is likely that the SarcoPhAge study recruited a sample that was in better overall condition, and that had better mobility, than a representative sample of pre-frail and frail participants. While this study has shown that the SarQoL questionnaire is a valid and reliable tool in frailty, additional investigations in samples with a different make-up need to confirm these results. Lastly, while the overall sample size was more than adequate for a psychometric study, the test–retest sample is relatively small with only 29 participants. This steep reduction from the overall sample size

is a result of the fact that only a subset of participants was invited to complete the questionnaire a second time. However, because we have the 95% CI around the ICC, we can judge that most values have adequate precision, apart from Domains 6 and 7.

In conclusion, the study evaluated the validity and reliability of the SarQoL questionnaire in frailty and found that it is a valid and reliable tool for the assessment of QoL. Because of the shared mechanism of physical weakness between sarcopenia and frailty, the SarQoL questionnaire can provide more specific information on QoL in frailty than the generic questionnaires available.

6. Annexes

Supplemental table: Frailty criteria and diagnosis		
Criteria	Cut-off	Scoring
Involuntary weight-loss	>4.5 kg in 1 year (if during follow-up visit: more than 5% weight loss since previous visit)	Yes=1 No=0
Handgrip strength	Dominant hand, 3 repetitions, highest value Men: BMI ≤24 => GS ≤29 kg BMI 24.1 - 26 => GS ≤30 kg BMI 26.1 - 28 => GS ≤30 kg BMI >28 => GS ≤32 kg Women: BMI ≤23 => GS ≤17 kg BMI 23.1 - 26 => GS ≤17.3 kg BMI 26.1 - 29 => GS ≤18 kg BMI >29 => GS ≤21 kg	Yes=1 No=0
Exhaustion "How many times during the past week have you felt like: Everything I do requires effort. I 'm not going to be able to continue like this."	A & B 0= rare or never (<1 day) 1= sometimes (1-2 days) 2= occasionally (3-4 days) 3= often (5-7 days)	If the participant gives response 2 or 3 for one or both questions, score 1.
Gait speed over 4.5m	Time to walk 4.5m recalculated from SPPB results on a 4m track. Men: Height ≤173 cm => ≥7 sec Height >173 cm => ≥6 sec Women Height ≤159 cm => ≥7 sec Height >159 cm => ≥6 sec	Yes=1 No=0
Physical activity level (Minnesota Leisure Time Activity questionnaire)	Men: <283 KCAL/week Women: <270 KCAL/week	Yes=1 No=0

7. References

1. World Health Organisation. Decade of healthy ageing. 2020 https://www.who.int/docs/default-source/decade-of-healthy-ageing/final-decade-proposal/decade-proposal-final-apr2020-en.pdf?sfvrsn=b4b75ebc_3
2. WHO Clinical Consortium on Healthy Ageing. Report of consortium meeting 1–2 December 2016 in Geneva, Switzerland. Geneva; 2017.
3. Vermeiren S, Vella-Azzopardi R, Beckwée D, Habbig AK, Scafoglieri A, Jansen B, et al. Frailty and the prediction of negative health outcomes: a meta-analysis. J Am Med Dir Assoc 2016; 17: 1163.e1– 1163.e17.
4. Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of frailty in community-dwelling older persons: a systematic review. J Am Geriatr Soc 2012; 60: 1487– 1492.
5. O'Caoimh R, Galluzzo L, Rodríguez-Loso A, Van der Heyden J, Ranhoff A, Lemprini-Koula M, et al. Prevalence of frailty at population level in European ADVANTAGE Joint Action Member States: a systematic review and meta-analysis. Ann Ist Super Sanita 2018; 54: 226– 238.
6. Buckinx F, Rolland Y, Reginster J-Y, Ricour C, Petermans J, Bruyère O. Burden of frailty in the elderly population: perspectives for a public health challenge. Arch Public Health 2015; 73: 19.
7. Crocker TF, Brown L, Clegg A, Farley K, Franklin M, Simpkins S, et al. Quality of life is substantially worse for community-dwelling older people living with frailty: systematic review and meta-analysis. Qual Life Res 2019; 28: 2041– 2056.
8. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. Pharmacoeconomics 2000; 17: 13– 35.
9. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol 2003; 56: 52– 60.
10. Beaudart C, Biver E, Bruyère O, Cooper C, Al-Daghri N, Reginster JY, et al. Quality of life assessment in musculo-skeletal health. Aging Clin Exp Res 2018; 30(5): 413– 418.
11. Beaudart C, Biver E, Reginster JY, Rizzoli R, Rolland Y, Bautmans I, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. Age Ageing 2015; 44: 960– 966.

12. Beaudart C, Edwards M, Moss C, Reginster JY, Moon R, Parsons C, et al. English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. Age Ageing 2017; 46: 271– 277.

13. Beaudart C, Biver E, Reginster JY, Rizzoli R, Rolland Y, Bautmans I, et al. Validation of the SarQoL®, a specific health-related quality of life questionnaire for sarcopenia. J Cachexia Sarcopenia Muscle 2017; 8: 238– 244.

14. Geerinck A, Scheppers A, Beaudart C, Bruyère O, Vandenbussche W, Bautmans R, et al. Translation and validation of the Dutch SarQoL® , a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact 2018; 18: 463– 472.

15. Konstantynowicz J, Abramowicz P, Glinkowski W, Taranta E, Marcinowicz L, Dymitrowicz M, et al. Polish validation of the SarQoL®, a quality of life questionnaire specific to sarcopenia. J Clin Med 2018; 7: 323.

16. Gasparik AI, Mihai G, Beaudart C, Bruyere O, Pop R-M, Reginster J-Y, et al. Correction to: psychometric performance of the Romanian version of the SarQoL®, a health-related quality of life questionnaire for sarcopenia. Arch Osteoporos 2018; 13: 98.

17. Fábrega-Cuadros R, Martínez-Amat A, Cruz-Díaz D, Aibar-Almazán A, Hita-Contreras F. Psychometric properties of the Spanish version of the sarcopenia and quality of life, a quality of life questionnaire specific for sarcopenia. Calcif Tissue Int 2020; 106: 274– 282.

18. Alekna V, Kilaite J, Tamulaitiene M, Geerinck A, Mastaviciute A, Bruyère O, et al. Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). Eur Geriatr Med 2019; 10: 761– 767.

19. Tsekoura M, Billis E, Gliatis J, Tsepis E, Matzaroglou C, Sakkas GK, et al. Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. Disabil Rehabil 2020; 42(7): 1006– 1012.

20. Geerinck A, Alekna V, Beaudart C, Bautmans I, Cooper C, De Souza Orlandi F, et al. Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: an analysis of subjects from 9 validation studies. PLoS One 2019; 14:e0216065.

21. Geerinck A, Bruyère O, Locquet M, Reginster JY, Beaudart C. Evaluation of the responsiveness of the SarQoL® questionnaire, a patient-reported outcome measure specific to sarcopenia. Adv Ther 2018; 35: 1842– 1858.

22. Cederholm T. Overlaps between frailty and sarcopenia definitions. Nestle Nutr Inst Workshop Ser 2015; 83: 65– 69.

23. Sieber CC. Frailty—from concept to clinical practice. Exp Gerontol 2017; 87: 160– 167.

Crossref PubMed Web of Science®Google Scholar

24. Cruz-Jentoft AJ, Michel JP. Sarcopenia: a useful paradigm for physical frailty. Eur Geriatr Med 2013; 4: 102– 105.

25. Beaudart C, Reginster JY, Petermans J, Gillain S, Quabron A, Locquet M, et al. Quality of life and physical components linked to sarcopenia: the SarcoPhAge study. Exp Gerontol 2015; 69: 103– 110.

26. Locquet M, Beaudart C, Hajaoui M, Petermans J, Reginster JY, Bruyère O. Three-year adverse health consequences of sarcopenia in community-dwelling older adults according to 5 diagnosis definitions. J Am Med Dir Assoc 2019; 20: 43– 46, e2.

27. Beaudart C, Locquet M, Touvier M, Reginster J-Y, Bruyère O. Association between dietary nutrient intake and sarcopenia in the SarcoPhAge study. Aging Clin Exp Res 2019; 31: 815– 824.

28. Beaudart C, Sanchez-Rodriguez D, Locquet M, Reginster JY, Lengelé L, Bruyère O. Malnutrition as a strong predictor of the onset of sarcopenia. Nutrients 2019; 11.

29. Locquet M, Beaudart C, Reginster JY, Bruyère O. Association between the decline in muscle health and the decline in bone health in older individuals from the SarcoPhAge cohort. Calcif Tissue Int 2019; 104: 273– 284.

30. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001; 56: M146– M156.

31. Taylor HL, Jacobs DR, Schucker B, Knudsen J, Leon AS, Debacker G. A questionnaire for the assessment of leisure time physical activities. J Chronic Dis 1978; 31: 741– 755.

32. Roadolff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Measur 1977; 1: 385– 401.

33. Safonova YA, Lesnyak OM, Baranova IA, Suleimanova AK, Zotkin EG. Russian translation and validation of SarQoL®—quality of life questionnaire for patients with sarcopenia. Nauchno-Prakticheskaya Revmatol 2019; 57: 38– 45.

34. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care 1992; 30: 473– 483.

35. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001; 33: 337– 343.

36. The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy (New York) 1990; 16: 199– 208.

37. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010; 63: 737– 745.

38. de Vet HCW, Terwee CB, Mokkink LB. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press; 2011.

39. Davidson M. Known-Groups Validity. In Encyclopedia of Quality of Life and Well-being Research. Dordrecht: Springer Netherlands; 2014. p 3481– 3482.

40. de Boer MR, Moll AC, de Vet HCW, Terwee CB, Völker-Dieben HJM, van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. Ophthalmic Physiol Opt 2004; 24: 257– 273.

41. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000; 53: 459– 468.

42. Terwee C, Bot S, de Boer M, van der Windt D, Knol D, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60: 34– 42.

43. Middel B, Van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. Int J Integr Care 2002; 2.

44. Cohen J. Quantitative methods in psychology. Psychol Bull 1992; 112: 155– 159.

45. Sprangers M, Schwartz C. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999; 48: 1507– 1515.

46. Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. Health Qual Life Outcomes 2004; 2: 1– 11.

47. Rodríguez-Mañas L, Féart C, Mann G, Viña J, Chatterji S, Chodzko-Zajko W, et al. Searching for an operational definition of frailty: a Delphi method based consensus statement. The Frailty Operative Definition-Consensus Conference Project. J Gerontol - Ser A Biol Sci Med Sci 2013; 68: 62– 67.

48. Guyatt GH, Eagle J, Sacket B, Willan A, Griffith L, McIlroy W, et al. Measuring quality of life in the frail. J Clin Epidemiol 1993; 46: 1433– 1444.

49. Xia W, Cooper C, Li M, Xu L, Rizzoli R, Zhu M, et al. East meets West: current practices and policies in the management of musculoskeletal aging. Aging Clin Exp Res 2019; 31: 1351– 1373.

50. Cesari M, De Carvalho IA, Thiyagarajan JA, Cooper C, Martin FC, Reginster JY, et al. Evidence for the domains supporting the construct of intrinsic capacity. J Gerontol - Ser A Biol Sci Med Sci 2018; 73: 1653– 1660.

51. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: a review. Eur J Intern Med 2016; 31: 3– 10.

52. Woolford SJ, Sohan O, Dennison EM, Cooper C, Patel HP. Approaches to the diagnosis and prevention of frailty. Aging Clin Exp Res 2020; 32: 1629– 1637.

53. Moreno-Ariño M, Torrente Jiménez I, Cartanyà Gutiérrez A, Oliva Morera JC, Comet R. Assessing the strengths and weaknesses of the Clinical Frailty Scale through correlation with a frailty index. Aging Clin Exp Res 2020; 32: 2225– 2232.

54. von Haehling S, Morley JE, Coats AJS, Anker SD. Ethical guidelines for publishing in the Journal of Cachexia, Sarcopenia and Muscle: update 2019. J Cachexia Sarcopenia Muscle 2019; 10: 1143– 1145.

Chapter 5: Screening for sarcopenia with the SarQoL questionnaire

Assessment of the performance of the SarQoL® questionnaire in screening for sarcopenia in older people.

Published as:

Geerinck, A., Dawson-Hughes, B., Beaudart, C. et al. Assessment of the performance of the SarQoL® questionnaire in screening for sarcopenia in older people. Aging Clin Exp Res 33, 2149–2155 (2021). https://doi.org/10.1007/s40520-021-01913-z

1. Abstract

Background: Because of its low prevalence and the need for physical tests to establish a diagnosis, recruiting sarcopenic people for clinical studies can be a resource-intensive process.

Aims: We investigated whether the SarQoL®, a 55-item questionnaire designed to measure quality of life in sarcopenia, could be used to identify older people with a high likelihood of being sarcopenic, and to compare its performance to the SARC-F tool.

Methods: We performed a secondary analysis of data from older, community-dwelling participants of the SarcoPhAge study, evaluated for sarcopenia according to the EWGSOP2 criteria, and who completed the SarQoL® and SARC-F questionnaires. We determined the optimal threshold to distinguish between sarcopenic and non-sarcopenic people with the Youden index. Screening performance was evaluated with the area under the curve (AUC) and by calculating sensitivity and specificity.

Results: The analysis of 309 participants provided an optimal threshold value of ≤ 52.4 points for identifying people with sarcopenia with the SarQoL® questionnaire, which resulted in a sensitivity of 64.7% (41.1–84.2%), a specificity of 80.5% (75.7–84.7%) and an AUC of 0.771 (0.652–0.889). Compared to the SARC-F, the SarQoL® has greater sensitivity (64.7% vs 52.39%), but slightly lower specificity (80.5% vs. 86.6%).

Discussion: The SarQoL® questionnaire showed acceptable screening accuracy, on par with the SARC-F. The optimal threshold of ≤ 52.4 points should be confirmed in other cohorts of older people.

Conclusions: This exploratory study showed that the SarQoL® could potentially be applied in a screening strategy, with the added benefit of providing a measure of QoL at the same time.2. Introduction

2. Introduction

Sarcopenia has been described by the 2nd European Working Group on Sarcopenia in Older People (EWGSOP2) as a "progressive and generalised skeletal muscle disorder that is associated with increased likelihood of adverse outcomes including falls, fractures, physical disability and mortality". In the same article, the EWGSOP2 also presented a revision of its diagnostic criteria for sarcopenia, presenting a new diagnostic algorithm and changing the threshold values for low muscle strength and low muscle mass [1]. This revision has increased the consistency between studies in the evaluation of sarcopenia, but some studies have observed that it lowers the prevalence of sarcopenia compared to the EWGSOP1 criteria [2, 3]. For clinical research and epidemiological studies, this means that more candidates need to be evaluated to achieve a sufficient number of sarcopenic participants to obtain the desired statistical power.

To help researchers recruit sarcopenic individuals in an efficient and cost-effective manner, multiple screening tools have been developed to identify those candidates with the highest probability of having sarcopenia. These come in different forms: there are questionnaires such as the Mini Sarcopenia Risk Assessment (MSRA—both a 7 and 5-item version available) and the SARC-F questionnaire (a 5 and 3-item version exist, as well as a version with calf circumference and a version which takes into account age and body mass) [4]. Other screening instruments rely solely on physical characteristics, such as the score developed by Ishii et al (age, grip strength and calf circumference), muscle mass prediction formulas or the chair stand test [4, 5].

Clinical studies in sarcopenia require a substantial amount of time and effort, because of the need to include and evaluate a large number of candidates to find sufficient sarcopenic subjects to achieve the required level of statistical power. A full diagnostic evaluation where muscle mass is evaluated by dual-energy X-ray absorptiometry (DXA) and muscle strength by dynamometer, as recommended, necessitates the use of qualified personnel and expensive instruments. Given the cost per patient for these evaluations, screening instruments that can significantly increase the proportion of sarcopenic persons within the pool of candidates invited for a full body composition assessment could greatly help the financial feasibility of large-scale clinical studies in sarcopenia. With this in mind, the hypothesis was raised that an existing instrument, developed to measure quality of life in sarcopenia, could potentially be of use in screening candidates for referral to full body composition evaluation and/or physical function assessment.

The instrument investigated in this study is the Sarcopenia Quality of Life (SarQoL®) questionnaire. It evaluates quality of life in sarcopenia through 55 items categorized into 7 domains of health-related dysfunction [6]. It is an auto-administered instrument and takes about 15 min to complete. Its clinimetric properties as a QoL questionnaire have been demonstrated in multiple validation studies

conducted in multiple languages [7,8,9,10,11,12,13,14,15,16,17,18]. Of particular interest in this context is the repeated observation that the SarQoL® questionnaire is able to discriminate between sarcopenic and non-sarcopenic groups, with the former scoring significantly lower on the overall QoL score of the questionnaire compared to the latter. Its focus on the impact of musculoskeletal health on quality of life contributes to our expectation that the overall QoL score produced by the SarQoL® questionnaire could be used to screen older people and identify those with a higher likelihood of sarcopenia.

The objective of this study is therefore to evaluate the capacity of the Overall QoL score of the SarQoL® questionnaire to detect individuals with sarcopenia according to the revised EWGSOP2 consensus criteria. The hypothesis linked to this objective is that the Overall QoL score of the SarQoL® questionnaire has an area under the ROC curve (AUC) greater than 0.7, indicating the test is useful in distinguishing between sarcopenic and non-sarcopenic people [19].

The secondary objective of this study is to compare the screening performance of the Overall QoL score of the SarQoL® questionnaire with the performance of the 5-item SARC-F questionnaire, the screening tool recommended by the EWGSOP2 [1]. The hypothesis linked to this objective is that the Overall QoL score is at least as accurate as the SARC-F, judged by AUC, sensitivity and specificity.

3. Material and methods

This study is a cross-sectional secondary evaluation of data collected at the third year of follow-up of the Sarcopenia and Physical Impairment with advancing Age (SarcoPhAge) prospective cohort study, carried out in the Liège province of Belgium [20]. The SarcoPhAge study was conducted in compliance with the principles outlined in the Declaration of Helsinki. The study protocol and its amendments received approval from the Ethics Committee of the University Teaching Hospital of Liège (n° 2012-277), and all participants provided written informed consent. This article was written to comply, as much as feasible, with the most recent version of the Standards for Reporting Diagnostic Accuracy (STARD) checklist [21].

3.1. Participants

The SarcoPhAge study enrolled a convenience sample of people who visited an outpatient clinic in Liège (Belgium) as well as people who responded to a press advertisement between June 2013 and July 2014. Participants in this study were 65 years of age or older, and, because of the limitations of the dual-

energy X-ray absorptiometry (DXA) instrument, people with a BMI above 50 kg/m^2 or with amputated limbs were not eligible. There were no additional inclusion criteria beyond these [20]. The third year of follow-up (July 2015–2016) was selected for inclusion because this was the first year that both the SarQoL® questionnaire and the SARC-F questionnaire were administered to all participants.

3.2. Measurements

For each participant, muscle mass was measured with a dual-energy X-ray absorptiometry instrument (Hologic Discovery A, USA) and grip strength with the Saehan hydraulic hand dynamometer (Saehan Corp., Masan, South Korea). Both instruments were calibrated according to the respective manufacturer's instructions at the recommended intervals. Appendicular skeletal muscle mass was calculated as the sum of all 4 limbs and divided by the squared height of the participant in question to obtain a skeletal muscle mass index (SMI = ASM/Ht^2). The grip strength of a person was defined as the highest value out of 6 measurements (3 for the dominant hand and 3 for the non-dominant hand). Detailed descriptions of both measurements are available in the article on the baseline results of the SarcoPhAge study [20]. These data allowed us to diagnose sarcopenia according to the EWGSOP2 criteria in participants with low muscle mass (ASM/Ht^2 < 7.0 kg/m^2 for men and < 5.5 kg/m^2 for women) and low muscle strength (grip strength < 27 kg for men and < 16 kg for women) [1]. Sarcopenia diagnosed with the EWGSOP2 criteria constitutes the reference standard in this study because of its status as the current consensus criteria and its applicability to samples recruited in Europe [1].

The index test in this study, the paper-based French-language SarQoL® questionnaire, was completed by the participants without assistance. An Overall QoL score (0–100 points) is calculated where lower scores indicate lower QoL, and thus also greater sarcopenia-related disability [6, 22]. The questionnaire is available in multiple languages from the website www.sarqol.org, and the Overall QoL score was calculated with an Access database developed for this purpose. Given the exploratory nature of this investigation, we did not pre-specify a test-positivity cut-off point.

We included a second index test in this analysis, so as to be able to compare the performance of the SarQoL® questionnaire against the current most widely used screening instrument in sarcopenia, the SARC-F [23]. It is composed of 5 questions on strength, locomotion, rising from a chair, climbing stairs and history of falls. A total score is calculated and ranges from 0 to 10 points, where higher scores are linked with a higher probability of being diagnosed with sarcopenia. A score of ≥ 4 points is used as a cut-off to identify individuals who require a full examination for sarcopenia in clinical practice [23]. The SARC-F was developed to be able to detect sarcopenia as diagnosed with the EWGSOP1 criteria, and a meta-analysis found a pooled sensitivity of 0.21 (0.13–0.31) combined with a specificity of 0.90 (0.83–0.94) [24]. With the publication of the revised EWGSOP2 criteria, several authors have looked

again at the performance of the SARC-F, and a meta-analysis that pooled the results from 4 studies found an AUC of 0.75 (95% CI 0.71–0.78) with a sensitivity of 0.77 (95% CI 0.49–0.92) and a specificity of 0.63 (95% CI 0.43–0.79), while the same meta-analysis found a pooled sensitivity of 0.32 (95% CI 0.19–0.47) and specificity of 0.86 (95% CI 0.77–0.92) for the EWGSOP1 criteria in 13 studies [25].

To compare the performance of the SarQoL® questionnaire and the SARC-F instrument with a screening instrument based on physical indicators, we calculated the probability of sarcopenia according to the Ishii formula, which was the best-performing screening instrument in a comparison of 5 with the EWGSOP1 criteria [5]. For men, we used the formula [0.62 × (age−64) −3.09 × (grip strength−50) −4.64 × (calf circumference−42)] to calculate the sum score and the formula [1 / (1 + $e^{-(\text{sum score} / 10 - 11.9)}$)] to calculate the probability of sarcopenia (expressed in percentage). For women, the formula [0.80 × (age−64) −5.09 × (grip strength−34) −3.28 × (calf circumference−42)] provided the sum score and the formula [1 / (1 + $e^{-(\text{sum score} / 10 - 12.5)}$) the probability of sarcopenia [26]. A sum score higher than 105 for men and 120 for women was used as the cut-off for a high probability of sarcopenia [26]. To the best of our knowledge, its performance when used to screen patients for sarcopenia with the EWGSOP2 criteria has not yet been established.

The reference test and the index tests were performed by the same investigator or completed by the participant during a single study visit. The study investigator also recorded clinical and demographic information needed for the Ishii formula. The results from the reference test and one of the index tests, the SARC-F, was directly available to the investigator. The SarQoL® Overall score and the Ishii score, were calculated some time after the end of the study visit.

3.3. Statistical analyses

Statistical analyses were carried out with the Statistical Package for the Social Sciences version 27.0.0.0 (SPSS Statistics; IBM, Armonk, NY). The distribution of variables in this analysis was examined by looking at the distance between median and mean, histogram, QQ-plot, and the Shapiro–Wilk test. Continuous variables are presented as mean ± standard deviation if normally distributed and as median (25th–75th percentile) if not normally distributed. The evaluation of the screening performance of the Overall QoL score of the SarQoL® questionnaire, the SARC-F tool and the Ishii screening test was based on their sensitivity (Se), specificity (Sp), positive likelihood ratio (LR +), negative likelihood ratio (LR-), positive predictive value (PPV), and negative predictive value (NPV) in relation to sarcopenia as diagnosed with the EWGSOP2 criteria. These values and the associated 95% confidence intervals were obtained through the GENLIN procedure, as outlined in document 422875 from IBM support [27]. Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC)

provided the overall accuracy of the three screening instruments. An AUC value above 0.9 indicates high accuracy of the screening instrument, between 0.8 and 0.9 excellent accuracy and between 0.7 and 0.8 acceptable accuracy [19]. The Youden J statistic (sensitivity + specificity −1) was used to find the optimal cut-point for the Overall SarQoL score [28]. The analyses presented in this article have been performed in all participants who were assessed for sarcopenia using the EWGSOP2 criteria, screened with the SARC-F questionnaire and who completed the SarQoL® questionnaire at the third follow-up of the SarcoPhAge study. A *p*-value of 0.05 was considered statistically significant.

4. Results

A total of 309 people were included in this analysis. All participants were assessed for sarcopenia with the EWGSOP2 criteria in the third yearly evaluation of the SarcoPhAge study, and 17 (5.5%) of them were diagnosed with sarcopenia. The sarcopenic participants were older than those not diagnosed as sarcopenic [80.07 (71.98 – 86.36) years versus 73.55 (69.68 –78.58) years, $p = 0.011$]. They also took more medication and had a lower gait speed than those not diagnosed with sarcopenia. The complete clinical characteristics for the sample are detailed in Table 1.

Table 1: Clinical characteristics

	Sarcopenic (n=17)	Not sarcopenic (n=292)	p-value [a]
Age (years)	80.07 (71.98 – 86.36)	73.55 (69.68 – 78.58)	0.011
Gender (women)	10 (58.8%)	170 (58.2%)	0.961
BMI (kg/m²)	23.91 (19.01 – 26.58)	26.74 (23.97 – 29.57)	0.001
N° of drugs	9.00 (3.50 – 12.50)	6.00 (4.00 – 8.00)	0.035
N° of comorbidities	4.00 (3.00 – 7.00)	4.00 (2.00 - 5.00)	0.462
Gait speed (m/s)	0.70 ± 0.27	1.14 ± 0.28	<0.001

[a] P-values from Mann-Whitney U-test, Pearson Chi-square or Student t-test, depending on variable characteristics.

The SARC-F questionnaire identified 48 participants (15.5% of the sample) with a score ≥ 4 points and thus suspected of having sarcopenia. A ROC curve of the SarQoL® Overall score and the SARC-F score is presented in Fig. 1. The AUC for the SarQoL® Overall score is 0.771 (95% CI: 0.652–0.889), and for the SARC-F 0.802 (95% CI: 0.696–0.909).

The Youden index was maximised at ≤ 52.4 points for the SarQoL® Overall score ($J_c = 0.452$, Se = 0.647, Sp = 0.805). This threshold value, together with the prespecified threshold for the SARC-F, were used for the construction of Table 2, detailing the screening accuracy of the two instruments.

The SarQoL® Overall score, dichotomized at ≤ 52.4 points, had, in absolute numbers, slightly greater sensitivity than the SARC-F score (64.7% vs. 52.9%), because it correctly identified 11 out of the 17 sarcopenic participants, whereas the SARC-F correctly identified 9 out of 17. In terms of their specificity, the SARC-F had, in absolute numbers, slightly greater specificity than the SarQoL® Overall score (80.5% vs. 86.6%), with 253 non-sarcopenic subjects correctly identified compared to the 235 found by the SarQoL® questionnaire.

The Ishii screening test outperformed both the SARC-F and the SarQoL® Overall score with an AUC of 0.884 (95% CI: 0.840–0.927). The Ishii screening test correctly identified all 17 sarcopenic individuals, and therefore had a sensitivity of 100%, and correctly identified 224 non-sarcopenic individuals for a specificity of 76.7%. It flagged a total of 85 people as being at high risk for sarcopenia, which is 27.5% of the total sample.

Figure 1: ROC curves for the SarQoL® Overall score and the SARC-F tool

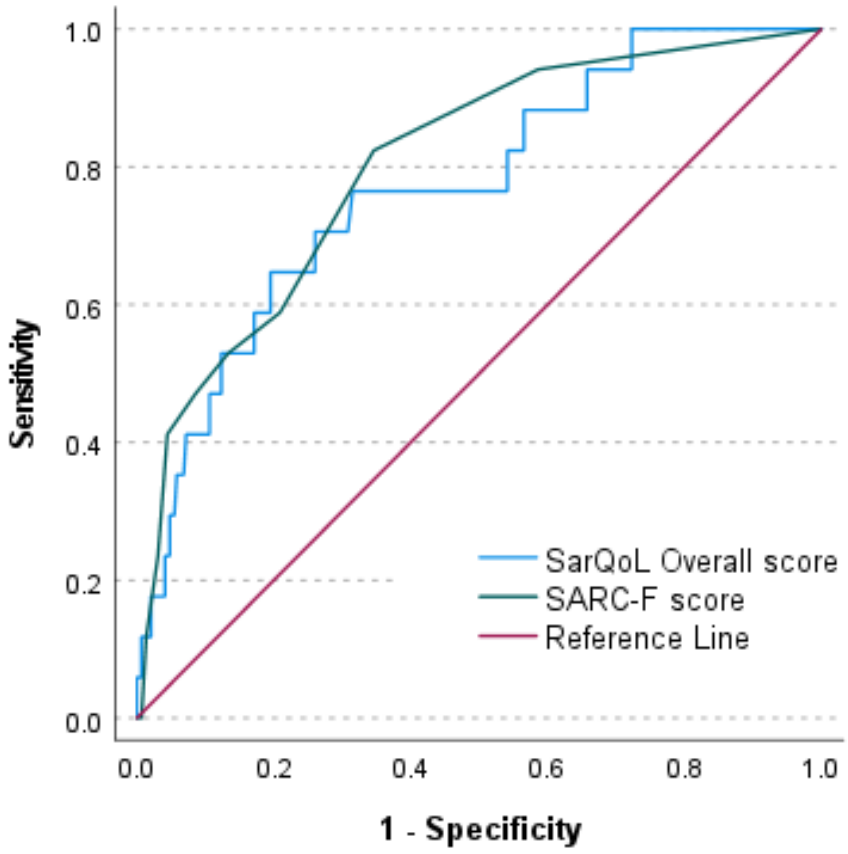


Table 2: Screening accuracy of the SarQoL® Overall score and the SARC-F instrument		
	SarQoL	SARC-F
True positives	11	9
False positives	57	39
True negatives	235	253
False negatives	6	8
Sensitivity	0.647 (0.411 – 0.842)	0.529 (0.301 – 0.750)
Specificity	0.805 (0.757 – 0.847)	0.866 (0.824 – 0.902)
Positive predictive value	0.162 (0.088 – 0.261)	0.188 (0.095 – 0.313)
Negative predictive value	0.975 (0.950 – 0.990)	0.969 (0.944 – 0.986)
Positive likelihood ratio	3.315 (2.175 – 5.051)	3.964 (2.322 – 6.768)
Negative likelihood ratio	0.439 (0.230 – 0.837)	0.543 (0.327 – 0.901)
AUC	0.771 (0.652-0.889)	0.802 (0.696-0.909)
AUC= area under the ROC curve		

We also looked at the sensitivity and specificity of a range of threshold values for the SarQoL® Overall score, which are displayed in table 3.

Table 3: Sensitivity and specificity for a range of threshold values for the SarQoL® Overall score				
Threshold value	Se	Sp	PPV	NPV
≤30 points	5.9%	100%	100%	94.8%
≤40 points	17.6%	95.9%	20.0%	95.2%
≤50 points	52.9%	85.6%	17.6%	96.9%
≤52.4 points (optimal threshold)	**64.7%**	**80.5%**	**16.2%**	**97.5%**
≤60 points	76.5%	65.8%	11.5%	98.0%
≤70 points	88.2%	40.1%	7.9%	98.3%
≤80 points	100%	21.2%	6.9%	100%
≤90 points	100%	7.9%	5.9%	100%
≤100 points	100%	NA	5.5%	NA
Se= sensitivity; Sp= specificity; PPV= positive predictive value; NPV= negative predictive value				

5. Discussion

This exploratory study showed that the SarQoL® questionnaire may be useful in screening potential candidates who are suspected of having sarcopenia for inclusion in clinical trials. The AUC of 0.771 (95% CI: 0.652–0.889) places it into the category of screening instruments with acceptable accuracy and confirms the primary study hypothesis. There might thus be a role for the SarQoL® questionnaire in a recruitment strategy of a clinical trial, certainly if it is already being considered to measure quality of life. We also found that the screening accuracy of the SarQoL® questionnaire in this sample was comparable to the SARC-F questionnaire but inferior to the Ishii screening test. The

SarQoL® questionnaire was able to correctly identify more sarcopenic participants than the SARC-F (64.7% vs. 52.9%), but at the cost of a slightly lower specificity (80.5% vs 86.6%). The Ishii screening test, which relies on physical parameters, correctly identified all 17 sarcopenic participants, giving it a sensitivity of 100%, but had the lowest specificity of all three tests at 76.7%.

That the Ishii screening test outperforms the SARC-F and Overall SarQoL® score should not be a great surprise. In fact, the items in the Ishii test closely resemble those that make up the diagnosis of sarcopenia according to the EWGSOP2 criteria, namely grip strength and calf circumference (as an indicator of muscle mass) [26]. The Ishii screening test has also shown, in a Polish study, that it possesses good screening accuracy when used to find sarcopenic people diagnosed with the EWGSOP2 criteria [29]. However, any comparison between the Ishii screening test and the SARC-F and Overall SarQoL® score needs to take into account that the Ishii screening test necessitates a face-to-face contact between the researcher and the potential candidate to obtain grip strength and calf circumference measurements, whereas the SARC-F and the SarQoL® questionnaire can be administered via the postal service, through the internet or via telephone.

The screening efficacy of the SARC-F, one of the most widely used tools and recommend by several organizations, has been investigated for multiple diagnostic criteria and summarized in a meta-analysis published in 2021. The authors found that the screening accuracy of the SARC-F was characterized by relatively low sensitivity (27–39%) combined with relatively high specificity (86–91%) when used in conjunction with the EWGSOP, Asian Working Group on Sarcopenia, International Working Group on Sarcopenia, and the Foundation for the National Institutes of Health Sarcopenia Project criteria. Interestingly, when they calculated the pooled sensitivity and specificity of the SARC-F based on the EWGSOP2 criteria, they found inverse results: moderate sensitivity (77%) and lower specificity (63%), although these results were only based on 4 studies. It is also important to mention that 3 of the 4 included studies focused on hospitalized patients, and that the pooled prevalence of sarcopenia was higher than in the general population at 21.56% [25]. We are aware of two other studies that are not included in this meta-analysis, namely Piotrowicz et al who reported a sensitivity of 35.3% and a specificity of 85.7%, and Nguyen et al, with a sensitivity of 64.9% and a specificity of 68.2%, both of which recruited community-dwelling older people [29, 30]. It has been argued that the SARC-F is better suited to ruling out sarcopenia rather than case-finding, which seems to be the case for the last two articles mentioned, but not so for the 4 included in the meta-analysis of Lu et al [31, 32].

In our study, the SarQoL® questionnaire performed similarly to the SARC-F questionnaire, with slightly greater sensitivity but slightly lower specificity. The SarQoL® questionnaire was able to correctly identify more sarcopenic patients in the sample, but the PPV of 16.2% was lower than the PPV of 18.8% of the SARC-F instrument. This means that 68 people would have been singled out for further investigation by the SarQoL® questionnaire, and 48 for the SARC-F, for two additional sarcopenic

subjects to be found. Therefore, in our example, the SarQoL® questionnaire would have been preferable if the recruitment strategy called for finding the greatest number of sarcopenic participants in the shortest amount of time, accepting the extra cost in performing complete body composition and/or physical performance assessments on more people. The SarQoL® questionnaire also has the advantage that it is self-administered and, therefore, requires fewer hours of study personnel time than the SARC-F, which is interviewer-administered.

The specific purpose for which a screening instrument is used can influence which of its characteristics to prioritize. In an ideal situation, a screening instrument would be inexpensive, easy to administer, without side effects, reliable, valid, and both highly sensitive and specific. Oftentimes, however, a trade-off needs to be made between these characteristics. Both the SARC-F and the SarQoL® questionnaire are inexpensive, easy to administer and without side effects given that they are questionnaires. The SarQoL® questionnaire has also demonstrated to be reliable in multiple studies [11]. However, both the SARC-F and the SarQoL® questionnaire are not highly sensitive nor highly specific, and are not as sensitive as the Ishii screening test. Nonetheless, if its limitations are taken into account, the SarQoL® questionnaire could be useful within certain contexts.

There are some limitations to take into account when interpreting the results of this study. First off, this study was a secondary analysis of data collected previously, and not specifically designed to answer the research question. This has led to certain issues around the reduction of risk of bias, such as the fact that the research assistant was not blinded to the results of the body composition analysis, grip strength measurement and SARC-F score. A second issue is the fact that, because no pre-specified cut-off exists, we determined the optimal threshold for the Overall QoL score of the SarQoL® questionnaire with the Youden index. This reflects the best balance between sensitivity and specificity, but may not necessarily be generalizable. The various studies performed with the SarQoL® questionnaire have already shown that absolute quality of life scores can significantly differ between countries. Normative population data or pilot studies will be needed to inform the appropriate threshold value in different situations. Lastly, because of the design of this study, we did not perform sample size calculations but provided confidence intervals around the main outcome values to provide a measure of precision. For both the SarQoL® and the SARC-F questionnaire, relatively large confidence intervals are observed around their point estimates, owing to the small number of people diagnosed with sarcopenia according to the EWGSOP2 criteria in this sample.

This study shows the feasibility of using the SarQoL® questionnaire as a tool to select those people who may benefit from a complete sarcopenia evaluation. While this study presents an interesting new use for the SarQoL® questionnaire, caution should be used in applying the threshold value used in this study ($\leq$ 52.4 points) to other populations.

6. Conclusion

In the population presented in this study, the SarQoL® Overall score, dichotomized at ≤52.4 points, performed roughly equal in terms of sensitivity and specificity to the SARC-F tool in identifying people considered sarcopenic with the EWGSOP2 criteria.

7. References

1. Cruz-Jentoft AJ, Bahat G, Bauer J, et al (2018) Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 16–31. https://doi.org/10.1093/ageing/afy169
2. Van Ancum JM, Alcazar J, Meskers CGM, et al (2020) Impact of using the updated EWGSOP2 definition in diagnosing sarcopenia: A clinical perspective. Arch Gerontol Geriatr 90:. https://doi.org/10.1016/j.archger.2020.104125
3. Locquet M, Beaudart C, Petermans J, et al (2019) EWGSOP2 Versus EWGSOP1: Impact on the Prevalence of Sarcopenia and Its Major Health Consequences. J Am Med Dir Assoc 20:384–385. https://doi.org/10.1016/j.jamda.2018.11.027
4. Mohd Nawi SN, Khow KS, Shiong Lim W, et al (2019) Screening Tools for Sarcopenia in Community-Dwellers: A Scoping Review Screening Tools for Sarcopenia-Siti N Mohd Nawi et al. Ann Acad Med Singapore 48:201
5. Locquet M, Beaudart C, Reginster JY, et al (2018) Comparison of the performance of five screening methods for sarcopenia. Clin Epidemiol 10:71–82. https://doi.org/10.2147/CLEP.S148638
6. Beaudart C, Biver E, Reginster JY, et al (2015) Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: The SarQoL. Age Ageing 44:960–966
7. Beaudart C, Biver E, Reginster J-Y, et al (2017) Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 8:. https://doi.org/10.1002/jcsm.12149
8. Beaudart C, Edwards M, Moss C, et al (2017) English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. Age Ageing 46:271–277. https://doi.org/10.1093/ageing/afw192
9. Fábrega R, Antonio C, Amat M, et al (2019) Psychometric Properties of the Spanish Version of the Sarcopenia and Quality of Life , a Quality of Life Questionnaire Specific for Sarcopenia. Calcif Tissue Int. https://doi.org/10.1007/s00223-019-00635-9
10. Geerinck A, Bruyère O, Locquet M, et al (2018) Evaluation of the Responsiveness of the SarQoL® Questionnaire, a Patient-Reported Outcome Measure Specific to Sarcopenia. Adv Ther 35:1842–1858. https://doi.org/10.1007/s12325-018-0820-z

11. Geerinck A, Alekna V, Beaudart C, et al (2019) Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (Sarqol) questionnaire: An analysis of subjects from 9 validation studies. PLoS One 14:e0216065. https://doi.org/10.1371/journal.pone.0216065

12. Erdogan T, Eris S, Avci S, et al (2021) Sarcopenia quality-of-life questionnaire (SarQoL)®: translation, cross-cultural adaptation and validation in Turkish. Aging Clin Exp Res. https://doi.org/10.1007/s40520-020-01780-0

13. Geerinck A, Scheppers A, Beaudart C, et al (2018) Translation and validation of the Dutch SarQoL ® , a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact 18:463–472

14. Tsekoura M, Billis E, Gliatis J, et al (2018) Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. Disabil Rehabil 1–7. https://doi.org/10.1080/09638288.2018.1514076

15. Konstantynowicz J, Abramowicz P, Glinkowski W, et al (2018) Polish Validation of the SarQoL®, a Quality of Life Questionnaire Specific to Sarcopenia. J Clin Med 7:323. https://doi.org/10.3390/jcm7100323

16. Gasparik AI, Mihai G, Beaudart C, et al (2018) Correction to: Psychometric performance of the Romanian version of the SarQoL(R), a health-related quality of life questionnaire for sarcopenia. Arch. Osteoporos. 13:98

17. Alekna V, Kilaite J, Tamulaitiene M, et al (2019) Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). Eur Geriatr Med

18. Safonova YA, Lesnyak OM, Baranova IA, et al (2019) Russian translation and validation of SarQoL® - quality of life questionnaire for patients with sarcopenia. Nauchno-Prakticheskaya Revmatol 57:38–45. https://doi.org/10.14412/1995-4484-2019-38-45

19. Mandrekar JN (2010) Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 5:1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

20. Beaudart C, Reginster JY, Petermans J, et al (2015) Quality of life and physical components linked to sarcopenia: The SarcoPhAge study. Exp Gerontol 69:103–110

21. Bossuyt PM, Reitsma JB, Bruns DE, et al (2015) STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. BMJ 351:1–9. https://doi.org/10.1136/bmj.h5527

22. Beaudart C, Biver E, Reginster JY, et al (2017) Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 8:238–244. https://doi.org/10.1002/jcsm.12149

23. Malmstrom TK, Miller DK, Simonsick EM, et al (2016) SARC-F: A symptom score to predict persons with sarcopenia at risk for poor functional outcomes. J Cachexia Sarcopenia Muscle 7:28–36. https://doi.org/10.1002/jcsm.12048

24. Ida S, Kaneko R, Murata K (2018) SARC-F for Screening of Sarcopenia Among Older Adults: A Meta-analysis of Screening Test Accuracy. J Am Med Dir Assoc 19:685–689. https://doi.org/10.1016/j.jamda.2018.04.001

25. Lu JL, Ding LY, Xu Q, et al (2021) Screening Accuracy of SARC-F for Sarcopenia in the Elderly: A Diagnostic Meta-Analysis. J Nutr Heal Aging 25:172–182. https://doi.org/10.1007/s12603-020-1471-8

26. Ishii S, Tanaka T, Shibasaki K et al (2014) Development of a simple screening test for sarcopenia in older adults. Geriatr Gerontol Int 14:93–101. https://doi.org/10.1111/ggi.12197

27. IBM Support (2018) Can SPSS Statistics produce epidemiological statistics from 2x2 tables such as positive and negative predictive values, sensitivity, specificity and likelihood ratios?

28. Youden WJ (1950) Index for rating diagnostic tests. Cancer 3:32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

29. Piotrowicz K, Głuszewska A, Czesak J, et al (2021) SARC-F as a case-finding tool for sarcopenia according to the EWGSOP2. National validation and comparison with other diagnostic standards. Aging Clin Exp Res. https://doi.org/10.1007/s40520-020-01782-y

30. Nguyen AT (2020) Reliability and Validity of SARC-F Questionnaire to Assess Sarcopenia Among Vietnamese Geriatric Patients. 879–886

31. Piotrowicz K, Gryglewska B, Gąsowski J (2021) The usefulness of SARC - F. Aging Clin Exp Res 40520. https://doi.org/10.1007/s40520-021-01839-6

32. Ünsal F, Murat M, Levent K (2021) SARC-F as a case-finding tool in sarcopenia: valid or unnecessary? Aging Clin Exp Res 3–4. https://doi.org/10.1007/s40520-021-01838-7

PART 2

Development and validation of a shorter version of the SarQoL questionnaire

Chapter 6: the short-form SarQoL

Development and validation of a short version of the Sarcopenia Quality of Life questionnaire: the SF-SarQoL

Published as:

Geerinck, A., Beaudart, C., Reginster, JY. *et al.* Development and validation of a short version of the Sarcopenia Quality of Life questionnaire: the SF-SarQoL. *Qual Life Res* (2021). https://doi.org/10.1007/s11136-021-02823-3

1. Abstract

Purpose : To facilitate the measurement of quality of life in sarcopenia, we set out to reduce the number of items in the previously validated Sarcopenia Quality of Life (SarQoL®) questionnaire, and to evaluate the clinimetric properties of this new short form.

Methods: The item reduction process was carried out in two phases. First, information was gathered through item-impact scores from older people (n = 1950), a Delphi method with sarcopenia experts, and previously published clinimetric data. In the second phase, this information was presented to an expert panel that decided which of the items to include in the short form. The newly created SFSarQoL was then administered to older, community-dwelling participants who previously participated in the SarcoPhAge study. We examined discriminative power, internal consistency, construct validity, test–retest reliability, structural validity and examined item parameters with a graded response model (IRT).

Results: The questionnaire was reduced from 55 to 14 items, a 75% reduction. A total of 214 older, community-dwelling people were recruited for the validation study. The clinimetric evaluation showed that the SF-SarQoL® can discriminate on sarcopenia status [EWGSOP2 criteria; 34.52 (18.59–43.45) vs. 42.86 (26.56–63.69); $p = 0.043$], is internally consistent ($\alpha = 0.915$, $\omega = 0.917$) and reliable [ICC = 0.912 (0.847–0.942)]. A unidimensional model was fitted (CFI = 0.978; TLI = 0.975; RMSEA = 0.108, 90% CI 0.094–0.123; SRMR = 0.055) with no misfitting items and good response category separation.

Conclusions : A new, 14-item, short form version of the Sarcopenia Quality of Life questionnaire has been developed and shows good clinimetric properties.

2. Background

The process of ageing is associated with numerous physiological changes. One of these changes is the age-related decrease in muscle mass and function known as sarcopenia, which has received a great deal of interest in the past decade [1, 2].

Sarcopenia is described by the European Working Group on Sarcopenia in Older People (EWGSOP) as "a progressive and generalized skeletal muscle disorder that is associated with increased likelihood of adverse outcomes including falls, fractures, physical disability and mortality" [3]. The most recent consensus criteria of the EWGSOP2 state that low muscle strength is an indicator of probable sarcopenia, low strength in combination with low muscle mass is confirmed sarcopenia, and low muscle strength, low muscle mass and low physical performance is severe sarcopenia [3].

Sarcopenia has been associated with increased mortality, functional decline, a higher rate of falls and a higher incidence of hospitalization [4, 5]. In the last few years, evidence has been accumulating on the adverse impact of sarcopenia on quality of life [6, 7].

In 2015, Beaudart and colleagues presented the Sarcopenia Quality of Life (SarQoL) questionnaire, an auto-administered patient-reported outcome measure specifically designed to measure quality of life in older, community-dwelling people [8]. It is still currently the only instrument measuring quality of life validated for sarcopenic samples and the only sarcopenia-specific QoL questionnaire available.

The clinimetric properties of the SarQoL questionnaire have been examined for 11 language-specific versions of the questionnaire and has demonstrated strong measurement properties [9,10,11,12,13,14,15,16,17,18,19,20]. The questionnaire has been extensively translated, and is available in 30 languages from the website www.sarqol.org.

The comprehensive nature of the SarQoL® questionnaire, which allows it to probe multiple facets of QoL in sarcopenia, means a trade-off has been made between its comprehensiveness and its response burden. Several factors may contribute to the perception of burden on the part of the respondent, such as the length of the questionnaire, the formatting, the instructions, the invasiveness of the questions and the cognitive load the questions put on the respondent [21]. While the developers estimated, based on the results of a pre-test in the target population, that it would take most patients about 10 min to complete the SarQoL®, in practice a considerable number of respondents need more time than this. Given that most clinical studies administer a number of tests and questionnaires, and thus need to take into consideration the response burden of each instrument so as not to jeopardize the accuracy of the obtained data and the percentage of missing responses, a shorter version of the SarQoL® questionnaire might prove valuable.

The first objective of this study was to extract a shorter version out of the 55 items of the SarQoL® questionnaire which safeguards the conceptual structure and the content validity of the original instrument. The second objective was to investigate the clinimetric properties of the newly developed short-form SarQoL.

3. Methods

3.1. Development phase

The SarQoL questionnaire

The short form described in this article was developed from the Sarcopenia Quality of Life (SarQoL) questionnaire. This auto-administered patient-reported outcome measure was developed with the specific aim of evaluating quality of life in sarcopenic, community-dwelling older people. The SarQoL measures QoL through 55 items categorized into seven domains of health-related dysfunction: physical and mental health, locomotion, body composition, functionality, activities of daily living, leisure activities, and fears [8]. The response options of the SarQoL questionnaire are a mix of Likert scales (3, 4, or 5 levels) and multiple-answer multiple-choice questions. The scoring algorithm calculates an overall QoL score which is scaled from 20 to 100 points (with complete data), and also provides seven domain scores, scaled from 0 (worst QoL possible) to 100 (best QoL possible) points. The scoring algorithm is not publicly available, but tools to calculate the scores are available by contacting info@sarqol.org. The clinimetric properties of the questionnaire have been evaluated in 11 different language-specific versions, and considerable information is available for known-groups validity, construct validity, internal consistency, floor and ceiling effects, test–retest reliability, standard error of measurement, smallest detectable change, and an evaluation of the responsiveness of the SarQoL has also been carried out [9,10,11,12,13,14,15,16,17,18,19,20]. Based on these results, the SarQoL is considered to be a valid, reliable and responsive instrument. The SarQoL questionnaire itself and additional information on the various publications are available from www.sarqol.org.

Item selection process

The objectives of the item reduction process were to create a significantly shorter version of the SarQoL questionnaire that would represent as much of the conceptual model of the Overall QoL score of the original questionnaire as possible, and thus also be highly correlated with the same score.

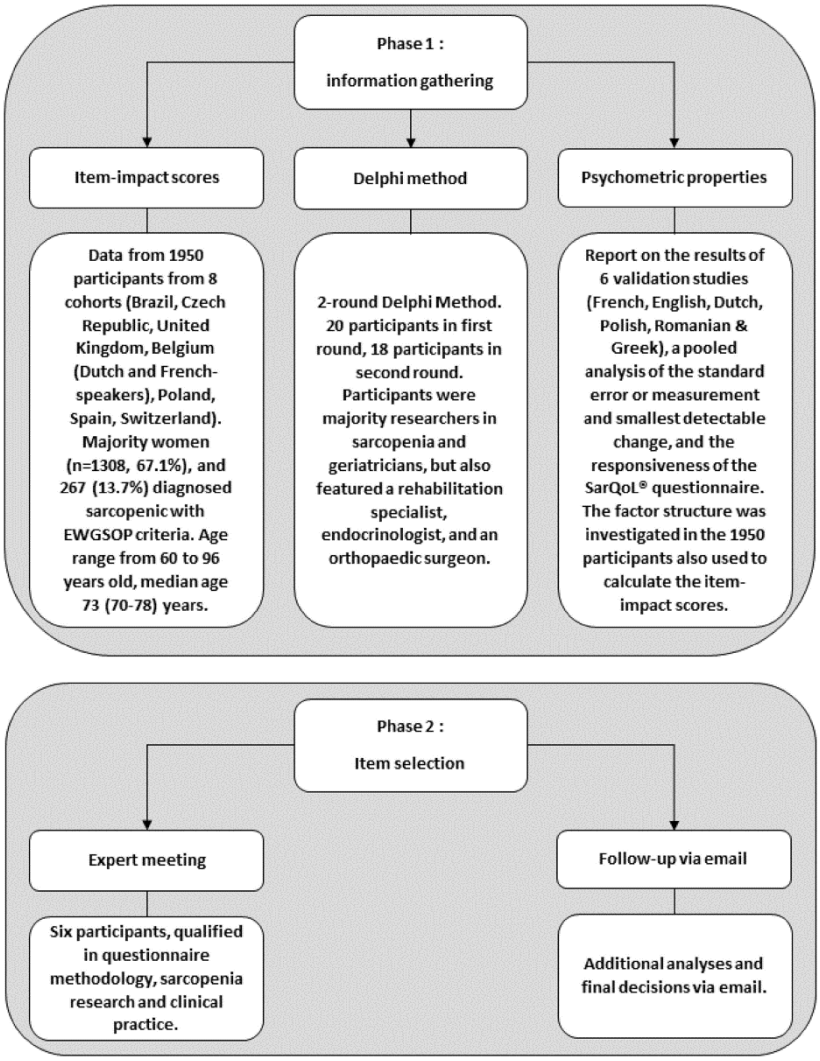
The item selection process was carried out in two phases, presented in Fig. 1. The first phase served to collect and collate as much information on the properties of the items and domains in the SarQoL questionnaire. This phase started off with the calculation of item-impact scores to determine which

items in the SarQoL questionnaire are the most relevant and impactful for sarcopenic people. For this purpose, we combined data collected in Brazil, the Czech Republic, the UK, Belgium (two separate cohorts), Poland, Spain and Switzerland. All data were collected in non-interventional studies (transversal and cohort) from community-dwelling older people (60 years and older) who were evaluated for sarcopenia according to the EWGSOP criteria [22]. In total, data from 1950 participants were included in this dataset, of which 267 were diagnosed as sarcopenic. By calculating the ratio of the number of participants experiencing an item to those who did not experience it, and dividing this by the mean impact, a ranking was established from most relevant and impactful to least [23]. The first phase of the item selection process continued with a 2-round modified Delphi method, so that the patient's perspective quantified by the item-impact scores could be complemented with the opinion of health care professionals and researchers. We targeted researchers and clinicians involved in sarcopenia research who had previous experience with the SarQoL questionnaire, through use, translation, validation or development, and invited them to participate. The participants were provided with an Excel file wherein they were able to categorize each of the 55 items as either "must absolutely be kept in a short form" or "could be discarded". Items were organized and presented per domain. In the second round, the participants were once again asked to categorize the items in the SarQoL questionnaire (keep or discard), but were now also provided the item-impact scores as well as the percentage of participants who agreed on whether to keep or discard an item in the first round. Consensus at the end of the second round was defined as 70% agreement. During both rounds, participants were able to add comments on their choices. The information from the Delphi method, the item-impact scores, and the already published information concerning the clinimetric properties of the SarQoL questionnaire was summarized into a report at the end of the first stage.

In the second phase of the item reduction process the report compiled at the end of phase one was presented to an expert group consisting of researchers specialized in sarcopenia and QoL, a clinical practitioner and a questionnaire methodologist (AG, CB, OB, ML, CM, SG). These discussed the available information and decided on the inclusion or exclusion of a number of items. As recommended in the guidelines formulated by Goetz et al., the expert group was asked to consider content validity (i.e. the results from the item-impact study and the Delphi method) as having the most weight in the decision-making process, followed by clinimetric properties and finally any additional analyses (factor analysis, correlations, or subgroup analyses) that were performed. To ensure an important reduction of the length of the questionnaire, an arbitrary goal of at least a 65% reduction was chosen at the start of the selection process, while maintaining the relative weight of the seven domains in the SarQoL questionnaire

.

Figure 1: Development process of the SF-SarQoL questionnaire



3.2. Validation phase

Population and study design

For the validation of the SF-SarQoL, we contacted the 314 participants who had previously participated in the fourth and/or fifth year of follow-up of the SarcoPhAge (Sarcopenia and Physical impairment with advancing Age) study [24]. In short, this study recruited older, community-dwelling volunteers from the Liège province of Belgium, and invited them once a year for a battery of physical and other measurements. Given that sarcopenia was the main focus of the SarcoPhAge study, body composition, muscle strength and physical performance were evaluated at each visit with dual-energy X-ray absorptiometry, a hydraulic hand-dynamometer and the Short Physical Performance Battery. Details on the SarcoPhAge study design and results have been reported previously [24, 25]

We provided the participants, through the postal service, with study packets composed of the short form SarQoL questionnaire, the EQ-5D and EQ-VAS questionnaire which are preference-based measures of

health status, and the original SarQoL questionnaire. The study packets were accompanied by an explanatory letter and a pre-stamped envelope with which to return the study documents [26]. The people who consented to participate and sent back the completed questionnaires received a second packet by mail about 10 days after the date on which they completed the first packet. The second study packet consisted of the SF-SarQoL and a query on whether their health had changed in the interval between the two administrations of the SF-SarQoL. Demographic and clinical data were obtained from the existing datasets collected during the fourth or fifth year follow-up visits of the SarcoPhAge study. Sarcopenia was diagnosed with the revised consensus criteria from the EWGSOP2 (handgrip strength below 27 kg for men or 16 kg for women, together with low muscle mass defined as appendicular skeletal muscle mass divided by height-squared (ASM/Ht^2) < 7.0 kg/m^2 for men or < 5.5 kg/m^2 for women) [3]. The research protocol (no 2012/277) and its amendment (dated 19/12/2019) were approved by the Ethics Committee of the University Teaching Hospital of Liège.

Clinimetric properties from classical test theory

The clinimetric properties of the SF-SarQoL have been examined with the following indicators from classical test theory:

1) Item characteristics have been evaluated with percentage of missing responses. Floor and ceiling effects for the overall QoL score of the SF-SarQoL were considered to be present if more than 15% of respondents obtained the lowest (0 points) or highest (100 points) score [27].
2) Discriminative power (also known as known-groups validity), which measures an instrument's ability to distinguish among distinct groups, has been examined in three separate comparisons: sarcopenic versus non-sarcopenic, probably sarcopenic (low grip strength in the EWGSOP2 algorithm) versus probably non-sarcopenic (normal grip strength), and at high risk of sarcopenia (SARC-F score ≥ 4) versus at low risk of sarcopenia [3, 28]. We expected to find significantly lower QoL scores on the SF-SarQoL for sarcopenic participants, those with low grip strength and those at high risk of sarcopenia. Significant differences in QoL were established with the Student *t* test or the Mann–Whitney *U* test, depending on normality of distribution of the scores. Point biserial correlation coefficients (*r*) were calculated to provide a measure of the strength of association between group status and QoL.
3) Internal consistency was measured with both the Cronbach's alpha value and the McDonald omega value. We decided on this approach because the alpha value allows comparison to previous validation studies, while the omega value avoids some of the problems associated with the alpha value and is considered to be a more accurate reflection of internal consistency [29]. For both indicators, values between 0.7 and 0.95 indicate that the items in the questionnaire are closely interrelated and measure the same concept [27].

4) Test–retest reliability has been quantified with the intraclass correlation coefficient (ICC—two-way mixed model and absolute agreement type) for the total score of the SF-SarQoL, and with weighted kappa coefficients (using quadratic weights) for the individual items. An ICC value greater than 0.7 indicates acceptable reliability [27]. For the weighted kappa coefficients, a value ≥ 0.8 is almost-perfect agreement, ≥ 0.6 and < 0.8 is substantial agreement, ≥ 0.4 and < 0.6 is moderate agreement, ≥ 0.2 and < 0.4 is fair agreement and < 0.2 is slight agreement [30]. Only those participants who participated in both administrations of the SF-SarQoL, whose health did not change in the interval period, and who completed the second questionnaire a maximum of 3 weeks after the first, were eligible for inclusion in this analysis. A Bland–Altman analysis was also carried out to detect whether there was systematic bias in the test–retest data [31].
5) The construct validity of the SF-SarQoL has been investigated through three approaches. First, we evaluated criterion validity, where the instrument scores are compared to those of a gold standard. This was measured with the ICC (two-way mixed model and consistency type) between the overall QoL scores of the short form and the original SarQoL questionnaire [27]. Secondly, we tested hypotheses on the expected correlation between the SF-SarQoL and the EQ-5D and EQ-VAS questionnaires, assuming that we will find strong correlations between them [27]. Lastly, we evaluated the structural validity of the SF-SarQoL. We hypothesized that the SF-SarQoL is unidimensional, with all items loading on the latent construct of quality of life, and have carried out a confirmatory factor analysis using the diagonally weighted least squares estimator (WSLMV) for ordinal data using the R package "Lavaan" (version 0.6–6). Model fit was evaluated with the Chi-square test ($p \geq 0.05$ indicates good fit), the comparative fit index (CFI; good fit if ≥ 0.95), the Tucker-Lewis index (TLI; good fit if ≥ 0.95), the root mean square error of approximation (RMSEA; good fit if ≤ 0.08) and the standardized root mean square residual (SRMR; good fit if ≤ 0.08) [32, 33].

Clinimetric properties from modern measurement theory

Before constructing and testing an IRT model, it is important to verify that the items meet the assumptions of unidimensionality, local independence and monotonicity [34].

1. Most IRT applications require a factor structure with a single latent trait, hence the need to establish whether the instrument in question is unidimensional. This was established using the results of the CFA described in the previous paragraph, supplemented with an exploratory factor analysis. Before launching the EFA, we inspected the suitability of the data using Bartlett's test of sphericity and the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy. The EFA was executed on the polychoric correlation matrix with the WLSMV estimator from the R package "Psych" (version

1.9.12.31). The number of factors present was evaluated with parallel analysis (PA) and Velicer's minimum average partial (MAP) test [35].

2. The second assumption, local independence, means that there should be no correlation between two items after the effect of the underlying trait is filtered out. In other words, the item responses should be entirely a function of the underlying trait, and not (partly) dependent on a second factor [34]. To determine this, we looked at the residual correlation matrix from the previously described single-factor CFA, and considered a value of 0.2 above the average residual correlation as the cut-off for local independence [36].
3. Lastly, the concept of monotonicity was examined. This concept states that the probability of endorsing a higher item response category should increase with increasing levels of the underlying construct [34]. Monotonicity was evaluated with Mokken scaling carried out with the R package "Mokken" (version 3.0.2), using the scalability coefficient *H* for each item and the questionnaire in its entirety. The assumption of monotonicity was confirmed if the item scalability coefficients were ≥ 0.3 and the scalability coefficient H_i for the entire questionnaire was ≥ 0.5 [36].

After confirming unidimensionality, local independence and monotonicity, a logistic Graded Response Model (GRM) was fit to the data using the R package "mirt" (version 1.32.1). This model calculates both item thresholds (b) as well as item slopes (a). For the purpose of this analysis, the response options "I do not undertake these types of physical activities" in item 2.1 and 2.2, "not applicable" in item 3.1 and 3.2, "I am unable to walk" in item 4, and "I have never participated in leisure activities" in item 8 were treated as missing responses. The encoding of the responses on item 8 was also re-ordered, going from decreased participation to increased participation. Item fit was examined with the $S - X^2$ indicator, where $p \leq 0.001$ indicates poor fit, and by examining the category characteristic curves. For all items, 3 thresholds were estimated, except for item 8, where only two thresholds were estimated.

3.3. Statistical analysis

All analyses were executed with SPSS version 27.0.0, R version 4.0.0. and JASP version 0.13.1.

In addition to the statistical manipulations described in the preceding paragraphs, we also verified normality of distribution for quantitative variables with the Shapiro–Wilk test, by comparing mean and median, and by evaluating the histogram and Q–Q plot. Continuous variables following a Gaussian distribution are reported as mean ± standard deviation, while skewed variables are reported as median (25th percentile–75th percentile). Nominal variables are reported as absolute (*n*) and relative (%) frequencies. All comparisons were considered significant at the 5% level ($p \leq 0.5$).

4. Results

4.1. Development

Twenty experts participated in the first round of the modified Delphi method, and eighteen of them participated in both rounds. The panel reached consensus on the inclusion of 13 items and the exclusion of 23 items, with 19 items not reaching the 70% agreement threshold for either option. Together with the item-impact scores, calculated separately for the sarcopenic ($n = 267$) and non-sarcopenic ($n = 1584$) participants, and the clinimetric information already available from previous validation studies, these allowed the expert panel to reach a final decision on the inclusion of 14 items from six domains (physical and mental health, locomotion, body composition, functionality, activities of daily living, and leisure activities), which together constitute the short-form SarQoL questionnaire. The expert panel made the decision to deviate from the original conceptual model by not including an item from domain seven (fears) because the format of the question (items are conditional upon the previous question) and the response options (only a positive answer is identified, a negative response or missing data cannot be separated) rendered item-level analysis problematic. The summarized results from the Delphi method, the item-impact ranking and the final decisions of the expert panel are shown in Table 1. The SF-SarQoL is available at www.sarqol.org.

Table 1: development SF-SarQoL

Domain/item		Delphi method [a]		Item-impact ranking [bc]		Final decision
		Consensus inclusion	Consensus exclusion	Sarcopenic group	Non-sarcopenic group	
Physical & mental health						
1.1	Loss of arm strength	x		3	6	IN
1.2	Loss of leg strength	x		1	4	IN
1.4	Loss of energy		x	4	5	
2	Muscle pain		x	2	3	
6	Feeling old		x	6	2	
7	Feeling of muscle weakness		x			
8	Feeling of being physically weak			5	1	IN
16	Feeling of being frail			7	7	
Locomotion						
9.1	Limitation in walking time	x		4	4	
9.2	Limitation in number of outings		x	6	6	
9.3	Limitation in walking distance	x		2	2	
9.4	Limitation in walking speed	x		1	1	IN
9.5	Limitation in steps length		x	7	7	
10.1	Feeling of fatigue when walking	x		3	3	IN
10.2	Need of recovery time when walking			7	8	
10.3	Difficulties to cross a road fast enough			9	9	
10.4	Difficulties to walk on uneven ground		x	4	5	
Body composition						
1.3	Loss of muscle mass			2	2	IN
13	Physical change		x	1	1	
14	Weight change (loss or gain)		x			
15	Upset with change		x			
Functionality						
1.5	Loss of physical capacity	x		2	2	IN
1.6	Loss of flexibility		x	3	1	
11	Balance problems			5	4	IN

Table 1: development SF-SarQoL

Domain/item		Delphi method [a]		Item-impact ranking [bc]		Final decision
		Consensus inclusion	Consensus exclusion	Sarcopenic group	Non-sarcopenic group	
12	Falls occurrence	x		13	8	
17.1	Climbing one flight of stairs	x		11	13	
17.2	Climbing several flights of stairs		x	6	6	
17.3	Climbing stairs without a banister			8	11	
17.4	Crouching or kneeling			4	5	
17.5	Stooping			10	10	
17.6	To stand up from the floor without any support			1	3	IN
17.7	Get up from a chair	x		9	7	
17.8	To stand from a sitting position			12	12	
18	Limitation of movement	x		7	9	IN
20	Sexuality		x	14	14	
Activities of daily living						
17.11	Take public transportation		x	14	14	
17.12	To get in/out a car		x	12	12	
3.1	Difficulty during light physical effort			5	9	
3.2	Fatigue during light physical effort	x		2	7	
3.3	Pain during light physical effort		x	4	8	
4.1	Difficulty during moderate physical effort			6	6	IN
4.2	Fatigue during moderate physical effort	x		3	3	IN
4.3	Pain during moderate physical effort		x	7	5	
5.1	Difficulty during intense physical effort		x	10	2	
5.2	Fatigue during intense physical effort		x	9	1	
5.3	Pain during intense physical effort		x	11	4	
17.9	Carrying heavy objects		x	1	10	IN
17.10	Open a bottle or a jar			8	11	
17.13	Shopping		x	15	15	
17.14	Household tasks			13	13	
Leisure activities						
21	Change in physical activities		x	1	1	

Table 1: development SF-SarQoL

Domain/item		Delphi method [a]		Item-impact ranking [bc]		Final decision
		Consensus inclusion	Consensus exclusion	Sarcopenic group	Non-sarcopenic group	
22	Change in leisure activities			2	2	IN
Fears						
19	Fear of getting hurt Fear of not succeeding Fear of being tired Fear of falling					

[a] Empty cells indicate that the 70% agreement threshold was not reached
[b] Because certain questions in the SarQoL questionnaire are conditional on other questions (i.e. "If yes on previous question, then …"), item-impact scores could not be calculated for items 7, 14, 15 and 19.
[c] Items are ranked from most impactful (1) to least impactful

4.2. Clinimetric evaluation

Participants

A total of 214 older people participated in the validation study for the SF-SarQoL. The median age of the participants was 76 (73–81) years and 63.1% were women. We found 70 (32.7%) participants with probable sarcopenia (low grip strength in the EWGSOP2 algorithm), of whom 21 (9.8%) had confirmed sarcopenia. With the help of the SARC-F questionnaire, we found 30 (14.0%) participants at high risk of sarcopenia. The complete clinical and QoL characteristics are reported in Table 2.

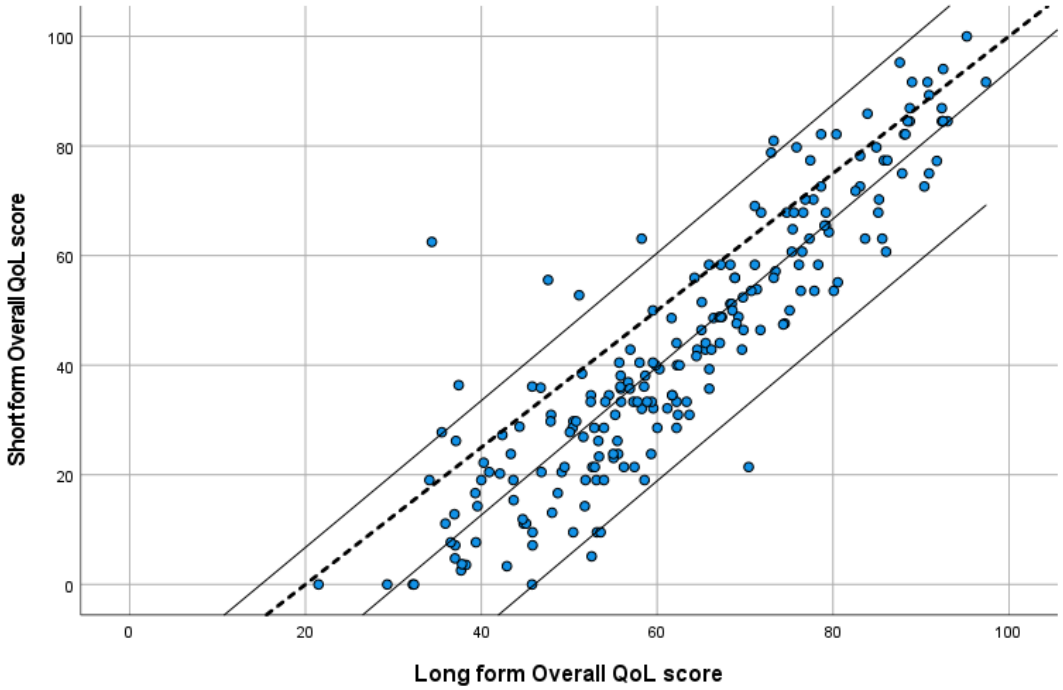
Table 2: characteristics of the sample

	N (%)	Median (P25- P75)
Gender		
Male	80 (36.9%)	
Female	137 (63.1%)	
Age (years)		76 (73 – 81)
Probable sarcopenia (with EWGSOP2)		
Yes	70 (32.7%)	
No	143 (66.8%)	
Sarcopenia (with EWGSOP2)		
Yes	21 (9.8%)	
No	193 (90.2%)	
At risk of sarcopenia (with SARC-F)		
Yes	30 (14.0%)	
No	184 (86.0%)	
EQ-5D index score		0.800 (0.747 – 0.827)
EQ-VAS		70 (60 – 80)
SarQoL – overall QoL score		
Physical and mental health		60.54 (48.87 – 73.30)
Locomotion		55.56 (41.67 – 75.70)
Body composition		62.50 (48.96 – 70.83)
Functionality		66.69 (55.36 – 82.28)
Activities of daily living		60.00 (48.21 – 76.67)
Leisure activities		33.25 (33.25 – 66.50)
Fears		87.50 (75.00 – 100.00)
Overall		61.97 (51.57 – 75.64)
SF-SarQoL overall QoL score first administration		40.24 (23.81 – 62.64)
SF-SarQoL overall QoL score second administration		47.62 (31.55 – 70.24)

Relationship between short and long form scoring algorithm

To ease interpretation of the QoL scores of the short form questionnaire, it was decided to use a scale going from zero to 100, a deviation from the 20–100 scale of the long form questionnaire. Within the scale, lower scores represent persons whose quality of life is significantly impacted by sarcopenia, and higher scores indicate people with better QoL and a smaller impact of sarcopenia. Figure 2 shows the scatter plot of the short and long form Overall QoL score. From this figure, it can be observed that the short form scores Overall QoL scores are roughly parallel but below the dotted equivalence line, which represents perfect correspondence between the 2 scores.

Figure 2: Relation between short form scores and the long form Overall QoL scores.



Legend: The dotted line provides a reference for equivalence between short and long form scores (from 0:20 to 100:100). A linear fit line with 95% prediction interval is also provided (R^2=0.816).

Clinimetric properties classical test theory

The per-item percentage of missing responses ranged between 0 and 5.6%. Five (2.3%) participants scored zero points on the Overall QoL score of the SF-SarQoL, and 1 (0.5%) person scored 100 points, indicating that there are no floor or ceiling effects in this sample. We found excellent discriminative power when comparing probably sarcopenic versus probably not [32.74 (20.15–43.15) vs. 48.81 (28.57–70.24); $p < 0.001$; $r = -0.342$], sarcopenic versus not sarcopenic [34.52 (18.59–43.45) vs. 42.86 (26.56–63.69); $p = 0.043$; $r = -0.144$] and at high risk of sarcopenia versus low risk [17.86 (6.64–24.05) vs. 46.43 (30.95–65.48); $p < 0.001$; $r = -0.444$]. Internal consistency among the items was excellent with a Cronbach's alpha of 0.915 (95% CI = 0.896–0.930) and a McDonalds' omega value of 0.917 (95% CI = 0.897–0.933). Test–retest reliability was calculated among 133 participants. Within this sub-sample, we found excellent test–retest reliability with an ICC of 0.912 (95% CI = 0.847–0.942) for the overall QoL score of the SF-SarQoL. On an item level, we found moderate to almost-perfect agreement between the first and second administration with weighted kappa coefficients, detailed in Table 3.

Table 3: Test-retest reliability and construct validity						
		Concordance of items between test and retest (n=133)		Standardized factor loadings		
		Weighted kappa (95% CI)	Interpretation [a]	Model 1	Model 2 [b]	
					Factor 1	Factor 2
1.1	Reduction strength arms	0.794 (0.658 – 0.840)	Substantial	0.695	0.725	
1.2	Reduction strength legs	0.735 (0.637 – 0.834)	Substantial	0.897	0.930	
1.3	Reduction muscle mass	0.682 (0.590 – 0.773)	Substantial	0.806	0.827	
1.4	Reduction physical capacity	0.613 (0.495 – 0.732)	Substantial	0.917	0.951	
1.5	Reduction length of walks	0.750 (0.673 – 0.828)	Substantial	0.867	0.873	
2.1	Difficulty moderate effort	0.691 (0.541 – 0.842)	Substantial	0.901		0.915
2.2	Tired moderate effort	0.646 (0.485 – 0.808)	Substantial	0.856		0.864
3.1	Get up from floor	0.683 (0.512 – 0.854)	Substantial	0.786		0.802
3.2	Carrying heavy objects	0.546 (0.335 – 0.756)	Moderate	0.821		0.833
4	Tired when walking	0.798 (0.732 – 0.865)	Substantial	0.874		0.866
5	Feel weak	0.791 (0.709 – 0.873)	Substantial	0.877		0.900
6	Balance problems	0.867 (0.812 – 0.921)	Almost perfect	0.673		0.689
7	Limit movements	0.728 (0.637 – 0.819)	Substantial	0.850		0.868
8	Leisure activities	0.406 (0.185 – 0.627)	Moderate	0.594		0.605
[a] Kappas interpreted according to Landis & Koch, where ≥0.8 is almost perfect agreement, ≥0.6 and <0.8 is substantial agreement, ≥0.4 and <0.6 is moderate agreement, ≥0.2 and <0.4 is fair agreement, and <0.2 is slight agreement. [b] Model 2 is a 2-factor model with correlated residual variance between items 1.5 and 4						

A Bland–Altman analysis revealed the presence of a systematic bias of 4.11 (95% CI 2.51; 5.72) points, with higher average scores for the retest scores (50.47 ± 24.82) compared to the test scores (46.36 ± 23.30).

The criterion construct validity, measuring the strength of relationship between the SarQoL overall QoL score and its short form equivalent, was excellent with an ICC of 0.835 (95% CI = 0.789–0.871). It should be noted that the scoring algorithm for the short form and the original SarQoL questionnaire are not on the same metric, and are thus not interchangeable. We also found strong correlations between the SF-SarQoL overall score and the EQ-5D index score ($r = 0.671$; $p < 0.001$) and the EQ-VAS

($r = 0.697$; $p < 0.001$). A confirmatory factor analysis of a one-dimensional model resulted in the following fit indices ($\chi^2 = 269.330$, df = 77, $p < 0.001$; CFI = 0.978; TLI = 0.975; RMSEA = 0.108, 90% CI = 0.094–0.123; SRMR = 0.055). As the five items of question 1 share a common stem, we hypothesized that they would be highly correlated, with would lead to a deterioration of fit indices. To overcome this issue, an alternative model was tested, with the five items of question 1 loading on a first latent variable, and the remaining questions on a second latent variable (factor 1: items 1.1 to 1.5; factor 2: items 2.1 to 8) and a correlated residual variance between items 1.5 and 4. This model obtained adequate fit indices: ($\chi^2 = 161.847$, df = 75, $p < 0.001$; CFI = 0.990; TLI = 0.988; RMSEA = 0.074, 90% CI = 0.058–0.089; SRMR = 0.042). The 2 latent variables in this model are highly correlated at $r = 0.894$. Standardized factor loadings for both models are reported in Table 3.

Table 4: Graded Response Model

Item		Mono-tonicity	Model fit	Item parameters			
		H_i	p-value S-X^2 [a]	a	b_1	b_2	b_3
1.1	Reduction strength arms	0.526	0.061	1.691	-1.519	0.277	1.579
1.2	Reduction strength legs	0.681	0.407	3.515	-0.618	0.314	1.388
1.3	Reduction muscle mass	0.590	0.460	2.278	-1.140	0.292	1.499
1.4	Reduction physical capacity	0.716	0.365	3.791	-1.012	0.494	1.594
1.5	Reduction length of walks	0.653	0.204	2.940	-0.543	0.415	1.461
2.1	Difficulty moderate effort	0.695	0.176	3.592	-0.478	0.361	1.262
2.2	Tired moderate effort	0.651	0.072	2.790	-0.416	0.618	1.756
3.1	Get up from floor	0.591	0.001	2.219	-1.002	0.193	1.378
3.2	Carrying heavy objects	0.653	0.497	2.544	-1.889	-0.314	1.012
4	Tired when walking	0.645	0.068	3.176	-0.673	0.247	1.425
5	Feel weak	0.687	0.476	3.386	-0.967	0.234	1.210
6	Balance problems	0.517	0.632	1.581	-1.362	0.022	1.298
7	Limit movements	0.697	0.269	2.954	-1.335	-0.110	0.709
8	Leisure activities	0.557	0.435	1.518	0.017	3.229	NA

[a] S-X^2 statistic calculated on 160 complete observations

Clinimetric properties modern measurement theory

Confirmatory factor analysis did not conclusively indicate that the SF-SarQoL is unidimensional. Therefore, we investigated further with an exploratory factor analysis, which was considered appropriate when the Bartlett's test returned a p value < 0.001 and the KMO test a value of 0.87. Parallel analysis identified a single factor in the data, as did the Velicer's MAP test, which achieved a minimum of 0.05 with 1 factor. There were no locally dependent items found, with no residual correlations greater than the cut-off of 0.184 or − 0.216 (average residual correlation = − 0.016). The monotonicity assumption was confirmed when scalability coefficients H_i between 0.517 ("balance problems") and 0.716 ("reduction physical capacity") were found, alongside a Mokken scalability coefficient H for the entire short form of 0.635.

After fitting the logistic Graded Response Model to the data, we found no misfitting items, as evidenced by the fact that no p values for the $S - X^2$ indicator were smaller than 0.001. The item with the lowest discriminative ability was found to be "leisure activities" ($a = 1.518$) and the most discriminative item was "reduction of physical capacity" ($a = 3.791$). The item thresholds were spread out from − 1.889 ("Carrying heavy objects") to 1.756 ("Tired moderate effort"). Detailed results on the model fit and item parameters are reported in Table 4. The category characteristics curves, a visual representation of the item parameters, are shown in Fig. 3.

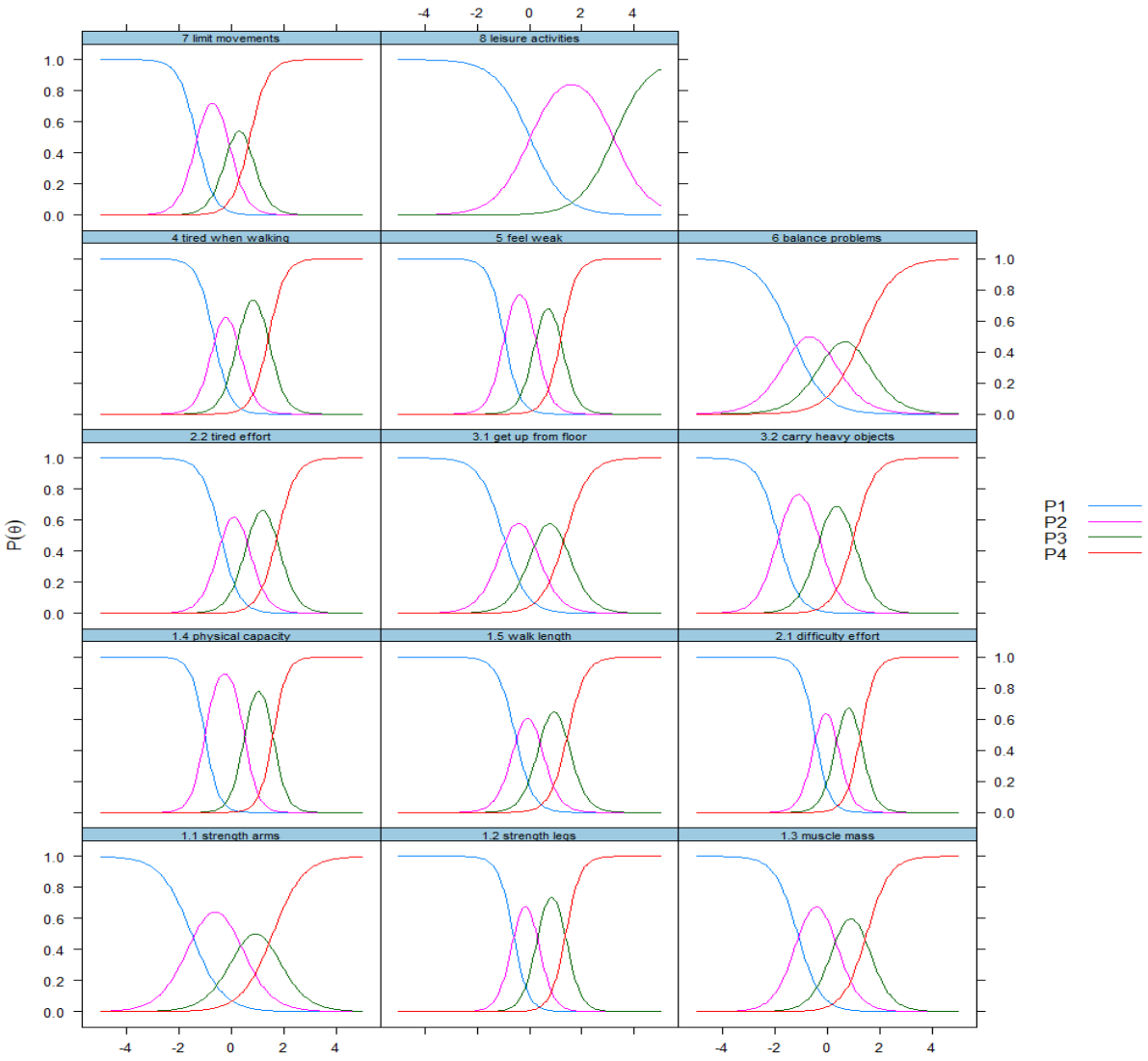


Figure 3: Category characteristic curves of the 14 items analysed in the SF-SarQoL

5. Discussion

This article describes the development of a 14-item short form version of the SarQoL® questionnaire, and the subsequent examination of its clinimetric properties.

The item reduction process follows the guidelines formulated by Goetz et al. by, among other things, prioritizing content validity over statistical properties [37]. The 2-phase process employed led to the inclusion of 14 items from six domains, preserving, as much as possible, the conceptual structure of the original SarQoL® questionnaire in the short form. One domain (D7: fears) did not contribute to the short form because, in the original questionnaire, it is dependent on the response of a different item that is not a part of domain seven. This type of conditional question ("If yes to the previous question, then …) combined with the fact that the response options for the items in question 19 make it impossible to distinguish between missing data and negative responses, made it inopportune in the eyes of the expert committee to include this domain. On top of the problems caused by its phrasing and response option, the participants in the Delphi method did not reach consensus on its inclusion, so these items and domain was not included in the short form. The questionnaire was thus reduced from 55 to 14 items, a 75% reduction.

In contrast with the original questionnaire, the newly created SF-SarQoL does not provide domain scores, but only an Overall QoL score. This is a conscious choice because, in our estimation, the original SarQoL® questionnaire is better suited when researchers wish to look at QoL on a domain-level. The SF-SarQoL is better suited to studies that use QoL as a secondary outcome, or in association with a general QoL instrument, and, in this vein, it privileges a single QoL score.

The validation part of this study found good to excellent results for discriminative power, construct validity, internal consistency, test–retest reliability and an absence of floor and ceiling effects. However, despite an ICC of 0.912 (95% CI = 0.847–0.942) for the test–retest reliability, we did find a systematic bias of 4.11 (95% CI = 2.51; 5.72) points. An earlier analysis of the original SarQoL® questionnaire in a sample of 274 sarcopenic participants demonstrated no such bias [0.18 (− 0.26; 0.63) points], so this result was unexpected [11]. It is unclear how this bias originated and whether it is a feature of the questionnaire or a one-off event, specific to this sample. It is possible that the higher QoL scores recorded during the second administration of the SF-SarQoL may be due to the packet length (19 pages for the first packet versus 6 pages for the second packet), or due to the information on sarcopenia received with the first packet, and which was absent in the second packet. Future validation studies should prioritize investigating test–retest reliability and, hopefully, clarify this issue. Confirmatory factor analysis did not conclusively confirm the unidimensional nature of the SF-SarQoL, with a 2-factor model showing better fit than the unidimensional model. The graded response model did not

indicate any misfitting items. The item trace lines show good separation between the different response categories.

Overall, the SF-SarQoL displays adequate to good clinimetric properties, allowing its use in research, clinical trials and clinical practice. Potential users should consider the objectives of their research when choosing between the 55-item or the 14-item SarQoL® questionnaire. If QoL is a primary outcome, the original SarQoL® questionnaire provides a superior level of detail and precision, as well as scores for the seven QoL domains on top of the overall QoL score. However, if QoL is not the main objective, and response burden is a serious consideration, the SF-SarQoL could be the more appropriate tool.

An important remark to make is that the scores on the original SarQoL® questionnaire and the newly developed SF-SarQoL are not interchangeable and should not be compared head-to-head. During the discussions on the scoring algorithm to be created for the short form SarQoL questionnaire, we examined the complexities of the original scoring algorithm, and a choice was made to place the SF-SarQoL on a 0–100 scale where the score range for the original SarQoL® questionnaire is about 20–100 points.

This study has several strengths: we followed the guidelines by Goetz et al., prioritized content validity, administered the SF-SarQoL in an independent sample and performed as complete a validation as possible with elements from both classical test theory and modern measurement theory.

However, this study also has some limitations: we did not perform differential item functioning analysis because of concerns about the sample size. We fully intend to rectify this once we are able to assemble sufficient data, preferably from multiple countries. We were unable to integrate the domain “fears” into the short-form, so a certain amount of content was lost during the item reduction process. Our sample size of 214 participants is sufficient for the performed statistical manipulations, but does not permit subgroup analyses. The members of the Delphi panel were selected for their previous knowledge of the SarQoL® questionnaire, and were not necessarily representative of the wider community of sarcopenia researchers and geriatricians. Due to the transversal nature of the performed validation study, we were unable to examine the responsiveness of the new SF-SarQoL. Evaluating this property of the SF-SarQoL should be a priority for future research.

In conclusion, this article presented the development process and the validation of a 14-item short form version of the SarQoL® questionnaire. In an independent sample, the SF-SarQoL demonstrated adequate measurement properties to allow its use. While its responsiveness should still be investigated, we fully recommend its use in situations where the original 55-item SarQoL® questionnaire is deemed to be too much of a burden on the respondents.

6. References

1. Peña Ordóñez, G. G., Bustamante Montes, L. P., Ramírez Duran, N., Sánchez Castellano, C., & Cruz-Jentoft, A. J. (2017). Populations and outcome measures used in ongoing research in sarcopenia. Aging Clinical and Experimental Research, 29, 695–700.
2. Reginster, J. Y., Beaudart, C., Al-Daghri, N., Avouac, B., Bauer, J., Bere, N., et al. (2020). Update on the ESCEO recommendation for the conduct of clinical trials for drugs aiming at the treatment of sarcopenia in older adults. Aging Clinical and Experimental Research. https://doi.org/10.1007/s40520-020-01663-4.
3. Cruz-Jentoft, A. J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., et al. (2018). Sarcopenia: Revised European consensus on definition and diagnosis. Age and Ageing, 48, 16–31.
4. Beaudart, C., Zaaria, M., Pasleau, F., Reginster, J.-Y., Bruyère, O., & Stenroth, L. (2017). Health outcomes of sarcopenia: A systematic review and meta-analysis. PLoS ONE, 12, e0169548.
5. Zhao, Y., Zhang, Y., Hao, Q., Ge, M., & Dong, B. (2019). Sarcopenia and hospital-related outcomes in the old people: A systematic review and meta-analysis. Aging Clinical and Experimental Research, 31, 5–14.
6. Woo, T., Yu, S., & Visvanathan, R. (2016). Systematic literature review on the relationship between biomarkers of sarcopenia and quality of life in older people. The Journal of Frailty & Aging, 5, 88–99.
7. Tsekoura, M., Kastrinis, A., Katsoulaki, M., Billis, E., & Gliatis, J. (2017). Sarcopenia and its impact on quality of life. Advances in Experimental Medicine and Biology, 987, 213–218.
8. Beaudart, C., Biver, E., Reginster, J. Y., Rizzoli, R., Rolland, Y., Bautmans, I., et al. (2015). Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: The SarQoL. Age and Ageing, 44, 960–966.
9. Beaudart, C., Biver, E., Reginster, J.-Y., Rizzoli, R., Rolland, Y., Bautmans, I., et al. (2017). Validation of the SarQoL, a specific health-related quality of life questionnaire for Sarcopenia. Journal of Cachexia, Sarcopenia and Muscle, 8, 238–244.
10. Beaudart, C., Edwards, M., Moss, C., Reginster, J. Y., Moon, R., Parsons, C., et al. (2017). English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. Age and Ageing, 46, 271–277.

11. Geerinck, A., Alekna, V., Beaudart, C., Bautmans, I., Cooper, C., De Souza, O. F., et al. (2019). Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: An analysis of subjects from 9 validation studies. PLoS ONE, 14, e0216065.

12. Erdogan, T., Eris, S., Avci, S., Oren, M. M., Kucukdagli, P., Kilic, C., et al. (2021). Sarcopenia quality-of-life questionnaire (SarQoL)®: Translation, cross-cultural adaptation and validation in Turkish. Aging Clinical and Experimental Research. https://doi.org/10.1007/s40520-020-01780-0.

13. Geerinck, A., Scheppers, A., Beaudart, C., Bruyère, O., Vandenbussche, W., Bautmans, R., et al. (2018). Translation and validation of the Dutch SarQoL ®, a quality of life questionnaire specific to sarcopenia. Journal of Musculoskeletal and Neuronal Interactions, 18, 463–472.

14. Tsekoura, M., Billis, E., Gliatis, J., Tsepis, E., Matzaroglou, C., Sakkas, G. K., et al. (2018). Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. Disability and Rehabilitation, 42, 1–7.

15. Konstantynowicz, J., Abramowicz, P., Glinkowski, W., Taranta, E., Marcinowicz, L., Dymitrowicz, M., et al. (2018). Polish validation of the SarQoL®, a Quality of Life Questionnaire specific to sarcopenia. Journal of Clinical Medicine, 7, 323.

16. Gasparik, A. I., Mihai, G., Beaudart, C., Bruyere, O., Pop, R.-M., Reginster, J.-Y., et al. (2018). Correction to: Psychometric performance of the Romanian version of the SarQoL(R), a health-related quality of life questionnaire for sarcopenia. Archives of Osteoporosis, 13, 98.

17. Alekna, V., Kilaite, J., Tamulaitiene, M., Geerinck, A., Mastaviciute, A., Bruyère, O., et al. (2019). Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). European Geriatric Medicine, 10(5), 761–767.

18. Safonova, Y. A., Lesnyak, O. M., Baranova, I. A., Suleimanova, A. K., & Zotkin, E. G. (2019). Russian translation and validation of SarQoL®—Quality of life questionnaire for patients with sarcopenia. Nauchno-Prakticheskaya Revmatol, 57, 38–45.

19. Fábrega-Cuadros, R., Martínez-Amat, A., Cruz-Díaz, D., Aibar-Almazán, A., & Hita-Contreras, F. (2020). Psychometric properties of the Spanish version of the sarcopenia and quality of life, a Quality of Life Questionnaire Specific for Sarcopenia. Calcified Tissue International, 106, 274–282.

20. Geerinck, A., Bruyère, O., Locquet, M., Reginster, J. Y., & Beaudart, C. (2018). Evaluation of the responsiveness of the SarQoL® questionnaire, a patient-reported outcome measure specific to sarcopenia. Advances in Therapy, 35, 1842–1858.

21. U.S. Department of Health and Human Services F and DA, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health (2009). Guidance for industry use in medical product development to support labeling claims guidance for industry. Retrieved from https://www.fda.gov/media/77832/download.

22. Cruz-Jentoft, A., Baeyens, J. P., Bauer, J., Boirie, Y., Cederholm, T., Landi, F., et al. (2010). Sarcopenia: European consensus on definition and diagnosis. Age and Ageing, 39, 412–423.

23. Broder, H. L., McGrath, C., & Cisneros, G. J. (2007). Questionnaire development: Face validity and item impact testing of the child oral health impact profile. Community Dentistry and Oral Epidemiology, 35, 8–19.

24. Beaudart, C., Reginster, J. Y., Petermans, J., Gillain, S., Quabron, A., Locquet, M., et al. (2015). Quality of life and physical components linked to sarcopenia: The SarcoPhAge study. Experimental Gerontology, 69, 103–110.

25. Locquet, M., Beaudart, C., Hajaoui, M., Petermans, J., Reginster, J. Y., & Bruyère, O. (2018). Three-year adverse health consequences of sarcopenia in community-dwelling older adults according to 5 diagnosis definitions. Journal of the American Medical Directors Association, 20, 43-46.e2.

26. The EuroQol Group. (1990). EuroQol: A new facility for the measurement of health-related quality of life. Health Policy (New York), 16, 199–208.

27. de Vet, H. C. W., Terwee, C. B., & Mokkink, L. B. (2011). Measurement in medicine: A practical guide. . University Press.

28. Malmstrom, T. K., & Morley, J. E. (2013). SARC-F: A simple questionnaire to rapidly diagnose sarcopenia. Journal of the American Medical Directors Association, 14, 531–532.

29. Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105, 399–412.

30. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159.

31. Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. Statistical Methods in Medical Research, 2802, 135–160.

32. Brown, T. (2015). Confirmatory factor analysis for applied research. (2nd ed.). The Guildford Press.

33. Kline, R. (2016). Principles and practice of structural equation modeling. (4th ed.). The Guildford Press.

34. DeMars, C. (2010). Item response theory. . Oxford University Press.

35. Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? International Journal of Human Computer Interaction, 32, 51–62.

36. Lameijer, C. M., Van Bruggen, S. G. J., Haan, E. J. A., Van Deurzen, D. F. P., Van Der Elst, K., Stouten, V., et al. (2020). Graded response model fit, measurement invariance and (comparative) precision of the Dutch-Flemish PROMIS® upper extremity V2.0 item bank in patients with upper extremity disorders. BMC Musculoskeletal Disorders, 21, 1–17.

37. Goetz, C., Coste, J., Lemetayer, F., Rat, A. C., Montel, S., Recchia, S., et al. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. Journal of Clinical Epidemiology, 66, 710–718.

Chapter 7: The relative importance of aspects of quality of life

Patients' preferences for quality of life aspects in sarcopenia: a best-worst scaling study

Submitted to *European Geriatric Medicine*

Geerinck A, Locquet M, Hiligsmann M, Reginster J-Y, Bruyère O, Beaudart C. Patients' preferences for quality of life aspects in sarcopenia: a best-worst scaling study

1. Abstract

Purpose: As information on patients' preferences regarding quality of life aspects in sarcopenia is lacking, this study aims to assess the relative importance of the 14 items of a QoL questionnaire designed for sarcopenia (the SF-SarQoL) using a best-worst scaling (BWS) survey.

Methods: Participants, aged 65 years or older and community-dwelling, who previously participated in the SarcoPhAge study, received a BWS survey via the mail. An object case BWS was selected in which participants completed 12 choice tasks, picking the most and least important aspect from 4 out of 14 SFSarQoL items for each task. Relative importance scores (RIS) were estimated using Hierarchical Bayes modelling. A cluster analysis was also conducted to investigate whether several profiles with regards to QoL preferences were present.

Results: A total of 163 participants were included, aged 75 (IQR: 73-81) years old, and mostly women (n=107; 65.6%). Two items were found to be significantly more important than others: "feeling a reduction of physical capacity" (RIS=11.26), and "having balance problems" (RIS=11.09). The least important items were "experiencing difficulty carrying heavy objects" (RIS=2.89), and "feeling a reduction in muscle mass" (RIS=3.82). We found relatively weak evidence for the presence of two clusters. One cluster prioritized items related to falls where the second prioritized items related to feeling physically capable.

Conclusion: Not all QoL aspects were equally important. The relative weight of each QoL aspect may be used to interpret QoL results obtained with the SF-SarQoL or to inform target outcomes in interventional studies.

2. Introduction

Sarcopenia, the skeletal muscle disorder characterized by a loss of muscle strength and function, can have a significant impact on those affected. It has been shown to be associated with a number of adverse outcomes such as mortality, functional decline, disability, falls and hospitalization [1]. This impact on a personal level cascades into impact on the health systems that provide care to people with sarcopenia, and economic studies have found significantly higher healthcare costs for sarcopenic people both in a hospital setting as well as in the community [2].

Previous research in sarcopenia has mainly focused on so-called hard outcomes (such as mortality or hospitalizations), but interest in the lived experience of sarcopenic patients has been steadily growing. More and more studies are reporting results for quality of life (QoL), mostly concluding that sarcopenic people have lower quality of life compared to non-sarcopenic people [3]. Other examples of patient-reported outcomes are pain, physical function, satisfaction with care, etc. A recent working group organized by the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO) emphasized that inclusion of a patient-reported outcome measure (PROM) in clinical trials of pharmaceutical interventions for sarcopenia is highly desirable [4]. The FDA has also encouraged the appropriate use of PROMs in regulatory studies, and has observed a 500% increase in the number of pre-market submissions that include PROMs between 2009 and 2015 [5].

A number of generic QoL questionnaires (i.e., designed for use across different populations) are regularly used in sarcopenia research, most notably the SF-36 and the EQ-5D. A QoL questionnaire specifically designed for sarcopenia, called the Sarcopenia Quality of Life (SarQoL®) questionnaire, has also been available since 2015 [6]. The SarQoL® is recommended for use with older, community-dwelling individuals experiencing a loss in muscle strength and function. It is based on a multidimensional concept of QoL, encompassing 55 items from 7 domains of health-related dysfunction: physical and mental health, locomotion, body composition, functionality, activities of daily living, leisure activities, and fears [6,7]. Recently, a shorter version of the SarQoL® questionnaire was developed, which reduced the length of the questionnaire from 55 to 14 items [8]. The SF-SarQoL questionnaire is available from the website www.sarqol.org in multiple languages.

Most QoL instruments translate the individual responses gathered with the tool in question into one or several scores, representing domains of QoL or the global level of QoL of the respondent. This approach is necessary for quantitative research on groups of people but reduces the complex concept of QoL to a number on a scale. While very useful, it should not be controversial to say that a single score does not tell us the whole story about a person's QoL. Researchers can often delve deeper into the gathered QoL results, by looking at domain scores or even the item responses themselves, which is already an

improvement over an overall score. However, this does not take into account that not all aspects of QoL are created equal: some items are likely to be considered more important by patients than others.

This type of information, the importance of one aspect/item/outcome in relation to others, can be obtained through choice modelling, of which the most frequently used designs are the discrete choice experiment (DCE) and best-worst scaling (BWS). DCE and BWS are already regularly employed to gauge patients' preferences regarding treatments [9,10]. Recently, a DCE was also used to look at which clinical outcomes were considered important by sarcopenic older persons, the first study of its kind in sarcopenia [11,12]. Interestingly, the participants of this study identified QoL as one of the 5 most important outcomes for sarcopenia interventions [12]. In comparison to a DCE, the BWS method is considered to be less cognitively demanding, gathers additional information on the least preferred option and is capable of capturing preferences for a longer list of items/attributes [10,13].

The primary objective of the present study was to establish a ranking from most to least important for the 14 aspects of QoL included in the SF-SarQoL® questionnaire using the best-worst scaling technique. The secondary objective of this study was to explore whether different profiles were present within the sample with regards to their ranking of the 14 aspects of QoL with the help of a cluster analysis.

3. Methods

3.1. Population

This study recruited older, community-dwelling people who had previously participated in the Sarcopenia and Physical Impairment with Advancing Age (SarcoPhAge) study. This is a 5-year cohort study, carried out in the Liège region of Belgium, which focused on a range of musculoskeletal indicators. All participants were aged 65 years and older at inclusion, with a body mass index below 50 kg/m² and without amputated limbs. Details on this study and several articles on different results have previously been published [14]. For the best-worst scaling study presented in this article, 314 individuals who had participated in the interviews for the 4th (July 2017 to September 2018) and/or 5th (June 2018 to November 2019) year of follow-up of the SarcoPhAge study, and for whom demographic and clinical data from these interviews were available, were contacted with an invitation to participate in February/March 2020. The research protocol (n° 2012/277) and its amendment dated 19/12/2019 were approved by the Ethics Committee of the University Teaching Hospital of Liège.

3.2. Study design

Patient preferences were elicited through an object (case 1) BWS survey. This type of choice experiment was first developed by Jordan J. Louvière in 1987, and its use in health care research was proposed in 2005 [13,15]. The objective of this type of choice experiment is to place objects (in this case different aspects of QoL) on an underlying, subjective, latent scale by having volunteers complete choice tasks in which they are asked to indicate the "best" (in this case: most important for QoL) and "worst" (in this case: least important for QoL) object from 3 or more options [15]. By analyzing choice frequency, for both "best" and "worst" choices, utility values can be calculated for each object, and a ranking from best to worst can be established [13].

The 14 items of the SF-SarQoL questionnaire constituted the list used to create the choice tasks in the BWS survey [8]. Twelve choice tasks of 4 items were presented to each participant to strike a balance between obtaining as much information as possible, without creating too much response burden. An example of a choice task from the BWS survey can be found in figure 1. Sawtooth Software was used to generate 2 versions of the BWS survey. The design algorithm of the Sawtooth software is considered to be similar to the Balanced Incomplete Block Design and takes into account frequency balance, orthogonality, connectivity and positional balance [16]. Participants were randomly assigned to receive either version A or B using IBM's SPSS software.

Figure 1: Example of a choice task in the BWS survey (translated from the original French).

In the table below, the participant has indicated that experiencing difficulty during activities of moderate effort is the most important aspects with regards to their quality of life, and reducing their leisure activities the least important.

Least important		**Most** important
	Feeling a reduction of the strength in your arms	
	Experiencing difficulty during activities of moderate effort	**X**
	Having problems with your balance	
X	Reducing your leisure activities	

Participants received a paper copy of the BWS questionnaire through the postal service. They completed the questionnaire at home and returned it through the mail using an included pre-paid envelope.

Participants also received and completed the SF-SarQoL questionnaire itself at the same time as the BWS survey. This shorter version of the SarQoL® measures overall QoL through 14 items and has been

validated for use in sarcopenia [8]. It provides a single score between 0 and 100 points, with greater scores indicating better QoL.

Clinical and demographic information was obtained from the interviews conducted at the 5th year of the SarcoPhAge study. If no data was collected at the 5th year interview (because of drop-out or missing data), the information collected at the 4th year of follow-up was used. Muscle mass was evaluated with dual x-ray absorptiometry, and muscle strength with a hydraulic hand dynamometer. We used the EWGSOP2 criteria to determine the sarcopenia status of each participant. Those with low grip strength, defined as <27 kg for men and <16 kg for women, were considered to be probably sarcopenic. If the persons with low grip strength also presented with low muscle mass, defined as an appendicular skeletal muscle mass divided by height-squared (ASM/Ht²) <7.0 kg/m² for men and <5.5 kg/m² for women, they were diagnosed as sarcopenic [17]. Participants also completed the SARC-F questionnaire, a screening tool which identifies those with a high probably of being sarcopenic through 5 questions on strength, assistance with walking, rising from a chair, climbing stairs and falls. Participants who scored 4 or more points (on a scale between 0 and 10) were considered to be likely sarcopenic [18].

3.3. Statistical analysis

The distribution of the continuous demographic and clinical variables was evaluated by looking at the Shapiro-Wilk test, histograms, Q-Q plots, and the distance between mean and median. Variables that were normally distributed are presented as mean ± standard deviation, those that are not presented as median (25th percentile – 75th percentile). Binary variables are presented as absolute and relative frequencies [n(%)].

Relative importance scores (RIS) were estimated using Hierarchical Bayes estimation modelled using multinomial logit. The raw RIS were rescaled so that the sum of all RIS was 100 [16]. RIS are presented as mean (95% confidence interval of the mean). A fit statistic was calculated for each respondent, quantifying the probability that a participant has answered in a random manner. Surveys with a fit statistic below 0.25, indicating a significant probability of random responses by the participant, were excluded [19].

Subgroup analyses were conducted between men and women, as well as between those with normal and low grip strength (defined as <27 kg for men and <16 kg for women). These two variables were chosen because of their importance in interpreting any QoL outcomes if these subgroups showed to place different importance on aspects of QoL. Additional subgroup analyses were performed (and presented in appendix 1) comparing RIS between version A and version B of the BWS survey, between SARC-F score ≥4 points and <4 points, between sarcopenic and non-sarcopenic participants (EWGSOP2 diagnostic criteria), between those aged ≤75 years and > 75 years old and between those with lower QoL

(≤47 points for the SF-SarQoL) and those with higher QoL (>47 points). P-values were calculated with Student T-test and Mann-Whitney U-test.

We carried out a cluster analysis on the obtained RIS using the Two-Step cluster strategy with the log-likelihood distance measure using logarithmically transformed versions of the 14 RIS. The number of clusters is selected by the software using the Bayesian Information Criterion. The overall goodness-of-fit of the cluster solution was evaluated with the silhouette measure of cohesion and separation, which ranges from -1 to 1. In a good cluster solution, the intra-cluster distances are small (high cohesion between elements in the same cluster) and the inter-cluster distances are large (good separation between elements from different clusters) [20]. A silhouette coefficient <0.25 indicates the absence of a substantial cluster structure; a value from 0.26 to 0.50 is considered a weak structure that could be artificial; from 0.51 to 0.70 translates to a reasonable structure; and from 0.71 to 1 the cluster solution is considered to be strong [21].

RIS were estimated and rescaled using Sawtooth Software. All statistical manipulations were carried out using SPSS v27.0.0.0. P-values ≤0.05 were considered to be significant.

4. Results

4.1. Population

Out of the 314 study candidates contacted, 163 (52%) sent back the completed BWS survey and were included in the dataset. Detail on the flow of participants throughout this study is provided in figure 2. Of these 163 people, 74 (45.4%) completed version A of the BWS survey, and 89 (54.6%) completed version B. The missing response rate per choice task ranged from 0 (%) to 6 (3.7%) observations, which makes for an average completion rate of 98.3% for the “best” choices and 98.1% for the “worst” choices. The mean fit statistic was 0.537 ± 0.110 and no participant was excluded because of a fit statistic below 0.25. Participants had a median age of 75.0 (73.0-81.0) years, and most were women (n=107, 65.6%). Additional characteristics are provided in table 1.

Figure 2: Flow-chart of participants eligible and included in this study.

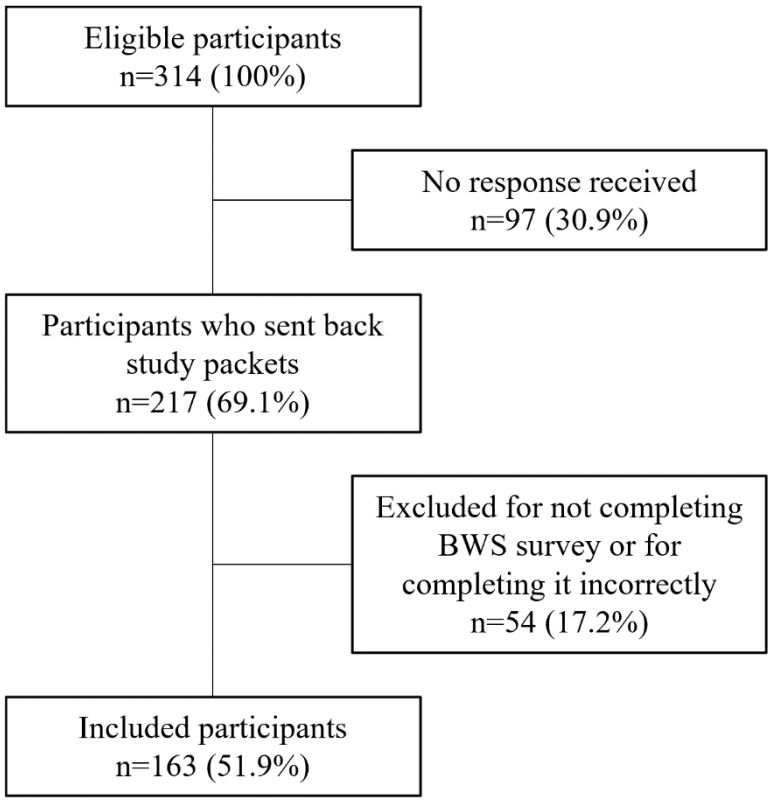


Table 1: clinical and demographic characteristics (n=163)	
	Median (IQR) or n(%)
Age (years)	75.0 (73.0 – 81.0)
SF-SarQoL QoL (0-100 points)	46.9 (27.0 – 66.1)
Gender	
Men	56 (34.4%)
Women	107 (65.6%)
Grip strength	
Low *	49 (30.1%)
Normal	113 (69.3%)
Sarcopenia	
Yes	11 (6.7%)
No	152 (93.3%)
* <27 kg for men and <16 kg for women	

4.2. Relative importance of the 14 QoL aspects

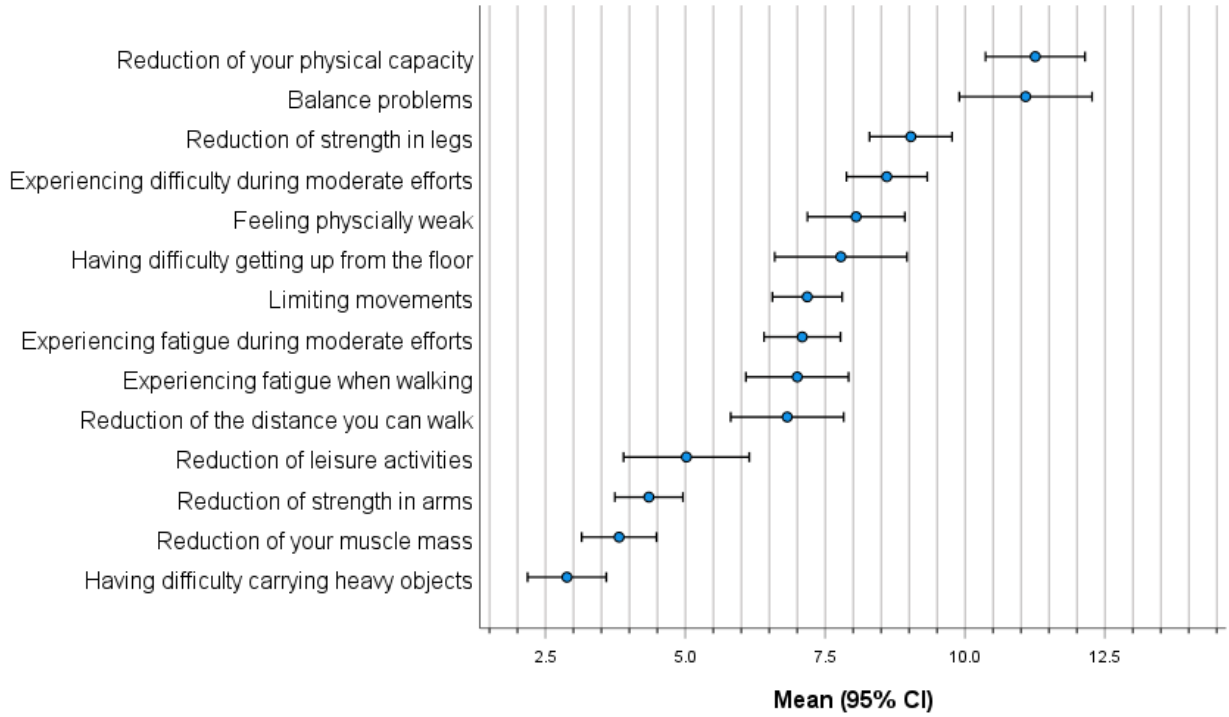
Relative importance scores calculated for the 14 aspects of QOL included in the SF-SarQoL questionnaire are presented in table 2 and figure 3. The participants considered that "feeling a reduction in their physical capacity" [11.26 (10.37-12.14)], "having balance problems" [11.09 (9.91-12.27)], and "feeling a reduction of the strength in your legs" [9.03 (8.30-9.77)] were the 3 most important aspects of QoL in sarcopenia. On the other end of the spectrum, they considered "feeling a reduction of the strength in your arms" [4.35 (3.75-4.96)], "feeling a reduction in your muscle mass" [3.82 (3.15-4.49)], and "having difficulty carrying heavy objects" [2.89 (2.19-3.59)] as the least important aspects of QoL. Relatively large 95% confidence intervals were found, and consequently an important number of items have overlapping intervals. Roughly speaking, items can be grouped together in 3 groups: the 2 items on feeling a reduction in physical capacity and experiencing balance problems are significantly more important than all other items. Next up are 8 items whose confidence intervals overlap: leg strength, difficulty during moderate effort, feeling weak, difficulty getting up from the floor, limiting movements, fatigue during moderate effort, fatigue while walking, and walking distance. Lastly, a third group of items are clearly less important than the items mentioned so far: leisure activities, arm strength, muscle mass and carrying heavy objects. It is important to add that this is a relative assessment, rating whether one item is more important than another, not an absolute assessment, rating whether an item is important or not.

Table 2 also presents the results separated by gender and grip strength. We did not find important differences in the way men and women or people with low and normal grip strength valued the 14 QoL aspects. Only the item "limiting movement" was significantly different in terms of relative importance score between men and women [8.29 (7.10 – 9.48) vs 6.60 (5.89 – 7.31); p=0.011]. The comparison on grip strength also yielded a single significantly different RIS, in this case for the item "difficulty getting up from the floor, which was considered more important by participants with low grip strength [9.89 (7.48 – 12.30) vs 6.94 (5.60 – 8.27); p=0.024].

The results of the additional subgroup analyses on BWS version (A versus B), SARC-F (≥4 points versus <4 points), sarcopenia status (sarcopenia versus no sarcopenia), age (≤75 years versus >75 years), and QoL (SF-SarQoL score ≤47 points versus >47 points) are available in appendix 1. In short, while we did find minor differences between the RIS values when comparing between groups on several characteristics, none of these differences upend the global results of the analysis on the complete sample.

Table 2: Relative Importance Scores – n=163							
Label		Gender			Grip strength		
	All	Men	Women	p-value	Normal	Low	p-value
Reduction physical capacity	11.26 (10.37 - 12.14)	11.35 (9.75 - 12.94)	11.21 (10.13 - 12.30)	0.889	11.55 (10.44 - 12.65)	10.55 (9.01 - 12.09)	0.314
Balance problems	11.09 (9.91 - 12.27)	10.96 (8.77 - 13.15)	11.15 (9.73 - 12.58)	0.881	11.05 (9.62 - 12.48)	11.34 (9.11 - 13.56)	0.826
Reduction strength legs	9.03 (8.30 - 9.77)	9.42 (8.12 - 10.71)	8.83 (7.93 - 9.74)	0.458	8.58 (7.71 - 9.45)	10.10 (8.69 - 11.50)	0.064
Difficulty during moderate efforts	8.60 (7.88 - 9.32)	8.85 (7.68 - 10.02)	8.48 (7.55 - 9.41)	0.634	8.89 (8.00 - 9.78)	7.96 (6.66 - 9.26)	0.247
Feeling physically weak	8.06 (7.19 - 8.92)	7.08 (5.52 - 8.63)	8.57 (7.52 - 9.62)	0.108	7.84 (6.74 - 8.93)	8.50 (7.03 - 9.97)	0.494
Difficulty getting up from the floor	7.78 (6.61 - 8.96)	7.56 (5.72 - 9.4)	7.90 (6.36 - 9.44)	0.786	6.94 (5.60 - 8.27)	9.89 (7.48 - 12.30)	0.024
Limiting movement	7.18 (6.56 - 7.8)	8.29 (7.10 - 9.48)	6.60 (5.89 - 7.31)	0.011	7.54 (6.76 - 8.32)	6.43 (5.40 - 7.45)	0.108
Fatigue during moderate effort	7.09 (6.41 - 7.77)	6.67 (5.55 - 7.8)	7.31 (6.44 - 8.18)	0.382	7.40 (6.54 - 8.26)	6.41 (5.28 - 7.53)	0.190
Fatigue while walking	7.00 (6.09 - 7.92)	6.98 (5.35 - 8.61)	7.02 (5.89 - 8.14)	0.971	6.85 (5.70 - 7.99)	7.34 (5.76 - 8.93)	0.627
Reduction walking distance	6.82 (5.82 - 7.83)	7.01 (5.17 - 8.86)	6.73 (5.51 - 7.95)	0.792	6.72 (5.49 - 7.96)	7.02 (5.17 - 8.86)	0.794
Reduction leisure activities	5.02 (3.90 - 6.14)	4.16 (2.38 - 5.95)	5.47 (4.03 - 6.92)	0.275	3.86 (6.62 - 4.55)	4.19 (2.29 - 6.08)	0.391
Reduction strength arms	4.35 (3.75 - 4.96)	4.63 (3.67 - 5.59)	4.20 (3.42 - 4.99)	0.512	4.53 (3.81 - 5.26)	3.96 (2.80 - 5.13)	0.398
Reduction muscle mass	3.82 (3.15 - 4.49)	3.93 (2.65 - 5.2)	3.76 (2.97 - 4.55)	0.820	3.67 (2.86 - 4.49)	4.12 (2.88 - 5.37)	0.546
Difficulty carrying heavy objects	2.89 (2.19 - 3.59)	3.12 (1.95 - 4.29)	2.76 (1.87 - 3.65)	0.630	3.20 (2.33 - 4.07)	2.20 (0.98 - 3.43)	0.201
Average importance (100/14) is 7.14. The double line indicates the cut-point between above and below average importance for RI scores. Items are presented from most important to least important according to RIS. CI= confidence interval							

Figure 3: relative importance scores



4.3. Cluster analysis

The cluster analysis detected 2 distinct clusters within the sample. The value for the silhouette measure of cohesion and separation was 0.3, indicating that the cluster solution found is relatively weak and should be interpreted with caution. The largest cluster had 88 members, while the second cluster was slightly smaller at 75 members. Relative importance scores and rank for the 14 aspects of quality of life are presented for each cluster in table 3.

Overall, cluster 1 found items related to falls (i.e. getting up from the floor, leg strength and balance) to be the most important and cluster 2 prioritized feeling physically capable. Both clusters shared the item "balance problems" in their top 3 of most important items, and "carrying heavy objects" as 1 of the 3 least important items.

Table 3 : RIS and ranking per cluster				
	Cluster 1 (n=88)		Cluster 2 (n=75)	
	RIS	Ranking	RIS	Ranking
Reduction physical capacity	8.84 (7.66 - 10.02)	4	14.1 (13.05 - 15.14)	1
Balance problems	10.04 (8.32 - 11.76)	3	12.31 (10.71 - 13.91)	2
Reduction strength legs	11.21 (10.2 - 12.22)	2	6.48 (5.73 - 7.23)	10
Difficulty during moderate efforts	7.43 (6.57 - 8.29)	5	9.98 (8.83 - 11.12)	4
Feeling physically weak	4.84 (3.98 - 5.71)	13	11.82 (10.72 - 12.93)	3
Difficulty getting up from the floor	11.54 (9.93 - 13.16)	1	3.37 (2.26 - 4.48)	11
Limiting movement	6.35 (5.63 - 7.08)	8	8.15 (7.12 - 9.18)	6
Fatigue during moderate effort	5.87 (5.1 - 6.63)	10	8.53 (7.41 - 9.65)	5
Fatigue while walking	6.64 (5.34 - 7.95)	7	7.43 (6.12 - 8.73)	7
Reduction walking distance	6.95 (5.54 - 8.35)	6	6.68 (5.19 - 8.17)	8
Reduction leisure activities	3.64 (2.22 - 5.05)	14	6.65 (4.89 - 8.4)	9
Reduction strength arms	6.26 (5.39 - 7.14)	9	2.11 (1.63 - 2.59)	12
Reduction muscle mass	5.39 (4.43 - 6.35)	11	1.97 (1.22 - 2.73)	13
Difficulty carrying heavy objects	4.98 (3.86 - 6.11)	12	0.43 (0.22 - 0.63)	14

5. Discussion

This study suggests that older people do not consider all items of musculoskeletal QoL represented in the BWS survey to be equally important. The ranking established in this study showed the QoL aspects “reduction of your physical capacity” and “experiencing balance problems” to be the most important. Within the sample described in this article, two different profiles were found with regards to the importance placed on certain aspects of musculoskeletal QoL. While the silhouette measure indicated that the structure found was weak, and that it could be artificial, it is not hard to imagine that there are likely different groups with different sets of priorities with regards to QoL. While we would caution against over-interpreting these results based on this sample alone, the choices made within the 2 clusters seem to make sense in that they coalesce around two themes: falls and physical capacity. The first one,

falls, had already been identified in a previous study using focus groups, but the second one, physical capacity, had not yet been put forward [12].

To our knowledge, this is the first study to demonstrate the relative importance of different aspects of QoL in a quantative manner in sarcopenia. Unfortunately, because of the highly specific nature of the SF-SarQoL, and its focus on musculoskeletal aspects of QoL that are relevant to sarcopenic patients, we are unable to directly compare our findings with other studies, because of the heterogeneity of the items studied under the umbrella of the concept of QoL. There are however a limited number of studies which have investigated how older people think about QoL and what aspects they consider to be more or less important, employing broader concepts of QoL than used in our own BWS survey.

A thematic synthesis by Van Leeuwen and colleagues compiled a number of qualitative studies on the subject and is the most thorough overview of what QoL means to older people. The authors included 48 studies, incorporating the perspectives of more than 3400 older community-dwelling people from Western countries. From this vast amount of information, they distilled nine QoL domains: health perception, autonomy, role and activity, relationships, attitude and adaptation, emotional comfort, spirituality, home and neighborhood, and financial security. They also stress the interconnections between domains and the ripple effect of changes in a particular domain on the other domains. This exhaustive synthesis however was not set up to indicate which aspects or domains of QoL are the most important, or to establish a hierarchy among the nine domains, favoring instead the broadest possible concept of QoL [22].

In terms of quantitative research, there are three studies that have surveyed the relative importance of different aspects of QoL in the specific population of older people. Molzahn and colleagues published the results of a secondary analysis of the WHOQOL-OLD pilot study in 2011. In this article, they present data collected from 7401 people aged 60 years or older from 22 countries on the importance of 31 facets of QoL. The participants in this study considered ADL, general health, sensory abilities, mobility, autonomy, and energy to be the most important QoL facets, in the order presented. With regards to the least important facets, they singled out sex-life, opportunity to learn new skills, social participation, and a positive body image and appearance [23]. While the items in the Molzahn study and our own survey are too dissimilar to compare head-to-head, it is interesting to note that the concepts considered important to the older people in the Molzahn study, such as ADL, general health, mobility, and autonomy, are well represented in the SF-SarQoL, while the concepts considered less important are not represented. A second study, carried out by Ratcliffe and colleagues and published in 2017, recruited 500 younger people (18-64 years) and 500 older people (65+ years) who performed two preference elicitation experiments (ranking and successive BWS) aimed at establishing a hierarchy of 12 quality of life dimensions. The older sample found the dimensions independence, physical mobility, control, and mental health particularly important in the ranking experiment, with similar results for the BWS

task. While the items in this study are again too dissimilar to our own BWS survey, we note the importance that the participants of our study placed on their physical capacity, balance, and strength in the legs, and hypothesize that these items may be considered as prerequisites for independence and physical mobility, considered important in the Ratcliffe study. This study also demonstrated that the preferences of younger and older people with regards to QoL are different [24]. Lastly, Uy and colleagues conducted a BWS experiment in Singapore of which they published the result in 2018. They sought to establish a ranking of 27 health-related QoL domains and recruited 603 participants aged between 21 and 88 years old to do this. The BWS results placed the domain "self-care" at the top of the hierarchy, followed by "healing and resistance to illness" and "social relationships". At the other end of the scale, the participants considered "having a satisfactory sex life" as the least important aspect of HRQoL, followed by "having a normal physical appearance", and "interacting with others" [25]. However, because of the earlier finding that QoL preferences are different between older and younger participants, these results should be interpreted with caution.

As with any study, there are some methodological and practical limitations that need to be addressed. A first limitation is that, because of the recruitment and administration methods of this study, there is the potential for non-response bias. A total of 314 potential participants were contacted, and we received responses from 217 of them, a 69% response rate. Out of those 217, a further 54 participants either did not complete the BWS survey at all, or failed to complete the survey correctly (e.g., multiple "best" choices for a single choice task). This means that we were able to include 52% of the people we contacted, and 75% of the people who participated, in the final analyses. When we compared the 163 participants included in this analysis with the 54 that responded but were excluded, we did not find a significant difference for age ($p=0.300$), gender ($p=0.183$) and probable sarcopenia/low grip strength ($p=0.155$). We did however find that a larger proportion of the sarcopenic participants in the sample were unable or unwilling to complete the choice tasks, compared to the non-sarcopenic participants (52% completion rate versus 78%; $p=0.011$). The 54 excluded participants also had significantly lower QoL [33.33 (18.27-44.55) vs. 45.99 (27.65-65.38) points on a scale from 0 to 100 measured with the SF-SarQoL; $p=0.001$] compared to the 163 included participants. This phenomenon may be related to the relative burden of the choice task, which may have been perceived as greater by sarcopenic participants and by those that already had substantially reduced quality of life. A second limitation is the sample size itself. Although there are currently no guidelines for minimum sample size for BWS surveys available in the literature, a review from 2016 found a median sample size of 175 participants (range: 15 to 803) for 26 object case BWS studies, in line with our own sample of 163 participants [10]. However, the relatively large confidence intervals found for the relative importance scores, which prohibit us from clearly separating some items, would likely have been narrower with a greater sample. This is especially noticeable for the items ranked at the middle of the importance hierarchy, where there are 8 items with overlapping confidence intervals. A third limitation of this study is that it was

conducted in a single setting, namely older, community-dwelling volunteers from the Liège province in Belgium. Without further data it is uncertain whether our results can be generalized to the wider population of older people in Belgium or whether the results of this study are transferable to other countries.

This study could however open up some perspectives for the future. The ranking established could assist in a more detailed analysis of QoL data obtained with the SF-SarQoL, either by an item-based analysis taking into account the relative ranking of the item in question, or by creating a preference-weighted overall QoL score for the SF-SarQoL. It could also inform specific targets for improvement in interventional studies or inspire the design of interventions so as to increase the effect on physical capacity, balance, and leg strength.

In conclusion, this study provides the first data on the relative importance of different aspects of QoL in the context of sarcopenia from the subjective perspective of the patient. We established a ranking of 14 aspects of QoL on importance and showed that there were two clusters present in the sample with different priorities with regards to QoL.

6. References

1. Beaudart C, Zaaria M, Pasleau F, Reginster J-Y, Bruyère O. Health outcomes of sarcopenia: a systematic seview and meta-analysis. PLoS One 2017;12:e0169548.
2. Bruyère O, Beaudart C, Ethgen O, Reginster JY, Locquet M. The health economics burden of sarcopenia: a systematic review. Maturitas 2019;119:61–9. https://doi.org/10.1016/j.maturitas.2018.11.003.
3. Woo T, Yu S, Visvanathan R. Systematic literature review on the relationship between biomarkers of sarcopenia and quality of life in older people. J Frailty Aging 2016;5:88–99. https://doi.org/10.14283/jfa.2016.93.
4. Reginster JY, Beaudart C, Al-Daghri N, Avouac B, Bauer J, Bere N, et al. Update on the ESCEO recommendation for the conduct of clinical trials for drugs aiming at the treatment of sarcopenia in older adults. Aging Clin Exp Res 2020. https://doi.org/10.1007/s40520-020-01663-4.
5. Food and Drug Administration. Value and Use of Patient-Reported Outcomes (PROs) in Assessing Effects of Medical Devices. 2017.
6. Beaudart C, Biver E, Reginster J-YJ-Y, Rizzoli R, Rolland Y, Bautmans I, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. Age Ageing 2015;44:960–6. https://doi.org/10.1093/ageing/afv133.
7. Beaudart C, Biver E, Reginster JY, Rizzoli R, Rolland Y, Bautmans I, et al. Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 2017;8:238–44. https://doi.org/10.1002/jcsm.12149.
8. Geerinck A, Beaudart C, Reginster J-Y, Locquet M, Monseur C, Gillain S, et al. Development and validation of a short version of the Sarcopenia Quality of Life questionnaire : the SF - SarQoL. Qual Life Res 2021. https://doi.org/10.1007/s11136-021-02823-3.
9. Hiligsmann M, Pinto D, Dennison E, Al-Daghri N, Beaudart C, Branco J, et al. Patients' preferences for osteoarthritis treatment: the value of stated-preference studies. Aging Clin Exp Res 2019;31:1–3. https://doi.org/10.1007/s40520-018-1098-3.
10. Cheung KL, Wijnen B, Hollin I, Janssen E, Bridges J, Evers S, et al. Using Best – Worst Scaling to Investigate Preferences in Health Care. Pharmacoeconomics 2016;34:1195–209. https://doi.org/10.1007/s40273-016-0429-5.

11. Hiligsmann M, Beaudart C, Bruyère O, Biver E, Bauer J, Cruz-Jentoft AJ, et al. Outcome Priorities for Older Persons With Sarcopenia. J Am Med Dir Assoc 2020;21:267-271.e2. https://doi.org/10.1016/j.jamda.2019.08.026.

12. Beaudart C, Bruyère O, Cruz-Jentoft AJ, Vaquero-Pinto MN, Locquet M, Bauer J, et al. Patient's Engagement in the Identification of Critical Outcomes in Sarcopenia. J Am Med Dir Assoc 2020;21:284–6. https://doi.org/10.1016/j.jamda.2019.09.004.

13. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: What it can do for health care research and how to do it. J Health Econ 2007;26:171–89. https://doi.org/10.1016/j.jhealeco.2006.04.002.

14. Beaudart C, Reginster JY, Petermans J, Gillain S, Quabron A, Locquet M, et al. Quality of life and physical components linked to sarcopenia: The SarcoPhAge study. Exp Gerontol 2015;69:103–10.

15. Louvière JJ, Flynn TN, Marley AAJ. Best-worst scaling: theory, methods and applications. Cambridge University Press; 2015.

16. Sawtooth Software. Lighthouse Studio Manual n.d. https://www.sawtoothsoftware.com/help/lighthouse-studio/manual/index.html?hid_web_maxdiff_numberofsets.html (accessed November 9, 2020).

17. Cruz-Jentoft AJ, Bahat G, Bauer J, Boirie Y, Bruyère O, Cederholm T, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019;48:16–31. https://doi.org/10.1093/ageing/afy169.

18. Malmstrom TK, Miller DK, Simonsick EM, Ferrucci L, Morley JE. SARC-F: A symptom score to predict persons with sarcopenia at risk for poor functional outcomes. J Cachexia Sarcopenia Muscle 2016;7:28–36. https://doi.org/10.1002/jcsm.12048.

19. Sawtooth Software. Identifying bad respondents: fit statistic and identifying random responders n.d. https://sawtoothsoftware.com/help/lighthouse-studio/manual/hid_web_maxdiff_badrespondents.html.

20. Norusis MJ. IBM SPSS Statistics 19 Statistical Procedures Companion: Cluster Analysis, Chapter 17 2011:361–91. https://doi.org/10.4135/9781412983648.

21. Kaufman L, Rousseeuw P. Finding groups in data: an introduction to cluster analysis. Wiley-Interscience; 1990.

22. Van Leeuwen KM, Van Loon MS, Van Nes FA, Bosmans JE, De Veti HCW, Ket JCF, et al. What does quality of life mean to older adults? A thematic synthesis. vol. 14. 2019. https://doi.org/10.1371/journal.pone.0213263.

23. Molzahn AE, Kalfoss M, Makaroff KS, Skevington SM. Comparing the importance of different aspects of quality of life to older adults across diverse cultures. Age Ageing 2011;40:192–9. https://doi.org/10.1093/ageing/afq156.

24. Ratcliffe J, Lancsar E, Flint T, Kaambwa B, Walker R, Lewin G, et al. Does one size fit all? Assessing the preferences of older and younger people for attributes of quality of life. Qual Life Res 2017;26:299–309. https://doi.org/10.1007/s11136-016-1391-6.

25. Uy EJ, Bautista DC, Xin X, Cheung YB, Thio ST, Thumboo J. Using best-worst scaling choice experiments to elicit the most important domains of health for health-related quality of life in Singapore. PLoS One 2018;13:1–18. https://doi.org/10.1371/journal.pone.0189687.

Discussion

Throughout this dissertation, our overarching objective was to contribute to the improvement of the measurement of quality of life in sarcopenia. This goal is important because of the role of quality of life as one of the rare indicators that translates patients' lived experiences into a quantitative variable that can be used in statistical analysis and group-based comparisons. It also refocuses attention from a disease or condition back to the patient, a perspective which can sometimes become lost in clinical trials. Because of this particular role, it is important that a valid, reliable, and precise instrument is available to measure quality of life with, and that the measurement properties of such an instrument are well-documented, which is, in part, what we set out to do within this dissertation. A famous quote, attributed to Lord Kelvin (a mathematical physicist and engineer born in 1824) states that "*When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be."* [1]. While he was talking about measuring electricity, and not a person's subjective assessment of their own well-being, his basic message is equally applicable to health-related quality of life. Precise and accurate measurement, in as simple a way as possible and with the lowest possible response burden, is vital to naming and understanding a phenomenon.

The projects within this dissertation were developed along two axes. In the first part of this dissertation, the measurement properties of the SarQoL questionnaire and its functioning in different contexts were documented. The second part builds upon this with the development and validation of a short form SarQoL questionnaire and an investigation of the relative importance of the 14 items included in this short form. In this way, we contributed to the aim of "precise and accurate measurement" in part 1, and "simple and with low response burden" in part 2.

Within the framework of the first part of this dissertation, we demonstrated that: (1) The SarQoL questionnaire is responsive to changes in quality of life over time; (2) the SarQoL questionnaire is a reliable instrument and has no systematic error; (3) the random error of the questionnaire is 2.65 points and the smallest detectable change for an individual patient is 7.35 points; (4) the SarQoL questionnaire can still discriminate between sarcopenic and non-sarcopenic people with the revised EWGSOP2 criteria for sarcopenia; (5) the SarQoL questionnaire demonstrated adequate measurement properties to allow its use in research focused on physical frailty diagnosed with the Fried criteria; and (6) the SarQoL questionnaire could prove to be useful as a screening instrument for sarcopenia. The results obtained in the first part of this dissertation complement the first table in the introduction section (page 11 & 12), which listed the validation studies performed on the SarQoL questionnaire. A complete overview of all

measurement properties documented for the SarQoL questionnaire, in order of publication, is displayed in the table on the next page.

In the second part of this dissertation, we reduced the number of items in the SarQoL questionnaire from 55 to 14, retaining 6 out of the 7 domains of health-related dysfunction, which resulted in the creation of the SF-SarQoL. Subsequently, we examined its measurement properties in a sample of 214 older, community-dwelling people and demonstrated that: (1) there were no floor or ceiling effects present; (2) the SF-SarQoL can discriminate between sarcopenic and non-sarcopenic participants diagnosed with the EWGSOP2 criteria, as well as between those with low grip strength (probable sarcopenia in the EWGSOP2 algorithm) and those with normal grip strength; (3) The SF-SarQoL was highly internally consistent; (4) The SF-SarQoL demonstrated excellent test-retest reliability, both for the total score as well as for the individual items. However, a small but significant systematic bias was present in the sample with participants scoring an average of 4.11 points higher on the second administration; (5) the construct validity of the SF-SarQoL was confirmed through the criterion approach, taking the original SarQoL questionnaire as the gold standard, and through strong correlations with the EQ-5D and EQ-VAS scores; (6) We presented two structural models for the SF-SarQoL: a single factor model where some fit indices were unsatisfactory, and a 2-factor model where the 5 items of the first question load on the first factor and the remaining 9 items load on the second factor, with a correlated residual variance specified between items 1.5 (reduction in the distance you can walk) and 4 (feeling tired when walking). This second model obtained satisfactory fit indices; (7) the Graded Response model showed no misfitting items and good category separation.

Also within the second part of this dissertation, we carried out a best-worst scaling study with the 14 items of the SF-SarQoL, and showed that: (1) not all aspects of quality of life are equally important to older people; (2) two aspects, namely “feeling a reduction of physical capacity” and “having balance problems” were significantly more important than the other items; (3) there were 2 clusters present, one considering items related to falls as the most important and the other prioritizing items related to feeling physically capable.

Table 1: Results from studies on the measurement properties of the SarQoL questionnaire

Version (year of publication)	Reference	Sample size (n)	Sarcopenic subjects [n(%)]	Sarcopenia diagnosis	Discriminative power (overall SarQoL score)	Internal consistency (α)	Construct validity	Test-retest reliability (ICC)	SEM and SDC	Floor and ceiling effects	Responsiveness
French (2016)	[2]	296	43 (14.5%)	EWGSOP	S: 54.7 (45.9-66.3) NS: 67.8 (57.3-79.0) P<0.001	0.87	Convergent and divergent hypotheses confirmed	0.91	**SEM: 4.06 SDC: 11.34**	absent	
English (2016)	[3]	297	14	EWGSOP	S: 61.9 ± 16.5 NS: 71.3 ± 12.8 P=0.01	0.88	Convergent and divergent validity confirmed	0.95 (0.92-0.97)	**SEM: 4.20 SDC: 11.65**	absent	
Romanian (2017)	[4]	100	13 (13.0%)	EWGSOP	S: 57.3 (34.4-70.7) NS: 68.4 (55.7-85.2) P=0.018	0.946	Convergent hypotheses confirmed; divergent hypotheses not confirmed			absent	
French - responsiveness (2018)	**[5]**	**42**	**42 (100%)**	**EWGSOP**							**8/9 hypotheses confirmed; SRM significantly larger than other questionnaires.**
Dutch (2018)	[6]	92	30 (32.6%)	EWGSOP	S: 67.15 (54.75-81.52) NS: 79.72 (70.10-86.88) P=0.003	0.883	6/8 hypotheses confirmed	0.976 (0.947-0.989)	**SEM: 2.54 SDC: 7.05**	Absent	
Polish (2018)	[7]	106	60 (56.6%)	EWGSOP	S: 54.9 ± 16.5 NS: 63.3 ± 17.1 P=0.013	0.92	Convergent validity confirmed	0.99 (0.995-0.999)	**SEM: 1.07 SDC: 2.96**	absent	

Table 1: Results from studies on the measurement properties of the SarQoL questionnaire

Version (year of publication)	Reference	Sample size (n)	Sarcopenic subjects [n(%)]	Sarcopenia diagnosis	Discriminative power (overall SarQoL score)	Internal consistency (α)	Construct validity	Test-retest reliability (ICC)	SEM and SDC	Floor and ceiling effects	Responsiveness
Greek (2018)	[8]	176	50 (28.4%)	EWGSOP	S: 52.12 ± 11.04 NS: 68.23 ± 14.1 P<0.001	0.96	Mixed results for convergent and divergent hypotheses	0.96 (0.95-0.97)	**SEM: 3.34 SDC: 9.24**	absent	
Lithuanian (2019)	[9]	176	58 (33.0%)	EWGSOP2	S: 50.32 ± 8.58 NS: 73.75 ± 13.51 P<0.001	0.95	8/8 hypotheses confirmed	0.976 (0.959-0.986)	**SEM: 0.18 SDC: 0.49**	absent	
Multiple – SEM & SDC (2019)	**[10]**	**278**	**278 (100%)**	**EWGSOP FNIH**				**0.969 (0.961-0.975)**	**SEM: 2.65 SDC: 7.35**		
Russian (2019)	[11]	100	50 (50.0%)	EWGSOP	S: 50.65 ± 14.23 NS: 75.10 ± 14.46 P<0.001	0.924	Mixed results for convergent hypotheses	0.935 (0.91-0.96)			
Ukrainian (2020)	[12]	49	28 (57.1%)	Ishii test	S: 58.43 ± 17.13 NS: 69.89 ± 13.31 p=0.014	0.898	Convergent validity confirmed; divergent validity refuted	0.997 (0.994-0.998)		absent	
Spanish (2020)	[13]	252	66 (26.2%)	EWGSOP2	S: 71.19 (57.51–78.89) NS: 76.04 (64.83–87.07) P=0.008	0.904	Convergent and divergent hypotheses confirmed	0.99 (0.98-0.99)		absent	
Serbian (2020)	[14]	699	12 (1.7%)	EWGSOP2	S: 60.31 (44.48–68.85) NS: 64.60	0.87	6/8 hypotheses confirmed			absent	

Table 1: Results from studies on the measurement properties of the SarQoL questionnaire

Version (year of publication)	Reference	Sample size (n)	Sarcopenic subjects [n(%)]	Sarcopenia diagnosis	Discriminative power (overall SarQoL score)	Internal consistency (α)	Construct validity	Test-retest reliability (ICC)	SEM and SDC	Floor and ceiling effects	Responsiveness
					(54.93–74.50) P=0.155						
Turkish (2021)	[15]	100	27 (27.0%) low muscle strength	EWGSOP2 – probable sarcopenia	S: 50 ± 16 NS: 68.9 ± 16.9 p < 0.001	0.88	Convergent validity confirmed, divergent validity not confirmed	0.97 (0.94-0.98)		absent	
French (2021)	**[16]**	**296**	**13 (4.4%)**	**EWGSOP2**	**S: 45.83 (38.62 - 60.26) NS: 66.43 (56.10 - 78.26) P<0.001**						
Korean (2021)	[17]	450	53 (11.8%)	EWGSOP2		0.886	Convergent and divergent validity confirmed	0.977 (0.975-0.979)		absent	
French – Frailty (2021)	**[18]**	**382**	**Robust: 172 (45.0%) Pre-frail: 167 (44.0%) Frail: 43 (11.0%)**	**Fried Frailty criteria**	**Robust: 77.1 (64.35–85.90) Pre-frail: 62.54 (53.33–69.57) Frail: 49.99 (40.45–56.06) P<0.001**	**0.866**	**Convergent and divergent validity confirmed**	**0.918 (0.834-0.961)**	**SEM: 3.88 SDC: 10.76**	**absent**	**Moderate: 5/7 hypotheses confirmed SRM: -1.44**

Results obtained within the framework of this dissertation are indicated in bold.

Quality of life is an ambiguous term, whose meaning and interpretation can change upon the context in which it is used, or the population in which it is being measured. In a way, there are as many conceptual models of quality of life as there are instruments to measure it, because each one of them builds their own construct, even if the mayor components such as physical health, emotional health, social relationships, and activities of daily living, to name just a couple, are shared between them. The US Food and Drug administration has avoided getting bogged down in this discussion by describing quality of life, within the context of clinical trials, as *"A general concept that implies an evaluation of all aspects of life on general well-being"* [19]. In the context of this dissertation, the terms quality of life and health-related quality of life have been used interchangeably, but they differ slightly in their meaning. Health-related quality of life is accepted to be changeable in function of the context, and may include general health, physical functioning, physical symptoms, emotional functioning, cognitive functioning, role functioning, social well-being and functioning, sexual functioning, and existential issues [20].

Quality of life and health-related quality of life cannot be observed directly and therefore they cannot be measured directly. To get around this, questionnaires such as the SarQoL questionnaire rely on observable indicators, that theoretically should be closely related to the construct the questionnaire claims to measure. The relationship between the construct and the indicators can work in two ways: either the construct is reflected in the indicators (called effect indicators in this model), or the construct is the result of the indicators (therefore called causal indicators) [21]. In reflective models, the covariation between the effect indicators is a function of the variation in the latent construct [22]. As an example, in a questionnaire measuring the concept fatigue, the responses to the indicators "I get tired very quickly", "I have enough energy" and "I feel no desire to do anything" are directly dependent on the level of fatigue the person experiences [23]. In formative models, changes in the causal indicators will generate a change in the latent construct. A prime example of this is the construct socio-economic status, which will be directly affected by indicators such as education, income, and occupation.

The SarQoL questionnaire was developed as a reflective multi-item questionnaire, where 55 indicators categorized into 7 domains of health-related dysfunction provide a comprehensive estimation of the sarcopenia-related quality of life of the respondent. However, upon closer inspection, it becomes clear that some items in the SarQoL questionnaire are causal indicators or may function in both a causal and effect manner depending on the respondent and the circumstances. While this is less than ideal for our clinimetric analyses, because most statistical methods employed in this dissertation are developed for reflective models, the SarQoL is certainly not the only quality-of-life questionnaire that combines effect and causal indicators [21]. The potential influence of the presence of causal indicators is mostly felt in analyses that employ correlation analysis, such as the evaluation of construct validity and internal

consistency. Other measurement properties, such as responsiveness and test-retest reliability, are unaffected. Coincidentally, one of the methods of judging the validity of causal items is how important patients rate an item, and how frequently they experience it, which means that the item-impact scores we calculated during the item selection process for the development of the SF-SarQoL also served to highlight the causal items with the highest relevance [24].

Several publications, detailed in table 1 on pages 161 to 163, have reported on the measurement properties of the SarQoL questionnaire within the framework of the measurement theory called Classical Test Theory (CTT). The basic concept behind classical test theory is that the score observed for an indicator (Y_i) consists of two parts: the true score of the construct to be measured (η) and a (random) error term associated with the specific indicator (ε_i) which translates to the formula $Y_i = \eta + \varepsilon_i$ [25]. This measurement theory primarily concerns itself with the behaviour of items as a group or scale, and not with the individual items. The advantages of classical test theory are its integration in most statistical packages and the fact that individual items can be suboptimal yet still be used successfully if there are enough of them. It has important disadvantages also: because of its emphasis on quantity (more items is more precision), scales can be long and repetitive. It is also sensitive to external communalities between items masking the true score, and parameter estimates are highly dependent on the sample used, rendering comparison across samples difficult [26].

Measurement properties associated with questionnaires under the principles of classical test theory are generally classified under validity, reliability, and responsiveness [27]. The validity of the SarQoL questionnaire, i.e., does it actually measure what it claims to measure, has mainly been investigated by formulating hypotheses on the expected correlation between the SarQoL scores and the scores of the EQ-5D, EQ-VAS and SF-36 questionnaires. If the concepts used in the hypotheses are theoretically closely related, we expect to find strong correlations, and the opposite for concepts that are not closely related in terms of content. This technique has provided important information on the construct validity of the questionnaire, but it is not perfect. To be able to be confident in the results of this technique, it is important to thoroughly assess the content and concept of the questionnaire used in the correlation hypotheses, namely the SF-36 and EQ-5D/VAS. While all three instruments (SarQoL, EQ-5D & SF-36) claim to measure quality of life, we have already discussed that this term can be used to cover substantially different content depending on the context and the instrument. A systematic description of the construct and content of these questionnaires, and the relationship between the different items and domains across these different questionnaires has not been performed, which could be considered a weakness. Furthermore, the exploratory factor analyses which were carried out for the original SarQoL questionnaire during the development phase of the SF-SarQoL were unable to confirm the structural validity of the questionnaire. However, through the translation and validation studies that

have been undertaken in the last 4 years and those that are still underway, we have been granted access to several datasets containing the SarQoL questionnaire, and the opportunity to see these analyses through to the end is present and should be one of our priorities for the future. A last aspect of validity, the cross-cultural stability of the items in the SarQoL questionnaire, has not received much attention because of the focus on analyses of the scores, both domain scores and the overall quality of life score, of the SarQoL questionnaire. Given the available validation studies conducted on different language versions of the SarQoL and in different countries, it could be worthwhile to investigate differences between cultures on the responses for the items included in the SarQoL questionnaire.

The reliability of the SarQoL questionnaire is well documented, although there are still aspects that can be further explored. The internal consistency of the questionnaire has been evaluated in multiple publications, but this was usually limited to the calculation of the Cronbach's alpha and, because of the length of the questionnaire, unlikely to show low consistency. In future studies, it would be interesting to assess the internal consistency per domain as well as to use the McDonalds omega value, which is less prone to overestimating reliability in questionnaires with high numbers of items [28]. The test-retest reliability of the questionnaire has been documented several times, but only for the auto-administered paper-based version. A team in Singapore is conducting a study in which they administer the SarQoL questionnaire through an interviewer, the results of which are eagerly awaited. Because of the availability of test-retest data, we were able to calculate the measurement error associated with the questionnaire in a sample from 9 different studies, as well as calculate the standard error of measurement for each included study. Overall, we found a measurement error of 2.65 points, which translated into a smallest detectable change of 7.35 points, or about 9 percent of the range of the overall quality of life score. One of the notable aspects of this study was the heterogeneity of the measurement error values found in the 9 separate studies, with the standard error of measurement ranging from of 0.18 to 4.20 points. This translated to a range of 0.49 to 11.65 points for the smallest detectable change. A number of factors may be in play to cause this variability, including but not limited to administration procedures, participant motivation, precise length of interval between administrations, cultural factors, or issues with the translations that have made the questions less straightforward. In retrospect, we might have benefitted from using a different statistical technique called generalizability analysis. This type of reliability study is more appropriate for complex situations where more than one source of variation is present, and can be viewed as a combination of classical test theory and an analysis of variance (ANOVA). This allows it to estimate multiple variance components behind the score variation and to compare these against each other in terms of their contribution to the overall variance [29].

The responsiveness of the SarQoL questionnaire is the least documented of its measurement properties. However, it is a key aspect to consider when using the SarQoL questionnaire to evaluate quality of life over time, and particularly important in interventional clinical trials. Clinical practitioners and health authorities rely on these kinds of trials to formulate treatment recommendations and develop health

programs, which can touch the lives of a great number of patients, so it is important that the longitudinal results obtained with the SarQoL questionnaire are precise in the way they detect a change in the quality of life of an individual patient. We performed 2 retrospective studies, one with sarcopenia as the main focus and one with physical frailty, in which the responsiveness of the questionnaire was evaluated, but the lack of an intervention causing a change in health status limits the evidence these studies can provide. However, one of the advantages of the design of the two studies we performed is that we were also able to look at the ability of the SF-36 and EQ-5D questionnaires to detect change under the exact same circumstances. It is encouraging to contrast the standardized response means of the SarQoL questionnaire with those of the SF-36 and EQ-5D and see that the SarQoL questionnaire reported larger effect sizes, indicating greater change. Given the positioning of the SarQoL questionnaire as an instrument to measure quality of life in interventional clinical studies, it is important that this measurement property of the SarQoL questionnaire is further documented in the future. There are several ways this could be envisaged: by using a global rating scale (often a question on how much the patients themselves feel they have changed on a Likert scale of 5 points), by linking the changes measured by the SarQoL questionnaire to a gold standard for change (for example, grip strength or gait speed) although the choice for a gold standard in this design would not be straightforward, and by re-evaluating the hypotheses used in the existing responsiveness studies in other studies [21]. For the 2 responsiveness studies carried out with the SarQoL questionnaire so far, we relied on the slow but progressive nature of sarcopenia to effect a change in health status, but if an opportunity presents itself to verify the responsiveness of the SarQoL questionnaire in an interventional setting this should be taken with both hands.

Some of the characteristics associated with measurement properties within classical test theory, both positive and negative, can be observed in the SarQoL questionnaire. Like other questionnaires developed with CTT in mind, the SarQoL questionnaire is relatively long at 55 items, which helps it to achieve a higher level of internal consistency and precision but is associated with response burden. The questionnaire also has an issue with external factors distorting the true score for some items, for example, the presentation of question 1, which has a common stem for 6 items (“Do you currently feel a reduction in”). This can lead to artificially large correlations between items, where these are not necessarily logical, for example between feeling a reduction in your energy and feeling a reduction in your muscle flexibility.

A second type of measurement theory has been steadily gaining ground in the domain of health-related measurement in the last 2 decades. Although Item Response Theory (IRT) has been around since the 1950’s, for a long time its application and development primarily took place in the domains of psychology and education, and its adoption in health measurement is still ongoing. In item response

theory, for unidimensional models, the latent trait is considered to be a scale upon which the ability of an individual (θ) and the difficulty of an item (b) can be placed. Item response theory then characterizes the probability that a respondent with a specific latent ability θ will respond in a certain manner [21]. A major difference between classical test theory and item response theory is that the former focuses on the quality of scales and domains, while the latter focuses on the quality of individual items within the instrument. There are also differences in the scoring of an instrument, where a questionnaire using classical test theory principles uses techniques such as sum scoring to quantify the latent construct for a particular respondent, item response theory estimates the level of the latent trait (θ) through probability functions [21]. Item response theory is particularly well suited to large item banks and computerized adaptive testing, but unfortunately such measures are time-consuming and expensive to develop.

Item response theory analysis was first used to evaluate the functioning of the SF-SarQoL questionnaire within this dissertation. These results documented the basic characteristics of the SF-SarQoL, i.e., model fit, discrimination and difficulty parameters, but do not yet utilize the full power of item response theory. The analyses presented in this dissertation provide a first glance and can hopefully be expanded in the future. There is also now an opportunity to evaluate the original SarQoL questionnaire using item response theory techniques, because sufficient data has been collected from international sources over the last few years to obtain a sufficiently large sample. While we have not yet been able to demonstrate that the 55-item version of the SarQoL questionnaire fulfils the preconditions of unidimensionality, local independence and monotonicity, an analysis on the domain-level is likely possible and appropriate. This approach could provide us with information on the functioning of the items within the 7 domains and could help in optimizing the current 55-item questionnaire by eliminating items with low discrimination parameters or some of the items with nearly identical discrimination parameters, provided that the content coverage of the questionnaire is not unduly impacted. Reliability parameters and measurement error can also be investigated through item response theory, which has the advantage of being able to provide the standard error (which is comparable to the standard error of measurement) for each individual item as well as over the scale as a whole. Finally, differential item functioning (for example, between men and women) could be investigated and taken into account [21].

Finally, it is important to recognize that the evaluation of measurement properties does not have a fixed endpoint. Multiple articles in this dissertation declare that the SarQoL questionnaire is a valid and reliable instrument, but this statement should be nuanced. The demonstrated reliability and validity are linked to the specific context in which they are investigated, including but not limited to language, culture, administration form, completion of the questionnaire at home or in a research centre, and the specific characteristics of the population themselves. While a certain amount of generalization from one context to another is necessary, it is also important not to take the measurement properties of the

questionnaire for granted. With the SarQoL questionnaire, a pattern of results is emerging pointing to its qualities as a measurement instrument. A formal analysis in the form of a systematic review and meta-analysis on the measurement properties of the SarQoL questionnaire could highlight its strengths and weaknesses.

Aside from its measurement properties, we also explored whether the SarQoL questionnaire could be used to screen for sarcopenia. Clinical studies always struggle with the balance between collecting as much information as possible and the burden put on the participants by the number of questionnaires and tests they are asked to complete and undergo. Having multiple uses associated with the SarQol questionnaire could help its adoption in clinical studies and potentially reduce this burden. We recognize that this is an unusual application for a quality-of-life questionnaire, and that specific circumstances are required to make using the SarQoL questionnaire to screen potentially sarcopenic people a viable option. This investigation came about after an inquiry from one of the co-authors, who wished to know whether she could screen people for inclusion in a clinical trial with the SarQoL questionnaire. At the time, we had not looked into this question, and could not provide her with a satisfactory answer. This prompted us to explore this application of the SarQoL questionnaire in a previously collected dataset. The use case we envisaged was that of a study in which the SarQoL questionnaire is administered to all participants, and where the results are analyzed to see which participants are more likely to be sarcopenic. These participants can then be invited for physical tests, hopefully reducing the number of people who are invited for body composition analysis but who turn out not to be sarcopenic. What we found was that the AUC value demonstrated that the SarQoL questionnaire was useful in screening for sarcopenia, and, to our surprise, that its screening accuracy was on par with the SARC-F questionnaire when used in conjunction with the revised EWGSOP2 criteria. While we do not necessarily expect that the SarQoL questionnaire will be used often for screening purposes, this study does open the door for a number of interesting possibilities. First off, if the full-length SarQoL questionnaire has a use as a screening tool, it is worthwhile to investigate whether the SF-SarQoL is equally useful. Given the easier administration and lower burden of the short form questionnaire, if it turns out to possess acceptable screening characteristics, it may be a viable option in an expanded range of circumstances. Secondly, we can approach these results from a different perspective, and interpret the threshold value found for the SarQoL questionnaire as an indicator for a situation where an intervention to improve the health status and quality of life of a patient would be justified. This might be particularly interesting for clinical practice or as a part of a comprehensive geriatric assessment.

While the relative length of the SarQoL questionnaire allows it to be comprehensive and helps it to be more precise, it places a burden on respondents and forces researchers to consider the cost-benefit

balance of including the questionnaire in their studies. The response burden is a function of its length, at 55 items, and the fact that making considered choices is a cognitively demanding task. What is more, it is not very often that the questionnaire is administered without any other tests or questionnaires before or after, so the ability of the respondent to stay concentrated and motivated may already be impacted. Researchers carefully consider what outcomes to study, and which instruments to use, during the design of the study protocol. Oftentimes, tough choices need to be made and not every instrument can be included. Quality of life is mostly included as a secondary outcome in sarcopenia research, and researchers often choose legacy instruments such as the SF-36 or the EQ-5D, prioritizing generalizability over precision [30, 31]. In creating a shorter version of the SarQoL questionnaire, we hoped to improve the acceptability of the SarQoL questionnaire through the following: (1) by reducing the cognitive burden on the respondent, (2) by reducing the complexity of the scoring procedure, (3) by reducing the probability of missing responses, (4) and by reducing the relative cost in time and effort of administering the questionnaire. A shorter version would also be more suitable to studies that want to include a generic instrument for generalizability purposes, and a specific questionnaire for measuring sarcopenia-related quality of life. The development of the SF-SarQoL also afforded us the opportunity to remove chained questions and questions with poor response options, which have complicated analysis of quality-of-life data and measurement properties previously.

While the reasons mentioned above are valid reasons to reduce the length of a questionnaire, there are other, equally valid reasons, that were not put forward as objectives of the item reduction process in this dissertation. Most notably, we did not make improving one or more measurement properties a primary objective, while this would have been a valid reason for modifying the SarQoL questionnaire.

At the start of this dissertation project, the search for a robust methodology for reducing the number of items in a health measurement questionnaire was one of our top priorities. Perhaps not surprisingly, there is no single, ready-made method of accomplishing this outcome, but rather a set of guidelines and principles to take into account. We based our approach on 3 guideline documents on item reduction for health status questionnaires. The first one, written by Joël Coste et al. and published in 1997, reviewed 49 articles describing the shortening of composite measurement scales and formulated recommendations based on their review [32]. Important to note is their finding that, although the large majority of the reviewed articles (64%) relied on a statistical approach alone in the selection of items for a short form questionnaire, this approach should only be used if the original instrument is considered a gold standard [32]. Three years later, Gregory Smith et al. published the second guideline document, describing 11 capital sins of short-form development, and elaborating a number of steps to be performed to assure rigorousness during the development of the short form itself, and during its validation afterwards [33]. The third guidance document, published in 2013 by Goetz et al., reviewed 103 articles reporting on the development of a short form measuring a health or psychological construct. The authors commented on the absence of a rigorous methodology apparent after review of the articles, and

formulated 6 recommendations that would, if adopted, greatly improve the development and validation of short form instruments [34]. For our own project of shortening the SarQoL questionnaire, we tried to follow and incorporate these guidelines as much as possible. We adopted this stance because this approach balances the expert-based design, where the content is prioritized, with a statistical approach, where statistical properties dictate the inclusion of items. In this way, the SF-SarQoL benefits from the strengths of both approaches while limiting the risks.

Below, we describe how we integrated each of the recommendation of the most recent guidance document by Goetz et al. into the development and validation of the SarQoL questionnaire.

- **Document the validity of the original composite measurement scale and the objective of its shortening.**

  At the time when the item reduction process was gathering pace, there were 6 validation studies (of the French, English, Dutch, Polish, Romanian, and Greek translations) available, and we further contributed an analysis of the standard error of measurement and smallest detectable change, and an evaluation of the responsiveness of the questionnaire. We also had the advantage of having access to multiple datasets gathered by other teams that had translated and validated the SarQoL questionnaire, for a total of 1,950 observations, which allowed us to explore the factor structure and differential item functioning. We formulated our objectives for the item reduction at the start of the project, namely (1) reducing the questionnaire by more than two-thirds, (2) conserving, in as much as possible, the 7 domains of health-related quality of life, and (3) obtaining similar measurement properties for the short form as the original questionnaire.

- **Take the conceptual model into account.**

  During the shortening process, it became clear that domain 7, which captures fear related to a limitation in movement due to muscle weakness, posed difficulties from a conceptual level. This domain consists of 4 items (fear of pain, fear of not being able to, fear of tiredness after activities, and fear of falling) and is chained to the previous question (on limiting movement due to muscle weakness) through an “if yes, then …” construction. Furthermore, the responses for these 4 items capture only a “yes” response, rendering us unable to distinguish between missing responses and “no” answers. For these reasons, we encountered problems establishing measurement properties for these items, and we would not be able to select one or more of these questions without incorporating the preceding question, which belonged to a different domain. Ultimately, the expert group decide to remove this domain from the short form. It might be that this domain could still have been incorporated by reformulating the items and changing the

response option, but this would have raised issues of consistency between the original formulation and whatever formulation would have been adopted in a short form.

- **Preserve content validity**

  We gathered information on the content of the SarQoL questionnaire from two different populations: healthcare professionals/researchers and older people themselves. The healthcare professionals and researchers provided their opinion on the importance of the 55 individual items of the SarQoL questionnaire through a two-round Delphi method. To gather information on the perspective of the target population on the importance of the different items in the SarQoL questionnaire we calculated item-impact scores, which allowed us to use a large database and obtain a ranking of the items. However, in an ideal situation, more could have been done to gather data from the patient perspective. While a best-worst scaling study, such as the one we performed with the 14 items of the SF-SarQoL, would probably not have been feasible with all 55 items of the SarQoL questionnaire, it is a very interesting technique and, could perhaps serve well for other questionnaires. Taking a general view on the content validity of the SF-SarQoL, it is difficult to judge how much of the content was lost in the reduction process. One of the strengths of the SarQoL questionnaire is its comprehensiveness, and its length allowed it to poll the respondents on a wide number of aspects. This is no longer the case for the SF-SarQoL, which is only 25% of the length of the original questionnaire, and consequently it represents a much less comprehensive assessment of sarcopenia-related quality of life.

- **Preserve psychometric properties.**

  The 14-item SF-SarQoL was validated in an independent sample and performed adequately. In terms of classical test theory analyses, we performed all evaluations that were previously performed on the original SarQoL questionnaire and did not find discrepancies between the two except for the presence of a slight systematic bias in the test-retest reliability of the SF-SarQoL, something we had not encountered in the original questionnaire. It is notable that we were not able to confirm a unidimensional model during the factor analysis for the SF-SarQoL, but instead found a 2-factor model to be better adapted.

- **Document the reasons for item selection**

  Throughout the item reduction process, we produced reports on the various meetings of the expert committee, and the information that was presented to them. We also presented this information in table form in the article included in this dissertation.

- **Validate the short form in an independent sample**

  We administered the SF-SarQoL after its development to a sample of previous participants of the SarcoPhAge cohort study, ensuring that the respondents' responses were not influenced by the answers of the other items included in the original questionnaire. We did, however, ask the respondents to complete the SF-SarQoL and the original questionnaire, presenting them with a study packet containing first the SF-SarQoL, then the EQ-5D and EQ-VAS questionnaire, followed by a best-worst scaling survey and finally the original SarQoL questionnaire. Goetz et al. recommend against this practice, because of the possibility of a halo effect. In auto-administered questionnaires, this effect is born from the fact that previous questions shape the context in which a question is answered, and responses may thus be shaped by how respondents answered previous questions [35]. We tried to limit the influence from the administration of the SF-SarQoL on the completion of the full-length SarQoL questionnaire by placing multiple questionnaires in between the two and by giving the respondents the explicit instruction not to look up their responses on the SF-SarQoL.

Over the last years, evidence on the measurement properties of the SarQoL questionnaire has steadily increased, and the research activities presented in this dissertation have contributed to this body of work. Other researchers from a multitude of countries have also invested time and effort in the SarQoL project though translations and validations. Certain investigations in this dissertation have only been possible because of the collaboration and help from these researchers, who contributed their perspective or shared data. Slowly but surely, we are seeing this collaborative effort start to bear fruit in that researchers are beginning to use the SarQoL questionnaire to measure quality of life in sarcopenia and reporting results in peer-review publications, while other studies should report results soon. Notably, several interventional clinical trials are underway that have incorporated the SarQoL questionnaire as a secondary outcome, and we eagerly await the results of these. We are also seeing some novel applications such as using the questionnaire to assess the impact of low muscle strength on quality of life in a sample of women diagnosed with systemic lupus erythematosus, or axial and peripheral spondyloarthropathy [36, 37]. However, the SarQoL has not been validated for use in these conditions, so the results obtained in these two studies should be interpreted with caution. This dissertation itself has investigated the application of the SarQoL questionnaire in frailty and validated its use in this population, opening the door for its use in studies investigating physical frailty diagnosed with the Fried criteria.

From its inception, it has been hoped that the SarQoL questionnaire would be of value and adopted by a wide range of user profiles: clinical practitioners, academic researchers, and pharmaceutical industry to name but a few. The results reported in this dissertation on the measurement properties of the SarQoL

questionnaire, and the creation of a shorter version, hopefully prove to be beneficial for each of these groups, whether their interest is in using the SarQoL as a signalling instrument, a screening instrument, to follow patients over time or to assess the effect of interventions. While the questionnaire cannot be everything to everyone, we have shown that it is a well-rounded tool and that it is not limited to a single type of application.

Both the European Medicines Agency and the Federal Drug Administration have stressed the need to incorporate patient perspectives in all aspects of trial design and outcome assessment, and we believe that, by documenting its measurement properties, and by creating a shorter version as well as establishing a ranking on importance of the 14 aspects of quality of life, we have contributed to the adoption of the SarQoL questionnaire in clinical trials. Because of the age and comorbidities generally associated with the target population of interventional trials aiming to treat sarcopenia, having information on how the patient experiences an intervention is essential information [29, 30]. Even if certain clinical indicators such as muscle mass do not improve, an intervention may still be beneficial if it improves the patients' quality of life. The importance of quality of life as an outcome in sarcopenia interventions has been demonstrated previously by Hiligsmann et al (2020), who showed that sarcopenic patients valued mobility, managing domestic activities, fall risk, fatigue, and quality of life, in that order, as outcome priorities for sarcopenia [38]. Beaudart et al looked at the same question in a group of experts in 2021 and found that they considered falls to be the most important outcome, followed by domestic activities, mobility, quality of life and fatigue [39].

Within the framework of this dissertation, we were the first to publish a ranking of the importance of a number of aspects of quality of life, determined by older people themselves. This opens up opportunities in terms of item-based evaluation of the evolution of quality of life and the possibility of a weighted score for the SF-SarQoL. The results from the best-worst scaling survey could also be employed to develop utility values, which would allow the calculation of quality-adjusted life years and thus facilitate economic analyses on the impact of sarcopenia. A methodology integrating the results from a best-worst scaling survey and 2 lead time trade-off experiments into an algorithm which allows a quality-of-life score to be converted to a utility value has been described by Essers et al. [40]. In the case of the SF-SarQoL, an additional study would have to be performed to obtain a preference valuation of a health state of "severe sarcopenia" and a state of "normal health", which could then be combined with the existing BWS results. It might also be valuable by introducing this type of study, the best-worst scaling survey, into the domain of sarcopenia and promoting its use to solicit patient opinion on several topics relevant to sarcopenia. A note of caution should be added to the results from the BWS study we performed, namely that there is a risk of bias related to the self-selection of the participants. The sample used in this study is the same as for the validation study of the SF-SarQoL, where 314 people agreed to participate and completed at least one of the questionnaires that they had received by mail. From these 314 people, only 163 (52%) sent back the BWS survey, with some not completing it at all, and others

misinterpreting the choice tasks and rendering the survey unusable by providing multiple answers. This self-selection may have reduced the representativeness of the sample. It may also be that there is a non-response bias, and that those 151 participants who did not complete, or completed incorrectly, the BWS survey would have rated the importance of the 14 aspects of quality of life differently.

Not surprisingly, there are still gaps in our knowledge with regards to the measurement properties of the SarQoL questionnaire and quality of life in sarcopenia. One of these gaps is the fact that there is not yet a value for what is considered to be a significant change in quality of life by the target population itself, often named minimal important change (MIC) or minimal clinically important difference (MCID). While we have calculated a value for what constitutes statistically significant difference in scores (the smallest detectable change of the questionnaire), calculating the minimal clinically important change necessitates a specific type of study, with a pre- and post-intervention design, and an anchor variable. Regulatory agencies, in particular the American Food and Drugs Administration (FDA) and the European Medicines Agency (EMA), place more and more emphasis on a demonstration of meaningful within-patient change for their decision-making process concerning medical products. Therefore, obtaining a value for a meaningful change in quality of life according to patients themselves is an important priority for the future of the SarQoL project. The FDA has provided guidance on this topic within the context of clinical trials and recommends the use of anchor-based methods to determine a meaningful change. Because the minimal important change is population-specific, it is advisable to design a clinical trial incorporating anchor variables in parallel with the outcome instrument in question, so that the meaningful change for an individual can be determined over the same time period in the same population [41].

For the interpretation of the scores collected by the SarQoL questionnaire in a transversal context, it would help researchers to have population norms so they can situate the scores of an individual respondent against the distribution on scores in a larger (representative) population. The responsiveness of the SarQoL questionnaire was investigated in a longitudinal cohort but should be confirmed in an interventional setting where a change in health status can be monitored, quantified, and controlled to a greater extent than in a longitudinal cohort study. To further stimulate and facilitate the adoption of the questionnaire, its behaviour in an interviewer-administered setting and as a computer-administrated questionnaire should be investigated. The story of the SF-SarQoL is only just beginning, and there are still plenty opportunities to learn more about its characteristic, in particular around its predictive value related to outcomes such as mortality, falls, loss of independence and others. Its utility as a screening instrument could also be investigated. The validation analyses presented in this dissertation should also be repeated in a larger, preferably multinational, sample of respondents. This would also allow the

evaluation of differential item functioning, for example between men and women, or between different nationalities.

While this dissertation is technical in nature, its implications on clinical practice are important. By documenting the measurement properties of the SarQoL questionnaire and by creating the SF-SarQoL, we can hope to have contributed to a better understanding of treatment effects in clinical trials in the future, and the construction of patient profiles that may guide the choice for treatment or other intervention for individual patients. In daily practice, the questionnaire could help to improve patient-doctor relations and provide a clearer picture of patients' global health and wellbeing as they perceive it themselves. It can also help to identify areas of concern and provide an opening to engage with the patient on aspects of their life that they may not bring up spontaneously. Because of the presence of multiple domains in the SarQoL questionnaire, priority areas can be defined, and progress in these specific areas can be quantified [42]. Of course, we do recognize that the use of quality-of-life questionnaires in clinical practice is not widespread, and that there are barriers that render the uptake more difficult. However, certainly with older patients and with chronic conditions, quality of life and general wellbeing may be equal to or trump any physical parameters in judging treatment success, and a wider use of this parameter in clinical practice could prove to be beneficial to the patient.

Over the past few years, information on the measurement properties of the SarQoL questionnaire has steadily increased, and this dissertation has done its bit to inspire users to be confident in using the instrument. We have also written a new chapter in the history of the SarQoL project by creating a shorter version, which we hope will be easier to integrate into clinical studies and other applications. Finally, we are proud to have worked on a tool designed to capture patients' lived experience, and in doing so giving a voice to their wants and needs. We are convinced that the SarQoL questionnaire and its short form are a valuable tool for measuring sarcopenia-related quality of life and hope to see many more results reported over the next few years.

References

1. Thomson W. (1889) Electrical units of measurement. In: Popular Lectures Vol. 1. p 73
2. Beaudart C., Biver E., Reginster J-Y., Rizzoli R., Rolland Y., Bautmans I., et al. (2017) Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 8:(2):238-244
3. Beaudart C., Edwards M., Moss C., Reginster JY., Moon R., Parsons C., et al. (2017) English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. Age Ageing 46:(2):271–277
4. Ildiko GA., Gabriela M., Charlotte B., Olivier B., Raluca-Monica P., Jean-Yves R., et al. (2017) Psychometric performance of the Romanian version of the SarQoL®, a health-related quality of life questionnaire for sarcopenia. Arch Osteoporos 12:(1):103
5. Geerinck A., Bruyère O., Locquet M., Reginster JY., Beaudart C. (2018) Evaluation of the Responsiveness of the SarQoL® Questionnaire, a Patient-Reported Outcome Measure Specific to Sarcopenia. Adv Ther 35:(11):1842–1858
6. Geerinck A., Scheppers A., Beaudart C., Bruyère O., Vandenbussche W., Bautmans R., et al. (2018) Translation and validation of the Dutch SarQoL® , a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact 18:(4):463–472
7. Konstantynowicz J., Abramowicz P., Glinkowski W., Taranta E., Marcinowicz L., Dymitrowicz M., et al. (2018) Polish Validation of the SarQoL®, a Quality of Life Questionnaire Specific to Sarcopenia. J Clin Med 7:(10):323
8. Tsekoura M., Billis E., Gliatis J., Tsepis E., Matzaroglou C., Sakkas GK., et al. (2018) Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. Disabil Rehabil 42:(7):1006-1012
9. Alekna V., Kilaite J., Tamulaitiene M., Geerinck A., Mastaviciute A., Bruyère O., et al. (2019) Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). Eur Geriatr Med 10:761-767
10. Geerinck A., Alekna V., Beaudart C., Bautmans I., Cooper C., De Souza Orlandi F., et al. (2019) Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (Sarqol) questionnaire: An analysis of subjects from 9 validation studies. PLoS One 14:(4):e0216065

11. Safonova YA., Lesnyak OM., Baranova IA., Suleimanova AK., Zotkin EG. (2019) Russian translation and validation of SarQoL® - quality of life questionnaire for patients with sarcopenia. Nauchno-Prakticheskaya Revmatol 57:(1):38–45

12. Dzhus M., Dzhus M., Masnyi M., Kulyk M., Mostbauer H., Ivashkivsky O., et al. (2020) Cross-sectional evaluation of the sarcopenia quality of life (Sarqol) questionnaire: Translation and validation of its psychometric properties. Ann Geriatr Med Res 24:(2):139–147

13. Fábrega-Cuadros R., Martínez-Amat A., Cruz-Díaz D., Aibar-Almazán A., Hita-Contreras F. (2020) Psychometric Properties of the Spanish Version of the Sarcopenia and Quality of Life, a Quality of Life Questionnaire Specific for Sarcopenia. Calcif Tissue Int 106:(3):274–282

14. Matijević R., Hrnjaković O., Đurđević A., Geerinck A., Beaudart C., Bruyère O., et al. (2020) Translation and psychometric performance of the Serbian version of the sarcopenia quality of life (SarQoL®) questionnaire. Srp Arh Celok Lek 148:(11–12):742–748

15. Erdogan T., Eris S., Avci S., Oren MM., Kucukdagli P., Kilic C., et al. (2021) Sarcopenia quality-of-life questionnaire (SarQoL)®: translation, cross-cultural adaptation and validation in Turkish. Aging Clin Exp Res (epub ahead of print)

16. Geerinck A., Locquet M., Reginster J-Y., Bruyère O., Beaudart C. (2020) Discriminative power of the Sarcopenia Quality of Life (SarQoL) questionnaire with the EWGSOP2 criteria. J Frailty Aging 10:193–194

17. Yoo J Il., Ha YC., Kim M., Seo SH., Kim MJ., Lee GY., et al. (2021) Translation and validation of the Korean version of the Sarcopenia Quality of Life (SarQoL-K®) questionnaire and applicability with the SARC-F screening tool. Qual Life Res 30:(2):603–611

18. Geerinck A., Locquet M., Bruyère O., Reginster JY., Beaudart C. (2021) Evaluating quality of life in frailty: applicability and clinimetric properties of the SarQoL® questionnaire. J Cachexia Sarcopenia Muscle 12:(2):319-330

19. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research., Center for Biologics Evaluation and Research., Center for Devices and Radiological Health. (2009) Guidance for Industry Use in Medical Product Development to Support Labeling Claims Guidance for Industry. https://www.fda.gov/media/77832/download. Accessed 2 May 2021

20. Fayers PM., Manchin D. (2016) Quality of life: the assessment, analysis and reporting of patient-reported outcomes (3rd edition). Wiley-Blackwell

21. de Vet HCW., Terwee CB., Mokkink LB. (2011) Measurement in Medicine: A Practical Guide. Cambridge University Press

22. Jarvis CB., Mackenzie SB., Podsakoff PM. (2013) A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. J Consum Res 30:(2):199–218

23. De Vries J., Michielsen H., Van Heck GL., Drent M. (2004) Measuring fatigue in sarcoidosis : The Fatigue Assessment Scale (FAS). Br J Health Psychol 9:279–291

24. Fayers P., Manchin D. (2013) Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes (2nd edition). Wiley

25. Lord FM., Novick MR. (1968) Statistical Theories of Mental Test Scores. Addison-Wesley

26. DeVellis RF. (2006) Classical test theory. Med Care 44:(11 SUPPL. 3):50–59

27. Mokkink LB., Terwee CB., Patrick DL., Alonso J., Stratford PW., Knol DL., et al. (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 63:(7):737–745

28. Dunn TJ., Baguley T., Brunsden V. (2014) From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 105:(3):399–412

29. Christophersen K.-N., Helseth S., Lund T. (2008) A Generalizability Study of the Norwegian Version of KINDL^{R} in a Sample of Healthy Adolescents. Qual Life Res 17:(1):87-93

30. Reginster JY., Cooper C., Rizzoli R., Kanis JA., Appelboom G., Bautmans I., et al. (2016) Recommendations for the conduct of clinical trials for drugs to treat or prevent sarcopenia. Aging Clin. Exp. Res. 28:47–58

31. Reginster JY., Beaudart C., Al-Daghri N., Avouac B., Bauer J., Bere N., et al. (2020) Update on the ESCEO recommendation for the conduct of clinical trials for drugs aiming at the treatment of sarcopenia in older adults. Aging Clin Exp Res 33:(1):3-17

32. Coste J., Guillemin F., Pouchot J., Fermanian J. (1997) Methodological approaches to shortening composite measurement scales. J Clin Epidemiol 50:(3):247–252

33. Smith GT., Mccarthy DM., Anderson KG. (2000) On the sins of short form development. Psychol Assess 12:(1):102–111

34. Goetz C., Coste J., Lemetayer F., Rat AC., Montel S., Recchia S., et al. (2013) Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. J Clin Epidemiol 66:(7):710–718

35. Bogner C., Landrock K. (2016) Response Biases in Standardised Surveys. GESIS Surv Guidel. https://www.gesis.org/fileadmin/upload/SDMwiki/BognerLandrock_Response_Biases_in_Standardised_Surveys.pdf. Accessed 2 May 2021

36. Sumantri S., Rengganis I., Laksmi PW., Hidayat R., Koesnoe S., Shatri H. (2021) The impact of low muscle function on health-related quality of life in Indonesian women with systemic lupus erythematosus. Lupus 30:(4):680–686

37. Sarcopenia in Axial and Peripheral Spondyloarthropathies. https://clinicaltrials.gov/ct2/show/NCT03319264. Accessed 2 May 2021

38. Hiligsmann M., Beaudart C., Bruyère O., Biver E., Bauer J., Cruz-Jentoft AJ., et al. (2020) Outcome Priorities for Older Persons With Sarcopenia. J Am Med Dir Assoc 21:(2):267-271.e2

39. Beaudart C., Bruyère O., Cruz-Jentoft AJ., Vaquero-Pinto MN., Locquet M., Bauer J., et al. (2020) Patient's Engagement in the Identification of Critical Outcomes in Sarcopenia. J Am Med Dir Assoc 21:(2):284–286

40. Essers I., Hiligsmann M., Kiltz U., Bansback N., Braun J., van der Heijde D., et al. (2019) Development of one general and six country-specific algorithms to assess societal health utilities based on ASAS HI. RMD Open 5:e000872

41. U.S. Food and Drug Administration (2019) Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making. https://www.fda.gov/media/132505/download. Accessed 2 August 2021

42. Jacobson P., Davis K., Cella D. (2002) Assessing Quality of Life in Research and Clinical Practice. Oncology 16:(9):133-139

List of abbreviations

95% CI	95% confidence interval
ADL	Activities of daily living
ALM	Appendicular lean mass
AUC	Area under the curve
AWGS	Asian working group on sarcopenia
BIA	Bio-electrical impedance analysis
BMI	Body mass index
BWS	Best-worst scaling
CASP-16	Control, autonomy, self-realization, and pleasure scale
CFA	Confirmatory factor analysis
CFI	Comparative fit index
COSMIN	Consensus-based standards for the selection of health Measurement instrument
CST	Chair stand test
CTT	Classical test theory
DCE	Discrete choice experiment
DXA	Dual-energy x-ray absorptiometry
EFA	Exploratory factor analysis
EQ-5D	EuroQoL 5-dimension questionnaire
ESCEO	European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases
ESPEN-SIG	Special interest Group on Sarcopenia of the European Society of Nutrition
EUROHIS-QOL	EUROHIS quality of life measure
EWGSOP	European working group on sarcopenia in older people
EWGSOP2	2nd European working group on sarcopenia in older people
FDA	Food and drug administration
FNIH	Foundation for the national institutes of health
GRM	Graded response model
HGS	Handgrip strength
HRQoL	Health-related quality of life
Ht	Height
ICC	Intraclass correlation coefficient
IF-VIG	Frail comprehensive geriatric assessment
IGF-1	Insulin-like growth factor 1
IQR	Interquartile range
IRT	Item response theory
IWGS	International working group on sarcopenia

Kcal	Kilocalorie
KMO	Kaiser-Meyer-Olkin test
LoA	Limits of agreement
LR-	Negative likelihood ratio
LR+	Positive likelihood ration
MAP	Minimum average partial test
MCS	Mental component summary
MIC	Minimal (clinically) important change
MSRA	Mini sarcopenia risk assessment
NPV	Negative predictive value
OR	Odds ratio
PCS	Physical component summary
PPV	Positive predictive value
PROM	Patient-reported outcome measure
QoL	Quality of life
QQ-plot	Quantile-quantile plot
RIS	Relative importance score
ROC	Receiver operating characteristic
RSMEA	Root mean square error of approximation
SARC-F	Strength – assistance with walking – rise from a chair – climb stairs – falls
SarcoPhAge	Sarcopenia and quality of life with advancing age
SarQoL	Sarcopenia quality of life questionnaire
SDC	Smallest detectable change
SD_{diff}	Standard deviation of the difference between 2 scores
SDOC	Sarcopenia definition and outcomes consortium
Se	Sensitivity
SEM	Standard error of measurement
SF-36	Short-form 36 item questionnaire
SF-SarQoL	Short-form sarcopenia quality of life questionnaire
SMI	Skeletal muscle index
Sp	Specificity
SPPB	Short physical performance battery
SRM	Standardized response mean
SRMR	Standardized root mean square residual
SSCWD	Society for sarcopenia, cachexia and wasting disorders
TLI	Tucker-Lewis index
TUG	Timed-up-and-go test

VAS	Visual analogue scale
WHOQOL-BREF	World Health Organization quality of life questionnaire – abbreviated version
WHOQOL-OLD	World Health Organization quality of life questionnaire – version for older people
WSLMV	Weighted least square mean and variance adjusted estimator
α	Alpha

Annexes

1. The SarQoL questionnaire
2. The Short-Form SarQoL
3. Best-worst scaling survey

The SarQoL questionnaire

Questionnaire I Time : ±10 min

SarQoL
Sarcopenia and Quality of Life

Quality of life in sarcopenia

This questionnaire asks about sarcopenia, which is a muscle weakness that comes about with ageing. Sarcopenia can affect your daily life. This survey will enable us to find out if the state of your muscles currently affects your quality of life.

Please choose the most appropriate response for each question. The questionnaire should take you approximately 10 minutes to complete.

1. Do you currently feel you have a reduction in:

	A lot	Some	A little	None
The strength in your arms?	☐	☐	☐	☐
The strength in your legs?	☐	☐	☐	☐
Your muscle mass?	☐	☐	☐	☐
Your energy?	☐	☐	☐	☐
Your physical capabilities?	☐	☐	☐	☐
Your general flexibility?	☐	☐	☐	☐

2. Do you have pain in your muscles?

- ☐ Often
- ☐ Sometimes
- ☐ Rarely
- ☐ Never

SarQoL
Sarcopenia and Quality of Life

3. When undertaking light physical activities (walking slowly, doing the ironing, dusting, washing-up, DIY, watering the garden, etc.), do you:

	Often	Occasionally	Rarely	Never	I do not undertake these types of physical activities
Have difficulty?	☐	☐	☐	☐	☐
Get tired?	☐	☐	☐	☐	☐
Experience pain?	☐	☐	☐	☐	☐

4. When undertaking moderate physical activities (fast walking, cleaning windows, hoovering, washing the car, pulling up weeds in the garden, etc.), do you:

	Often	Occasionally	Rarely	Never	I do not undertake these types of physical activities
Have difficulty?	☐	☐	☐	☐	☐
Get tired?	☐	☐	☐	☐	☐
Experience pain?	☐	☐	☐	☐	☐

5. When undertaking intense physical activities (running, hiking, lifting heavy objects, moving furniture, digging the garden, etc.), do you:

	Often	Occasionally	Rarely	Never	I do not undertake these types of physical activities
Have difficulty?	☐	☐	☐	☐	☐
Get tired?	☐	☐	☐	☐	☐
Experience pain?	☐	☐	☐	☐	☐

6. Do you currently feel old?

- ☐ Yes, very
- ☐ Yes, somewhat
- ☐ Yes, a little
- ☐ No, not at all

SarQoL
Sarcopenia and Quality of Life

7. If yes to question 6, what gives you that impression?
(Choose as many answers as you like)

- ☐ I become unwell easily
- ☐ I take many medications
- ☐ I feel a weakness in my muscles
- ☐ I have problems with my memory
- ☐ I've had to face the death of several people close to me
- ☐ I do not have much energy, I am often tired
- ☐ My eyesight is poor
- ☐ Other:

8. Do you feel physically weak?

- ☐ Yes, completely
- ☐ Yes, somewhat
- ☐ Yes, a little
- ☐ No, not at all

9. Do you feel you are limited in:

	A lot	Some	A little	None
The length of time you can walk for?	☐	☐	☐	☐
How often you go out walking?	☐	☐	☐	☐
The distance you can walk?	☐	☐	☐	☐
The speed at which you can walk?	☐	☐	☐	☐
The length of your steps?	☐	☐	☐	☐

10. When you are walking:

	Often	Occasionally	Rarely	Never	I am unable to walk
Do you feel very tired?	☐	☐	☐	☐	☐
Do you need to sit down regularly to recover?	☐	☐	☐	☐	☐
Do you have difficulty crossing roads quickly enough?	☐	☐	☐	☐	☐
Do you have difficulties with uneven surfaces?	☐	☐	☐	☐	☐

SarQoL
Sarcopenia and Quality of Life

11. Do you have problems with your balance?

- ☐ Often
- ☐ Occasionally
- ☐ Rarely
- ☐ Never

12. How often do you fall?

- ☐ Very often
- ☐ Occasionally
- ☐ Rarely
- ☐ Never

13. Do you think that your physical appearance has changed?

- ☐ Yes, very
- ☐ Yes, somewhat
- ☐ Yes, a little
- ☐ No, not at all

14. If yes to question 13, in what way? (Choose as many answers as you like)

- ☐ Change in your weight (you've put on weight or you've lost weight)
- ☐ Appearance of wrinkles
- ☐ Loss of height
- ☐ Loss of muscle mass
- ☐ Hair loss
- ☐ Getting white or grey hair
- ☐ Other: _____

15. If yes to question 13, are you upset by this change?

- ☐ Yes, very
- ☐ Yes, somewhat
- ☐ Yes, a little
- ☐ No, not at all

SarQoL
Sarcopenia and Quality of Life

16. Do you feel frail?

- ☐ Very much so
- ☐ A little
- ☐ Not at all

17. Do you currently have difficulty in undertaking any of the following daily activities:

	Unable to do	Great difficulty	A little difficulty	No difficulty	Not applicable
Climbing a flight of stairs?	☐	☐	☐	☐	☐
Climbing several flights of stairs?	☐	☐	☐	☐	☐
Going up one or several steps without holding on to the banister?	☐	☐	☐	☐	☐
Squatting or kneeling?	☐	☐	☐	☐	☐
Stooping or leaning down to pick up an object off the floor?	☐	☐	☐	☐	☐
Getting up from the floor without holding on to anything?	☐	☐	☐	☐	☐
Getting out of a low chair without armrests?	☐	☐	☐	☐	☐
Moving, generally, from a sitting position to a standing position?	☐	☐	☐	☐	☐
Carrying heavy objects (large bags full of shopping, saucepan filled with water, etc.)?	☐	☐	☐	☐	☐
Opening a bottle or a jar?	☐	☐	☐	☐	☐
Using public transport?	☐	☐	☐	☐	☐
Getting in or out of a car?	☐	☐	☐	☐	☐
Doing your shopping?	☐	☐	☐	☐	☐
Doing the housework (making the bed, hoovering, doing the ironing, washing the dishes, etc.)?	☐	☐	☐	☐	☐

SarQoL
Sarcopenia and Quality of Life

18. Does your muscle weakness limit your movement?

- ☐ Yes, a lot
- ☐ Yes, somewhat
- ☐ Yes, a little
- ☐ No, not at all

19. If yes to question 18, for what reasons? (Choose as many answers as you like)

- ☐ Fear of pain
- ☐ Fear that you might not be able to
- ☐ Fear of feeling tired after these activities
- ☐ Fear of falling
- ☐ Other: _____

20. Does your muscle weakness limit your sex life?

- ☐ I am not sexually active
- ☐ Yes, completely
- ☐ Yes, somewhat
- ☐ Yes, a little
- ☐ No, not at all

21. How has your participation in physical activities/sport changed?

- ☐ Increased
- ☐ Decreased
- ☐ Unchanged
- ☐ I have never participated in physical activities or sports

22. How has your participation in leisure activities (going out to eat, gardening, doing DIY, shooting/fishing, senior citizens clubs, playing bridge, going for a walk, etc.) changed?

- ☐ Increased
- ☐ Decreased
- ☐ Unchanged
- ☐ I have never participated in leisure activities

The Short-Form SarQoL questionnaire

SarQoL
Sarcopenia and Quality of Life

Questionnaire | Time : ±5 min

Short Form

Quality of life in sarcopenia

This questionnaire asks about sarcopenia, which is a muscle weakness that comes about with ageing. Sarcopenia can affect your daily life. This survey will enable us to find out if the state of your muscles currently affects your quality of life.

Please choose the most appropriate response for each question. The questionnaire should take you approximately 5 minutes to complete.

1. Do you currently feel you have a reduction in:

	A lot	Some	A little	None
The strength in your arms?	☐	☐	☐	☐
The strength in your legs?	☐	☐	☐	☐
Your muscle mass?	☐	☐	☐	☐
Your physical capabilities?	☐	☐	☐	☐
The distance you can walk?	☐	☐	☐	☐

2. When undertaking moderate physical activities (fast walking, cleaning windows, hoovering, washing the car, pulling up weeds in the garden, etc.), do you:

	Often	Occasionally	Rarely	Never	I do not undertake these types of physical activities
Have difficulty?	☐	☐	☐	☐	☐
Get tired?	☐	☐	☐	☐	☐

3. Do you currently have difficulty in undertaking any of the following activities:

	Unable to do	Great difficulty	A little difficulty	No difficulty	Not applicable
Getting up from the floor without holding on to anything?	☐	☐	☐	☐	☐
Carrying heavy objects (large bags full of shopping, saucepan filled with water, etc.)?	☐	☐	☐	☐	☐

SarQoL
Sarcopenia and Quality of Life

4. When you are walking, do you feel very tired?

Often	Occasionally	Rarely	Never	I am unable to walk
☐	☐	☐	☐	☐

5. Do you feel physically weak?

Yes, completely	Yes, somewhat	Yes, a little	No, not at all
☐	☐	☐	☐

6. Do you have problems with your balance?

Often	Occasionally	Rarely	Never
☐	☐	☐	☐

7. Does your muscle weakness limit your movement?

Yes, alot	Yes, somewhat	Yes, a little	No, not at all
☐	☐	☐	☐

8. Has your participation in leisure activities (going out to eat, gardening, doing DIY, shooting/fishing, senior citizens clubs, playing bridge, going for a walk, etc.) changed?

Increased	Decreased	Unchanged	I have never participated in leisure activities
☐	☐	☐	☐

Best-Worst Scaling survey

Questionnaire meilleur-pire (version A)

Dans chacune des tables suivantes, veuillez identifier l'aspect **le plus important** par rapport à votre qualité de vie et l'aspect **le moins important**. **Veuillez n'identifier qu'un seul aspect comme étant le plus important (une seule croix à droite) et un seul aspect comme étant le moins important (une seule croix à gauche).** Chaque table devra avoir une case cochée dans la colonne de gauche et une case cochée dans la colonne de droite.

Exemple : dans la table ci-dessous, la personne a indiqué que ressentir des difficultés lors de la réalisation d'efforts physiques modérés est l'aspect le plus important par rapport à sa qualité de vie, et que de diminuer ses activités de loisir est, par contre, l'aspect le moins important par rapport à sa qualité de vie.

Le moins important	Exemple	Le plus important
	Ressentir une diminution de la force dans ses bras	
	Ressentir de la difficulté lors de la réalisation d'efforts physiques modérés	**X**
	Avoir des problèmes d'équilibre	
X	Diminuer ses activités de loisir	

Maintenant, à votre tour. Merci de compléter les 12 tables ci-dessous en indiquant, dans chacune de ces tables, l'aspect qui vous semble être **le moins important** parmi les 4 choix proposés et celui qui vous parait être **le plus important**.

Le moins important	Question 1	Le plus important
	Ressentir une diminution de ses capacités physiques	
	Ressentir de la difficulté lors de la réalisation d'efforts physiques modérés	
	Avoir des difficultés à se relever du sol sans appui	
	Ressentir une diminution de sa masse musculaire	

Le moins important	Question 2	Le plus important
	Diminuer ses activités de loisir	
	Ressentir une diminution de la force dans ses jambes	
	Ressentir de la fatigue lorsque vous marchez	
	Limiter ses mouvements, à cause de faiblesse musculaire	

Le moins important	Question 3	Le plus important
	Ressentir de la difficulté lors de la réalisation d'efforts physiques modérés	
	Diminuer ses activités de loisir	
	Avoir des problèmes d'équilibre	
	Ressentir une diminution de la force dans ses bras	

Le moins important	Question 4	Le plus important
	Ressentir une diminution de la force dans ses jambes	
	Ressentir une diminution de ses capacités physiques	
	Avoir des difficultés à porter des objets lourds	
	Ressentir une limitation de votre distance de marche	

Le moins important	Question 5	Le plus important
	Ressentir de la fatigue lorsque vous marchez	
	Ressentir une limitation de votre distance de marche	
	Diminuer ses activités de loisir	
	Ressentir de la difficulté lors de la réalisation d'efforts physiques modérés	

Le moins important	Question 6	Le plus important
	Avoir des difficultés à porter des objets lourds	
	Ressentir une diminution de sa masse musculaire	
	Ressentir de la fatigue lorsque vous marchez	
	Avoir des problèmes d'équilibre	

Le moins important	Question 7	Le plus important
	Ressentir une diminution de sa masse musculaire	
	Ressentir de la fatigue lors de la réalisation d'efforts physiques modérés	
	Se sentir faible physiquement	
	Diminuer ses activités de loisir	

Le moins important	Question 8	Le plus important
	Ressentir une diminution de la force dans ses bras	
	Ressentir de la fatigue lorsque vous marchez	
	Ressentir de la difficulté lors de la réalisation d'efforts physiques modérés	
	Se sentir faible physiquement	

Le moins important	Question 9	Le plus important
	Ressentir de la fatigue lors de la réalisation d'efforts physiques modérés	
	Ressentir une diminution de la force dans ses bras	
	Ressentir une diminution de la force dans ses jambes	
	Avoir des difficultés à porter des objets lourds	

Le moins important	Question 10	Le plus important
	Ressentir une limitation de votre distance de marche	
	Limiter ses mouvements, à cause de faiblesse musculaire	
	Ressentir une diminution de la force dans ses bras	
	Avoir des difficultés à se relever du sol sans appui	

Le moins important	Question 11	Le plus important
	Avoir des difficultés à se relever du sol sans appui	
	Avoir des problèmes d'équilibre	
	Ressentir de la fatigue lors de la réalisation d'efforts physiques modérés	
	Ressentir une diminution de la force dans ses jambes	

Le moins important	Question 12	Le plus important
	Limiter ses mouvements, à cause de faiblesse musculaire	
	Avoir des problèmes d'équilibre	
	Ressentir une diminution de ses capacités physiques	
	Se sentir faible physiquement	

Avez-vous bien indiqué <u>le meilleur ET le pire aspect</u> pour chaque question ?

Merci beaucoup !

Publications

Responsiveness of the SarQoL questionnaire

Adv Ther (2018) 35:1842–1858
https://doi.org/10.1007/s12325-018-0820-z

ORIGINAL RESEARCH

Evaluation of the Responsiveness of the SarQoL® Questionnaire, a Patient-Reported Outcome Measure Specific to Sarcopenia

Anton Geerinck · Olivier Bruyère · Médéa Locquet · Jean-Yves Reginster · Charlotte Beaudart

Received: September 12, 2018 / Published online: October 29, 2018

ABSTRACT

Introduction: The Sarcopenia Quality of Life (SarQoL®) questionnaire was developed to provide a patient-reported outcome measure specific to sarcopenia. Its psychometric properties indicate that it is a valid and reliable instrument. However, until now, its ability to detect change over time has not been examined. Therefore, the objective of this study is to evaluate the responsiveness (also known as sensitivity to change) of the SarQoL® questionnaire in a prospective, longitudinal cohort of community-dwelling, older, sarcopenic subjects.

Methods: Sarcopenic subjects from the SarcoPhAge (Sarcopenia and Physical impairment with advancing Age) study were included. Responsiveness was evaluated with nine pre-specified hypotheses on the correlation between the evolution of the SarQoL® scores after a 2-year interval and the evolution of the scores on the Short Form-36 (SF-36) and the Euroqol 5-dimension (EQ-5D) questionnaires. This technique considers responsiveness to be a form of longitudinal validity. Additionally, standardized response means were also calculated to compare the quantity of change measured by the different questionnaires.

Results: A total of 42 sarcopenic subjects were included. The median age of the sample was 72.9 (68.9–78.8) years, 59.5% were female, and the mean body mass index was 23.3 (20.4–25.7) kg/m^2. A good responsiveness was observed, as evidenced by the confirmation of eight out of nine hypotheses, well above the 75% confirmation threshold. The standardized response mean of the Overall SarQoL® score was significantly higher than those of the SF-36 Physical Component Summary ($p = 0.005$), the EQ-5D Utility Index ($p < 0.001$) and the Euroqol visual analogue scale ($p = 0.003$).

Conclusion: The first data available on the ability of the SarQoL® questionnaire to detect change over time indicates that the questionnaire has good responsiveness. This, together with the previously established psychometric properties, confirms that the SarQoL® questionnaire is a relevant instrument for the assessment of quality of life in sarcopenic populations.

Enhanced digital features To view enhanced digital features for this article go to https://doi.org/10.6084/m9.figshare.7195778.

A. Geerinck (✉) · O. Bruyère · M. Locquet · J.-Y. Reginster · C. Beaudart
Department of Public Health, Epidemiology and Health Economics, University of Liège, Liège, Belgium
e-mail: anton.geerinck@uliege.be

Keywords: Older people; Patient-reported outcome measure; Psychometrics; Quality of life; Questionnaire; Responsiveness; Sarcopenia

INTRODUCTION

Background

Sarcopenia, defined as *"a syndrome characterised by progressive and generalised loss of skeletal muscle mass and strength and with a risk of adverse outcomes such as physical disability, poor quality of life and death"* by the European Working Group on Sarcopenia in Older People (EWGSOP), is a growing public health problem [1]. It has recently been recognized as a geriatric condition with an ICD-10-CM code (M62.84) [2]. Sarcopenia has been shown to be associated with negative health outcomes, such as a higher rate of mortality and functional decline, a higher rate of falls, and a higher incidence of hospitalization [3]. Other research has shown an association between sarcopenia and depression [4]. Not much is yet known about the relationship between sarcopenia and quality of life. Although several studies have incorporated quality of life outcomes in their designs, the results are difficult to compare because of the different diagnostic criteria used to establish sarcopenia. Some studies that diagnosed sarcopenia with the EWGSOP criteria have found lower health-related quality of life (HRQoL) scores for sarcopenic subjects in select domains of the Short-Form 36-item (SF-36) questionnaire, but other studies (using other diagnostic criteria) have found no difference in SF-36 scores between sarcopenic and non-sarcopenic subjects [5].

Until recently, researchers only had generic questionnaires, such as the SF-36, available to assess quality of life in sarcopenic patients. These questionnaires are designed for use in broad populations and may thus not be sensitive enough to accurately measure quality of life in sarcopenic populations [6]. To address this problem, Beaudart et al. developed the Sarcopenia Quality of Life (SarQoL®) questionnaire in 2015 [7].

Until now, no study has evaluated the responsiveness, defined as *"the ability of an instrument to detect change over time in the construct to be measured"*, of the SarQoL® questionnaire [8]. When an instrument is used for evaluative purposes, i.e. when the aim is to detect and measure longitudinal change in subjects or populations, responsiveness is a key psychometric property [9, 10]. This situation is often present in clinical studies aimed at testing the effect of an intervention, where an accurate assessment of HRQoL before and after the intervention is an important outcome. Researchers need to have valid data on the responsiveness of the instrument they wish to use to be certain of the results they obtain.

The psychometric properties of the SarQoL® questionnaire have been evaluated in several cross-sectional studies, but until now, its ability to detect change over time (responsiveness) had not yet been examined [11–14]. This study aimed to evaluate the responsiveness of the SarQoL® questionnaire in a sample of older, community-dwelling, sarcopenic subjects from the SarcoPhAge (Sarcopenia and Physical impairment with advancing Age) cohort.

METHODS

Design

The current article describes an instrument validation study that examined data collected at the 2nd and 4th annual visit of the SarcoPhAge study, an ongoing 5-year prospective, longitudinal, observational cohort study being carried out in Liège (Belgium) [15, 16]. Participants in the SarcoPhAge study all provided written informed consent. The research protocol and its amendments were approved by the Ethics Committee of the University Teaching Hospital of Liège (no. 2012-277).

Participants

Participants from the SarcoPhAge study with valid data from the 2nd (T1) and 4th (T3) study visit (a 2-year interval) who were diagnosed as sarcopenic according to the EWGSOP criteria were included [1]. This 2-year interval was chosen because it covers the first and last available administrations of the questionnaire and because the SarcoPhAge study is an

observational study; therefore, we relied on the natural progression of sarcopenia to cause a change in health status between the two measurements. The details of this study have been reported previously [11, 15, 17, 18].

Sarcopenia was diagnosed according to the EWGSOP algorithm, which demands the presence of low muscle mass in combination with low muscle strength and/or low physical performance [1]. Muscle mass was measured by dual-energy X-ray absorptiometry (DXA) (Hologic Discovery A, USA), which was calibrated daily by scanning a spine phantom. Male subjects with a skeletal muscle mass index (SMI = appendicular lean mass/height^2) below 7.26 kg/m^2 and women with an SMI below 5.5 kg/m^2 were considered to have low muscle mass. Muscle strength was measured with a hydraulic hand dynamometer (Saehan Corporation, Korea), calibrated at the beginning of the study for 10, 40 and 90 kg. Men with a maximal handgrip strength below 30 kg and women below 20 kg were considered to have low muscle strength. Physical performance was examined with the help of the Short Physical Performance Battery (SPPB), with a value of 8 or less being considered low [15].

Participants were included in the current analysis when diagnosed as sarcopenic at T1 and/or T3 and when both SarQoL® questionnaires (T1 and T3) had less than 20% missing data for the calculation of the Overall score.

Measures

The SarQoL® Questionnaire

The SarQoL® questionnaire is a patient-reported outcome measure (PROM) specific to sarcopenia. The SarQoL® questionnaire consists of 22 questions incorporating 55 items, which fall into seven domains of HRQoL. These domains are "Physical and Mental Health", "Locomotion", "Body Composition", "Functionality", "Activities of Daily Living", "Leisure activities" and "Fears". Each domain is scored from 0 to 100, and an Overall score is calculated. The questionnaire is auto-administered and takes 10 min to complete [7]. The questionnaire is available in 16 languages and can be found on its webpage [19].

Several psychometric properties of the SarQoL® questionnaire have been examined previously. The questionnaire has demonstrated its ability to distinguish between sarcopenic and non-sarcopenic subjects (discriminative power). It has good internal consistency and construct validity, and its test–retest reliability is excellent. Furthermore, it has been demonstrated that there are no floor or ceiling effects for the Overall score [11–14].

The Short-Form 36-Item (SF-36) Questionnaire

The SF-36 is a multi-item generic health survey that uses 36 questions to measure functional health and wellbeing from the patient's perspective. It measures eight domains: "Physical Functioning", "Role limitation due to physical problems", "Bodily Pain", "General Health Perceptions", "Vitality", "Social Functioning", "Role limitations due to emotional problems" and "Mental Health", each of which provides a score between 0 and 100. Additionally, two composite scores can be calculated: the Physical Component Summary (PCS) and the Mental Component Summary (MCS) [20–22].

The EuroQol 5-Dimension 3-Level (EQ-5D-3L)

The EQ-5D-3L is a standardized measure of health status developed by the EuroQol Group in 1990. The instrument consists of two pages: the EQ-5D descriptive system, which is composed of five questions encompassing five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression); and the Visual Analogue Scale (EQ-VAS), which records the respondent's self-rated health on a vertical scale going from best (100) to worst imaginable health (0). The EQ-5D descriptive system is used to calculate an index score, which represents the utility value for current health [23, 24].

Physical Parameters

Parameters related to muscle mass, muscle strength and physical performance were collected. Apart from the SMI, we also determined appendicular lean mass (ALM) and ALM divided

by body mass index (ALM/BMI) by DXA. As mentioned previously, muscle strength was determined with a hydraulic hand dynamometer. For physical performance, the patients performed the SPPB test, which also includes the usual gait speed on a 4-m track. The subjects also performed the timed-up-and-go (TUG) test, which uses the time that a subject takes to rise from a chair, walk three metres, turn around, walk back to the chair, and sit down to determine a subject's mobility. Lastly, the chair stand test (CST) was administered as part of the SPPB. In this test, the subjects are asked to stand up from a chair and sit back down five times as fast as they can.

Methodological Approach

Hypotheses Testing

It is recommended to treat responsiveness as the longitudinal form of construct validity and to evaluate it in much the same way as the construct validity of a questionnaire [25]. Thus, we formulated hypotheses between the changes in the scores of the SarQoL® questionnaire and the changes observed for the SF-36 and the EQ-5D. AG, CB and OB were responsible for the formulation of the hypotheses, on the basis of similarity in the construct of the different domains, and previously found results for the construct validity of the questionnaire. The data used in this analysis were collected before the formulation of the hypotheses, but no statistical manipulations in relation to the evaluation of responsiveness were carried out before the final set of hypotheses was agreed upon.

The hypotheses used for the evaluation of the responsiveness, the expected strength of the correlations and the rationale for their formulation are detailed in Table 1.

We employed the criteria formulated by De Boer et al. to evaluate the results of the hypotheses testing. These state that a questionnaire has high responsiveness when less than 25% of hypotheses are refuted, moderate responsiveness when 25–50% are refuted and poor responsiveness when more than 50% are refuted [26].

Standardized Response Means (SRMs)

We also calculated SRMs for the different questionnaires, by dividing the mean difference between T1 and T3 by the standard deviation of the differences between the paired measurements [27]. The SRM reflects the magnitude of the change measured by the different questionnaires. Consequently, when greater SRMs are obtained, this is an indication of better responsiveness. To allow the use of the thresholds for responsiveness formulated by Cohen et al., which are designed for use with the effect size and which categorize an observed change, we applied the correction developed by Middel and Van Sonderen [28, 29]. After correcting the SRMs with the formula $[(\mathrm{SRM}/\sqrt{2})/\sqrt{(1-r)}$; with r = correlation between baseline and follow-up score], we categorized them as trivial when $\mathrm{SRM} < 0.20$, small when $0.20 \leq \mathrm{SRM} < 0.49$, moderate when $0.50 \leq \mathrm{SRM} < 0.79$ and large when $\mathrm{SRM} \geq 0.80$ [29].

A selection of SRMs were compared in pairs to evaluate whether they were significantly different. This was carried out using the modified jack-knife method, which uses linear regression to determine whether a significant difference exists between two SRMs [30]. For this measure, an individual SRM is first calculated for each subject by dividing their change score by the standard deviation of the change scores in the whole sample. Next, a "centred" SRM is calculated for each subject by subtracting the mean SRM score of the sample from the individual SRMs. With these variables, a linear regression is carried out with the individual SRMs of the two quality-of-life scores of interest as dependent variables and the "centred" SRM of one of the quality-of-life scores (either one will work) as the independent variable. A significant difference is demonstrated when the p value of the intercept is at most 0.05 [30, 31].

Correlations Between Physical Parameters and QoL

We investigated the relationship between the evolution of physical parameters linked to sarcopenia and the changes observed by the different questionnaires with the help of correlations. We selected the five summary/total scores available (SarQoL® Overall score, SF-

Table 1 Hypotheses for the evaluation of responsiveness

Hypotheses	Expected strength of correlation	Rationale
1. Δ SarQoL Overall score and Δ SF-36 General Health domain are correlated	$r > 0.4$	The SarQoL Overall score and the SF-36 General Health score have been shown to be correlated in the French ($r = 0.67$) and the English ($r = 0.49$) validations. These domains are similar in that they both measure a subject's general view of either their HRQoL or health. Because of the strong interaction between general health status and HRQoL, we expect a correlation of at least 0.4, despite the difference in underlying construct
2. Δ SarQoL Overall score and Δ SF-36 Vitality domain are correlated	$r > 0.3$	Here also, two different constructs are measured, but they have been shown to be correlated (FR: $r = 0.72$; ENG: $r = 0.74$). Since the underlying constructs are less similar than in hypothesis 1, and we expect the influence of a change in vitality to be less impactful than one in General Health, the expected correlation was set to at least 0.3
3. Δ SarQoL Overall score and Δ SF-36 Physical Functioning domain are correlated	$r > 0.5$	The domain Physical Functioning covers a significant portion of the content used to calculate the Overall score of the SarQoL®, although the Overall score also takes into account other aspects of HRQoL. The English validation confirmed this similarity with a correlation of 0.82, although the French validation found a smaller correlation of 0.49. Nevertheless, we expect changes on both measures to be correlated at a strength of at least 0.5
4. Δ SarQoL Overall score and Δ EQ-VAS are correlated	$r > 0.4$	The Overall score and the EQ-VAS both give a general view of the subjects' current health or HRQoL, and should thus, in theory, be correlated. We expect the difference in health as measured by the EQ-VAS to be reflected in changes in HRQoL (as evidenced by a cross-sectional correlation of $r = 0.597$) but, since they measure two different but related constructs, it was decided to fix the expected strength of this association to at least 0.4

Table 1 continued

Hypotheses	Expected strength of correlation	Rationale
5. Δ SarQoL domain 1 (Physical and Mental Health) and Δ SF-36 General Health domain are correlated	$r > 0.3$	Domain 1 of the SarQoL® questionnaire carries significant weight in the calculation of the Overall score. Since we know a correlation exists between the Overall score and the General Health domain for the construct validity (see hypothesis 1), we theorized that this same correlation should exist between Physical and Mental Health and General Health. We did expect this correlation to be weaker, although the cross-sectional correlation was $r = 0.655$, since some aspects covered in the Overall score are not represented in Physical and Mental Health. It was decided to expect a correlation of at least 0.3
6. Δ SarQoL domain 1 (Physical and Mental Health) and Δ EQ-VAS are correlated	$r > 0.3$	In the same vein as hypothesis 5, we expected changes on Physical and Mental Health to be associated with changes on the EQ-VAS, as shown by a cross-sectional correlation of $r = 0.562$. However, since a part of the content is lost when focusing on a single domain of the SarQoL®, it was decided to expect a weaker correlation than hypothesis 5, and to adopt at least 0.3 as the threshold
7. Δ SarQoL domain 2 (Locomotion) and Δ SF-36 Physical Functioning domain are correlated	$r > 0.4$	The ability to walk and the ease with which a person can walk are an important factor that influences the totality of how a person functions physically, demonstrated by a cross-sectional correlation of $r = 0.558$. While the domain Locomotion is a much narrower construct than Physical Functioning, we expect both domains to be significantly correlated at a strength of at least 0.4
8. Δ SarQoL domain 4 (Functionality) and Δ SF-36 Physical Functioning domain are correlated	$r > 0.5$	The underlying constructs of the domains Functionality and Physical Functioning are, in theory, similar, and it was therefore felt that a relatively strong correlation of at least 0.5 was to be expected, even if the cross-sectional correlation was lower at $r = 0.420$

Table 1 continued

Hypotheses	Expected strength of correlation	Rationale
9. Δ SarQoL domain 5 (Activities of Daily Living) and Δ SF-36 Physical Functioning domain are correlated	$r > 0.5$	While these two domains represent different underlying constructs, we theorized that a change in physical functioning would be equally reflected in a change in a person's Activities of Daily Living, because one is a prerequisite for the other. It was felt that we should expect a relatively strong correlation of at least 0.5 since we expected these two domains to be interwoven even if the cross-sectional correlation was lower at $r = 0.460$

Δ = change in; r = correlation

36 PCS and MCS, EQ-5D Utility Index and EQ-VAS) to represent the HRQoL of the subjects and constructed correlations with usual gait speed, handgrip strength, SPPB score, ALM, ALM/BMI, SMI, TUG test and the chair stand test. The strength of the association was judged as excellent when larger than 0.81, very good when between 0.61 and 0.80, good when between 0.41 and 0.60, acceptable when between 0.21 and 0.40 and insufficient when less than 0.20 [32].

Statistical Analysis

Data were analysed using IBM SPSS Statistics, version 24.0.0.0 for Windows (Armonk, NY: IBM Corp).

The distribution of variables was determined by examining the histogram, the quantile–quantile plot, the Shapiro–Wilk test and the difference between mean and median. Gaussian variables are reported as the mean ± standard deviation and non-Gaussian variables as median (P25–P75). Nominal variables are reported as absolute (n) and relative frequencies (%). The presence of significant differences between T1 and T3 was examined with the paired samples t test for variables with normal distribution, the Wilcoxon matched-pair signed-rank test for non-Gaussian variables and the chi-squared test for nominal variables. Pearson correlations were calculated when both groups/variables had normal distributions. Spearman correlations were calculated when this was not the case.

Change scores were calculated by subtracting the scores from T1 from those obtained at T3. For quality of life, this means that a positive change score indicates an improvement and a negative change score a decline. The calculation of the SRMs, their correction with the technique from Middel and Van Sonderen and the modified jack-knife method used to detect significant differences between SRMs have been described in the preceding paragraphs.

A post hoc power analysis was conducted on the Pearson and Spearman correlations used in the primary outcome with the G*Power software, version 3.1.9.2 [33]. This analysis computes the achieved power for a bivariate normal model with an α-error of 0.05 and a sample size of 42 subjects.

Results were considered significant at $p \leq 0.05$.

RESULTS

In total, 42 sarcopenic participants from the SarcoPhAge study fulfilled the inclusion criteria, which is a moderate sample size according to the COSMIN checklist [34]. The subjects had a median age of 73 (69–79) years at T1, and 25 out of 42 (59.5%) were women. The median number of drugs taken by the participants increased

significantly ($p = 0.001$) from 6 (5–9) at T1 to 8 (6–10) at T3, as did the proportion of subjects who fell in the year before the study visits, from 8 (19.0%) at T1 to 16 (38.1%) at T3 ($p = 0.017$). The gait speed of the participants diminished significantly from a median of 1.02 (0.80–1.21) m/s at T1 to 0.89 (0.76–1.09) m/s at T3 ($p = 0.032$). In the sample as a whole, a slight but significant reduction in handgrip strength was observed, from a median of 19.75 (18.00–28.00) kg at T1 to 19.00 (16.75–22.50) kg at T3 ($p = 0.010$). This change was attributable to the female subjects ($p = 0.030$). No significant changes between T1 and T3 were found for BMI ($p = 0.393$), number of comorbidities ($p = 0.763$), proportion of subjects who experienced a fracture in the year before the study visits ($p = 0.268$), independence in activities of daily living as measured by the Katz scale (0.942), SPPB score ($p = 0.083$), TUG test ($p = 0.081$), ALM/BMI ($p = 0.197$) and SMI ($p = 0.451$). The ALM of the whole sample diminished significantly ($p = 0.035$), but this effect was lost when the sample was divided into men ($p = 0.287$) and women ($p = 0.072$).

The three different questionnaires obtained different results for quality of life. The SarQoL® questionnaire measured a significant reduction for three domains (Body Composition, $p = 0.023$; Functionality, $p = 0.002$; Activities of Daily Living, $p < 0.001$) and the Overall score, which diminished from a median of 61.15 (51.15–71.76) at T1 to 54.56 (42.31–68.44) at T3 ($p = 0.002$). The SF-36 PCS and MCS, the EQ-5D Utility Index and the EQ-VAS, however, did not detect a significant change (respectively, $p = 0.679$, $p = 0.062$, $p = 0.231$ and $p = 0.716$). The complete clinical characteristics and the evolution of quality of life can be found in Table 2.

Responsiveness

Of the nine formulated hypotheses, 8 (89%) were confirmed. Hypothesis 9 was rejected when a correlation of $r = 0.467$ was found, just under the threshold of $r > 0.5$. In total, three very good correlations were found, five good correlations and two acceptable correlations. The results of this evaluation as well as of the power analysis are reported in Table 3.

According to the criteria by De Boer et al., the SarQoL® questionnaire possesses high responsiveness because fewer than 25% of hypotheses are refuted [26].

Standardized Response Means

The magnitude of change observed in the sample was examined by calculating SRMs. The SarQoL® questionnaire had three domains with SRMs below 0.20, indicating that no change was observed, two domains with an SRM between 0.20 and 0.49 (small change) and three domains with an SRM between 0.50 and 0.79 (moderate change). In contrast, only one domain of the SF-36 had a moderate SRM (Physical Functioning; SRM = − 0.50), and six domains reported an SRM indicating small change. A further three domains of the SF-36 had SRMs indicating no change had occurred. For the EQ-5D, small SRMs were observed for two domains, with the remaining five domains having SRMs indicating no change. All obtained SRMs can be found in Table 4.

The SRM of the SarQoL® Overall score was significantly larger than the SF-36 PCS ($p = 0.005$), the EQ-5D Utility Index ($p < 0.001$) and the EQ-VAS ($p = 0.003$). The SRMs of the SarQoL® Overall score and the SF-36 MCS were not significantly different ($p = 0.150$). The results of this analysis are reported in Table 5.

Correlations Between Physical Parameters and QoL

Good correlations were found between change in the SarQoL® Overall score and change in gait speed ($r = 0.50$), SPPB score ($r = 0.47$) and the chair stand test ($r = - 0.42$). Good correlations were also found between change in ALM/BMI and change on the EQ-VAS ($r = - 0.48$) as well as between change on the timed up-and-go test and change on the SF-36 PCS ($r = - 0.44$). Acceptable correlations were found between change in gait speed and change on the SF-36 PCS ($r = 0.39$), between change on the chair stand test and change on the SF-36 PCS

Table 2 Clinical characteristics and quality of life scores for sarcopenic sample (n = 42)

	T1	T3	Change	*P* value
Age (years)	72.90 (68.85–78.81)	NA	NA	NA
Gender				
Male	17 (40.5%)	NA	NA	NA
Female	25 (59.5%)	NA	NA	NA
BMI (kg/m^2)	23.25 (20.35; 25.68)	23.09 (20.06; 25.84)	− 0.03 (− 0.67; 0.58)	0.393[a]
Number of drugs	6.00 (5.00; 9.00)	8.00 (6.00; 10.00)	1.00 (0.00; 3.00)	0.001[a]
Number of comorbidities	4.00 (3.00; 6.25)	4.00 (2.75; 7.00)	0.00 (0.00; 0.00)	0.763[a]
Fall in last year				
Yes	8 (19.0%)	16 (38.1%)	NA	0.017[b]
No	34 (81.0%)	26 (61.9%)	NA	
Fracture in last year				
Yes	4 (9.5%)	4 (9.5%)	NA	0.268[b]
No	38 (90.5%)	38 (90.5%)	NA	
Katz score	8.00 (8.00; 9.00)	8.00 (8.00; 9.00)	0.00 (0.00; 0.00)	0.942[a]
SPPB score	9.50 (8.00; 11.00)	8.00 (6.75; 11.00)	− 0.50 (− 2.00; 0.25)	0.083[a]
Gait speed (m/s)	1.02 (0.80; 1.21)	0.89 (0.76; 1.09)	− 0.10 (− 0.26; 0.14)	0.032[a]
Chair stand test (s)	14.57 (11.97; 18.29)	16.07 (11.06; 20.94)	1.06 (− 0.86; 3.34)	0.083[a]
Timed up-and-go (s)	10.67 (8.66; 13.31)	12.23 (9.15; 16.27)	0.88 (− 1.18; 3.14)	0.081[a]
Hand grip strength (kg)	19.75 (18.00; 28.00)	19.00 (16.75; 22.50)	− 1.50 (− 5.25; 1.00)	0.010[a]
HGS men (kg)	28.00 (21.00; 37.00)	25.00 (20.50; 31.50)	− 1.00 (− 7.50; 1.00)	0.146[a]
HGS women (kg)	19.00 (14.50; 20.25)	18.00 (12.00; 19.25)	− 2.00 (− 4.25; 1.00)	0.030[a]
ALM (kg)	14.31 (13.09; 18.73)	14.02 (12.94; 18.14)	− 0.29 (− 0.57; 1.16)	0.035[a]
ALM men	18.92 (17.44; 20.26)	19.10 (16.79; 19.97)	− 0.30 (− 0.63; 0.30)	0.287[a]
ALM women	13.40 (12.47; 14.15)	13.21 (12.03; 14.00)	− 0.29 (− 0.48; 0.16)	0.072[c]
ALM/BMI	0.69 (0.57 (0.74)	0.67 (0.58–0.74)	− 0.01 (− 0.04; 0.02)	0.197[a]
ALM/BMI men	0.74 (0.70; 0.88)	0.74 (0.69; 0.89)	− 0.01 (− 0.03; 0.03)	0.940[a]
ALM/BMI women	0.60 (0.53; 0.70)	0.60 (0.51; 0.67)	− 0.02 (− 0.04; 0.02)	0.141[a]
SMI (kg/m^2)	5.57 (5.25–6.70)	5.51 (5.14–6.60)	− 0.02 (− 0.18; 0.15)	0.451[a]
SMI men	6.86 (6.37–7.16)	6.84 (6.26; 7.30)	− 0.01 (− 0.20; 0.17)	0.454[a]
SMI women	5.26 (5.11; 5.52)	5.32 (4.94; 5.47)	− 0.04 (− 0.19; 0.10)	0.440[c]
SarQoL D1				
Physical and Mental Health	58.87 (45.53; 69.15)	51.09 (41.37; 67.19)	− 5.00 (− 12.51; 4.72)	0.107[a]

Table 2 continued

	T1	T3	Change	*P* value
SarQoL D2				
Locomotion	55.56 (46.53; 72.22)	55.56 (38.20; 70.14)	− 2.78 (− 11.81; 5.55)	0.331[c]
SarQoL D3				
Body Composition	58.33 (45.83; 67.71)	50.00 (41.67; 60.63)	− 4.16 (− 12.92; 4.17)	0.023[c]
SarQoL D4				
Functionality	70.24 (59.49; 82.85)	63.46 (47.60; 75.89)	− 4.55 (− 10.70; 1.78)	0.002[a]
SarQoL D5				
Activities of Daily Living	61.61 (43.33; 75.00)	48.22 (37.29; 65.42)	− 6.43 (− 20.00;− 3.12)	< 0.001[a]
SarQoL D6				
Leisure activities	33.25 (29.09; 49.88)	33.25 (16.62; 66.50)	0.00 (− 16.62; 16.62)	0.645[a]
SarQoL D7				
Fears	87.50 (75.00; 100.00)	87.50 (75.00; 100.00)	0.00 (− 12.50; 0.00)	0.382[a]
SarQoL Overall score	61.15 (51.15; 71.76)	54.56 (42.31; 68.44)	− 5.23 (− 12.46; 1.61)	0.002[a]
SF-36 PCS	42.08 (31.86; 49.14)	37.65 (30.47; 48.24)	1.40 (− 5.36; 4.78)	0.679[a]
SF-36 MCS	44.71 (33.86; 53.31)	38.91 (30.55; 50.40)	− 2.18 (− 10.13; 3.77)	0.062[a]
EQ-5D Utility Index	0.800 (0.517–0.827)	0.800 (0.708–0.827)	0.00 (− 0.193; 0.1557)	0.231[a]
EQ-VAS	70.00 (60.00–75.00)	70.00 (60.00–75.00)	0.00 (− 7.50; 5.00)	0.716[a]

NA not applicable, *PCS* Physical Component Summary, *MCS* Mental Component Summary
[a] Wilcoxon matched-pair signed-rank test
[b] Chi-squared test
[c] Paired samples *t* test

($r = -0.37$) and the SF-36 MCS ($r = -0.36$). No other correlations were statistically significant. The full analysis can be found in Table 6.

DISCUSSION

The aim of this study was to evaluate the responsiveness of the SarQoL® questionnaire in a population of older, community-dwelling, sarcopenic subjects by formulating hypotheses on the correlations between change scores, and by calculating the standardized response means. Additionally, we examined the correlations between changes in physical parameters and the evolution of the quality-of-life scores.

The results from the hypotheses reveal that the SarQoL® questionnaire has high responsiveness according to the criteria of De Boer et al., with only one hypothesis out of nine (11%) refuted [26]. The most notable results are the strong correlations found for the Overall score and domain 4 (Functionality) of the SarQoL® questionnaire, and the Physical Functioning domain of the SF-36. These correlations, respectively $r = 0.669$ and $r = 0.680$, were larger than the expected correlation of $r = 0.5$ but make sense in light of the similarity of their content and the relatively important weight of domain 4 in the calculation of the Overall score of the SarQoL® questionnaire.

The SRMs show that the change measured by the Overall score of the SarQoL® questionnaire was significantly larger than that measured by the SF-36 PCS, the EQ-5D utility index and the

Table 3 Evaluation of responsiveness with hypotheses

Hypothesis	Expected strength of correlation	Observed correlation		Confirmation/ rejection	Power ($1 - \beta$)
		r	*p* value		
1. Δ SarQoL Overall score and Δ SF-36 General Health domain are correlated	$r > 0.4$	0.442[a]	0.005	Confirmed	0.851
2. Δ SarQoL Overall score and Δ SF-36 Vitality domain are correlated	$r > 0.3$	0.454[b]	0.004	Confirmed	0.872
3. Δ SarQoL Overall score and Δ SF-36 Physical Functioning domain are correlated	$r > 0.5$	0.669[a]	< 0.001	Confirmed	0.999
4. Δ SarQoL Overall score and Δ EQ-VAS are correlated	$r > 0.4$	0.404[a]	0.009	Confirmed	0.773
5. Δ SarQoL domain 1 (Physical & Mental Health) and Δ SF-36 General Health domain are correlated	$r > 0.3$	0.610[a]	< 0.001	Confirmed	0.994
6. Δ SarQoL domain 1 (Physical & Mental Health) and Δ EQ-VAS are correlated	$r > 0.3$	0.312[a]	0.047	Confirmed	0.531
7. Δ SarQoL domain 2 (Locomotion) and Δ SF-36 Physical Functioning domain are correlated	$r > 0.4$	0.412[a]	0.010	Confirmed	0.791
8. Δ SarQoL domain 4 (Functionality) and Δ SF-36 Physical Functioning domain are correlated	$r > 0.5$	0.680[a]	< 0.001	Confirmed	0.999
9. Δ SarQoL domain 5 (Activities of Daily Living) and Δ SF-36 Physical Functioning domain are correlated	$r > 0.5$	0.467[a]	0.003	Rejected	0.893

Δ = change in; *r* = correlation
[a] Spearman correlation
[b] Pearson correlation

EQ-VAS, but not the SF-36 MCS. The absence of a significant difference between the SRM of the Overall score and the SF-36 MCS indicates a very large 95% confidence interval of the latter. The SRM obtained for the SarQoL® Overall score is in accordance with the change in physical parameters of the subjects. Participants lost approximately 10% of their original gait speed (from a median of 1.02 m/s to 0.89 m/s), and the female participants lost a median of 2 kg of grip strength in the 2-year interval. It is also interesting to note that the number of falls experienced in the year preceding the administration of the test doubled from 8 (19.0%) to 16 (38.1%). The SarQoL® Overall score more accurately reflects these changes, more so than the SF-36 and the EQ-5D.

The SarQoL® questionnaire measured an SRM indicating moderate change for domain 4 (Functionality) and domain 5 (Activities of Daily Living), highlighting that the effects of diminished muscle strength and physical performance manifest themselves most in all the physical tasks performed on a regular basis. SRMs indicating small change were reported for domain 1 (Physical and Mental Health) and domain 3 (Body Composition). The smaller SRM for domain 1 may result from the way the questions are formulated, with many more abstract concepts (energy, physical capacity, muscle mass, etc.) instead of the very relatable examples from domains 4 and 5 (climbing a flight of stairs, opening a bottle or jar, etc.). Subjects may have more difficulty finding the

Table 4 Standardized response means

Domains	SRM	Corrected SRM	Interpretation[a]
1. Δ SarQoL D1 Physical and Mental Health	− 0.31	− 0.34	Small change
2. Δ SarQoL D2 Locomotion	− 0.15	− 0.19	No change
3. Δ SarQoL D3 Body Composition	− 0.37	− 0.47	Small change
4. Δ SarQoL D4 Functionality	− 0.50	− 0.62	Moderate change
5. Δ SarQoL D5 Activities of Daily Living	− 0.57	− 0.56	Moderate change
6. Δ SarQoL D6 Leisure activities	0.04	− 0.04	No change
7. Δ SarQoL D7 Fears	− 0.01	− 0.01	No change
8. Δ SarQoL Overall score	− 0.54	− 0.72	Moderate change
9. Δ SF-36 Physical Functioning	− 0.44	− 0.50	Moderate change
10. Δ SF-36 Social Functioning	− 0.41	− 0.48	Small change
11. Δ SF-36 Role Limitations due to Physical Health	0.02	− 0.02	No change
12. Δ SF-36 Role Limitations due to Emotional Problems	− 0.28	− 0.26	Small change
13. Δ SF-36 Mental Health	− 0.27	− 0.35	Small change
14. Δ SF-36 Vitality	− 0.03	− 0.03	No change
15. Δ SF-36 Bodily Pain	− 0.17	− 0.15	No change
16. Δ SF-36 General Health	− 0..23	− 0.28	Small change
17. Δ SF-36 Physical Component Summary	− 0.18	− 0.20	Small change
18. Δ SF-36 Mental Component Summary	− 0.29	− 0.34	Small change
19. Δ EQ-5D Mobility	0.10	− 0.08	No change
20. Δ EQ-5D Autonomy	− 0.36	NA[b]	Small change
21. Δ EQ-5D Usual activities	0.20	− 0.33	Small change
22. Δ EQ-5D Pain	− 0.07	− 0.06	No change
23. Δ EQ-5D Anxiety	− 0.19	− 0.17	No change
24. Δ EQ-5D Utility Index	0.19	0.18	No change
25. Δ EQ-VAS	− 0.11	− 0.09	No change

[a] Interpretation of corrected SRMs: $0.20 \leq SRM < 0.49$ = small change; $0.50 \leq SRM < 0.79$ = moderate change; $SRM \geq 0.80$ = large change

[b] Correction for SRM of EQ-5D Autonomy cannot be computed because Δ EQ-5D Autonomy at T3 is constant (all subjects responded with the same answer)

right answers for them because these changes are much less perceptible in absolute terms. The SRM for domain 3 (Body Composition) covers an area where drastic change is not necessarily expected given that the median age in the sample is 73 years old and that many of the age-related changes to the way one looks have already manifested themselves. Finally, three domains reported SRMs that indicate no change has occurred. Domain 6 (Leisure Activities) and domain 7 (Fears) are represented by, respectively, two and four items in the questionnaire

Table 5 Exploration of significant differences between SRMs

Hypothesis	Intercept	*p* value	Interpretation	Larger SRM
The SRMs of SarQoL Overall score and SF-36 PCS score are significantly different	− 0.326	0.005	Different	SarQoL
The SRMs of SarQoL Overall score and SF-36 MCS score are significantly different	− 0.236	0.150	Not different	None
The SRMs of SarQoL Overall score and EQ-5D Utility Index are significantly different	− 0.724	< 0.001	Different	SarQoL
The SRMs of SarQoL Overall score and EQ-VAS are significantly different	− 0.443	0.003	Not different	SarQoL

Calculation of *p* values carried out with modified jack-knife method [30]

and may be much less sensitive than domains with more items. For domain 2 (Locomotion), this reasoning does not apply. This domain asks pointed questions connected to walking (length, frequency, difficulties, tiredness, etc.), and given that the usual gait speed has significantly diminished, one would expect to see an effect in this domain. However, the questions in this domain may be affected by the phenomenon of response shift, whereby the internal standards of measurement of the subject are recalibrated.

The SF-36 reported moderate change for the domain Physical Functioning, and small change for the domains Social Functioning, Role limitations due to emotional problems, Mental Health, and General Health, and reported no change for the other domains. These results are in line with our hypothesis that the SarQoL® questionnaire, being specific to sarcopenia, should detect a greater change than generic questionnaires such as the SF-36. The EQ-5D reported a small change for the domains Autonomy and Usual Activities and no change for all other scores. This should not be surprising given the distance between the response options for the EQ-5D, which means a significant change needs to occur in real life for it to be registered in the change scores.

Lastly, the correlations between changes in physical parameters and the changes on the different overall/composite scores revealed three good correlations for the SarQoL® Overall score, one good and two acceptable correlations for the SF-36 PCS, one acceptable correlation for the SF-36 MCS, no correlations for the EQ-5D Utility Score and one good correlation for the EQ-VAS. In general, the SarQoL® Overall score correlates well with physical performance, with good correlations for change in gait speed, SPPB and CST. However, these results should be interpreted with caution given the multidimensional nature of sarcopenia, which is unlikely to be covered in a single test.

This study has several strengths. The methodology we adopted supplied us with evidence from different sources and allowed us to show both the quality and quantity of responsiveness. We were able to draw upon the data collected within the SarcoPhAge study, which allowed us to have a moderate sample size (n = 42) despite the relatively low prevalence of sarcopenia. Furthermore, the SarcoPhAge study collected muscle mass data with DXA, which is, in practice, the most reliable method, and collected data on a number of tests for physical performance, which allowed us to compare the changes on several physical parameters [35].

There are, however, several limitations in this study. The SarcoPhAge study was not specifically designed to allow the evaluation of the responsiveness of the SarQoL® questionnaire, lacking both a known intervention and a transition question. A second limitation is that the primary methodology used in this study, the testing of hypotheses, has only been

Table 6 Correlations between changes in physical parameters and evolution of quality of life

Domains	*r*	*p* value	Interpretation
1. Δ Gait speed and Δ SarQoL Overall	0.50	0.001	Good
2. Δ Gait speed and Δ SF-36 PCS	0.39	0.017	Acceptable
3. Δ Gait speed and Δ SF-36 MCS	0.02	0.926	NS
4. Δ Gait speed and Δ EQ-5D Utility Index	− 0.09	0.560	NS
5. Δ Gait speed and Δ EQ-VAS	0.16	0.324	NS
6. Δ Grip strength and Δ SarQoL Overall	0.08	0.592	NS
7. Δ Grip strength and Δ SF-36 PCS	0.27	0.104	NS
8. Δ Grip strength and Δ SF-36 MCS	− 0.14	0.393	NS
9. Δ Grip strength and Δ EQ-5D Utility Index	0.22	0.165	NS
10. Δ Grip strength and Δ EQ-VAS	0.08	0.626	NS
11. Δ SPPB and Δ SarQoL Overall	0.47	0.002	Good
12. Δ SPPB and Δ SF-36 PCS	0.30	0.068	NS
13. Δ SPPB and Δ SF-36 MCS	0.25	0.131	NS
14. Δ SPPB and Δ EQ-5D Utility Index	0.12	0.450	NS
15. Δ SPPB and Δ EQ-VAS	0.12	0.450	NS
16. Δ ALM and Δ SarQoL Overall	0.15	0.355	NS
17. Δ ALM and Δ SF-36 PCS	0.04	0.829	NS
18. Δ ALM and Δ SF-36 MCS	0.19	0.264	NS
19. Δ ALM and Δ EQ-5D Utility Index	0.03	0.832	NS
20. Δ ALM and Δ EQ-VAS	< − 0.01	0.986	NS
21. Δ ALM/BMI and Δ SarQoL Overall	− 0.02	0.901	NS
22. Δ ALM/BMI and Δ SF-36 PCS	− 0.14	0.807	NS
23. Δ ALM/BMI and Δ SF-36 MCS	− 0.11	0.537	NS
24. Δ ALM/BMI and Δ EQ-5D Utility Index	− 0.06	0.726	NS
25. Δ ALM/BMI and Δ EQ-VAS	− 0.48	0.002	Good
26. Δ ALM/Ht^2 and Δ SarQoL Overall	0.11	0.477	NS
27. Δ ALM/Ht^2 and Δ SF-36 PCS	0.10	0.570	NS
28. Δ ALM/Ht^2 and Δ SF-36 MCS	0.22	0.192	NS
29. Δ ALM/Ht^2 and Δ EQ-5D Utility Index	0.01	0.964	NS
30. Δ ALM/Ht^2 and Δ EQ-VAS	< 0.01	0.989	NS
31. Δ TUG and Δ SarQoL Overall	− 0.17	0.279	NS
32. Δ TUG and Δ SF-36 PCS	− 0.44	0.007	Good
33. Δ TUG and Δ SF-36 MCS	− 0.23	0.174	NS

Table 6 continued

Domains	*r*	*p* value	Interpretation
34. Δ TUG and Δ EQ-5D Utility Index	− 0.02	0.923	NS
35. Δ TUG and Δ EQ-VAS	− 0.02	0.882	NS
36. Δ CST and Δ SarQoL Overall	− 0.42	0.013	Good
37. Δ CST and Δ SF-36 PCS	− 0.37	0.032	Acceptable
38. Δ CST and Δ SF-36 MCS	− 0.36	0.040	Acceptable
39. Δ CST and Δ EQ-5D Utility Index	− 0.13	0.470	NS
40. Δ CST and Δ EQ-VAS	− 0.11	0.546	NS

Δ = change in; *r* = correlation

NS not significant, *SPPB* Short Physical Performance Battery, *ALM* appendicular lean mass, *ALM/BMI* ALM divided by body mass index, *ALM/Ht*^{2} ALM divided by height squared, *TUG* timed up-and-go test, *CST* chair stand test

introduced a few years ago and that several questions about this process have not yet found a consensus, such as how many hypotheses should be tested, what percentage should be confirmed for good responsiveness and how to set the strength of the expected correlations. We have tried to address these issues by using pre-defined, specific and challenging hypotheses but recognize that this methodology should be considered an ongoing process and hope that other studies can re-evaluate our hypotheses and add their own. Lastly, the SF-36 PCS and MCS scores were used in the evaluation of the SRMs but not in the hypotheses. We acknowledge that the PCS and MCS would have made good targets for the formulation of hypotheses, but unfortunately, the choice to calculate these scores was made after the hypotheses were formulated and after the statistical manipulations had started. It was therefore impossible for us to include the PCS and MCS scores in the hypotheses. It is our hope that future responsiveness studies will include the PCS and MCS in their hypotheses.

CONCLUSIONS

This study contributed data on the last major psychometric property of the SarQoL® questionnaire not yet studied. The questionnaire has good responsiveness, measured both in an evaluation with hypotheses (8/9 confirmed) and by the strength of its standardized response means. The SarQoL® questionnaire appears to be the optimal tool for the assessment of quality of life in sarcopenic populations. Its use in clinical trials assessing biochemical entities for the management of sarcopenia should be recommended, as patient-related outcomes are encouraged to be included as co-primary endpoints in such studies [36].

ACKNOWLEDGEMENTS

We thank the participants of the SarcoPhAge study, whose continued participation made this article possible.

Funding. No funding or sponsorship was received for this study.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Authorship Contributions. AG, CB, OB and JYR developed the study design. CB and ML collected the data. AG was responsible for data analysis and drafted the article. All authors

provided revision of the draft article, and read and approved the final manuscript.

Prior Presentation. The data in this article have been presented as a poster at the ICFSR 2018 congress (Miami, USA) and the WCO-IOF-ESCEO 2018 congress (Krakow, Poland).

Disclosures. OB, CB and J-YR are shareholders of SarQoL® sprl. J-YR is the president of the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO), which has endorsed the SarQoL® questionnaire. AG and ML declare they have no conflicts of interest.

Compliance with Ethics Guidelines. Participants in the SarcoPhAge study all provided written informed consent. The research protocol and its amendments were approved by the Ethics Committee of the University Teaching Hospital of Liège (no. 2012-277).

Data Availability. The dataset used for this study is available from the corresponding author on reasonable request.

REFERENCES

1. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, et al. Sarcopenia: European consensus on definition and diagnosis. Age Ageing. 2010;39:412–23.

2. Cao L, Morley JE. Sarcopenia is recognized as an independent condition by an International Classification of Disease, tenth revision, clinical modification (ICD-10-CM) code. J Am Med Dir Assoc. 2016;17:675–7.

3. Beaudart C, Zaaria M, Pasleau F, Reginster J-Y, Bruyère O. Health outcomes of sarcopenia: a systematic review and meta-analysis. PLoS One. 2017;12:e0169548.

4. Chang K-V, Hsu T-H, Wu W-T, Huang K-C, Han D-S. Is sarcopenia associated with depression? A systematic review and meta-analysis of observational studies. Age Ageing. 2017;46:738–46.

5. Beaudart C, Reginster J-Y, Geerinck A, Locquet M, Bruyère O. Current review of the SarQoL®: a health-related quality of life questionnaire specific to sarcopenia. Expert Rev Pharmacoecon Outcomes Res. 2017;17:335–41.

6. Rizzoli R, Reginster JY, Arnal JF, et al. Quality of life in sarcopenia and frailty. Calcif Tissue Int. 2013;93:101–20.

7. Beaudart C, Biver E, Reginster J-Y, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. Age Ageing. 2015;44:960–6.

8. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737–45.

9. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? J Clin Epidemiol. 1992;45:1341–5.

10. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40:171–8.

11. Beaudart C, Biver E, Reginster J-Y, Rizzoli R, Rolland Y, Bautmans I, et al. Validation of the SarQoL, a specific health-related quality of life questionnaire for sarcopenia. J Cachexia Sarcopenia Muscle. 2017;8:238–44.

12. Beaudart C, Edwards M, Moss C, et al. English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. Age Ageing. 2017;46:271–7.

13. Ildiko GA, Mihai G, Beaudart C, et al. Psychometric performance of the Romanian version of the SarQoL®, a health-related quality of life questionnaire for sarcopenia. Arch Osteoporos. 2017;12:103.

14. Geerinck A, Scheppers A, Beaudart C, et al. Translation and validation of the Dutch SarQoL®, a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact. 2018. http://www.ismni.org/jmni/accepted/jmni_aa_BAUTMANS.pdf.

15. Beaudart C, Reginster J-Y, Petermans J, et al. Quality of life and physical components linked to sarcopenia: the SarcoPhAge study. Exp Gerontol. 2015;69:103–10.

16. Locquet M, Beaudart C, Bruyère O, Kanis JA, Delandsheere L, Reginster J-Y. Bone health assessment in older people with or without muscle health impairment. Osteoporos Int. 2018;29:1057–67.

17. Beaudart C, Reginster J-Y, Petermans J, Bruyère O. Qualité de vie du patient sarcopénique: apport de l'étude liégeoise SarcoPhAge. Geriatr Psychol Neuropsychiatr Vieil. 2015;13:391–5.

18. Locquet M, Beaudart C, Reginster J-Y, et al. Prevalence of concomitant bone and muscle wasting in elderly women from the SarcoPhAge cohort: preliminary results. J Frailty Aging. 2017;6:18–23.

19. The SarQoL questionnaire [Internet]. http://www.sarqol.org. Accessed 7 Sep 2018.

20. Ware JJ, Sherbourne C. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30:473–83.

21. Ware JJ, Konsinski M. SF-36 physical and mental health summary scales: a manual for users of version 1. Lincoln: QualityMetric; 2001.

22. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in orthopaedics: a brief guide. J Bone Jt Surg. 2015;97:1628–34.

23. Van Reenen M, Oppe M. EQ-5D-3L User guide. Basic information on how to use the EQ-5D-3L instrument. 2015. EuroQol Group. https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-3L_UserGuide_2015.pdf.

24. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001;33:337–43.

25. De Vet HCW, Knol DL, Terwee CB, Mokkink LB. Measurements in medicine. Cambridge: Cambridge University Press; 2011.

26. De Boer M, Moll A, de Vet H, Terwee CB, Völker-Dieben HJM, van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. Ophthalmic Physiol Opt. 2004;24:257–73.

27. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000;53:459–68.

28. Middel B, Van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. Int J Integr Care. 2002;2:e15.

29. Cohen J. Quantitative methods in psychology. Psychol Bull. 1992;112:155–9.

30. Bessette L, Sangha O, Kuntz K, et al. Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. Med Care. 1998;36:491–502.

31. Angst F, Verra ML, Lehmann S, Aeschlimann A. Responsiveness of five condition-specific and generic outcome assessment instruments for chronic pain. BMC Med Res Methodol. 2008;8:26.

32. Evans JD. Straightforward statistics for the behavioral sciences. Pacific Grove: Brooks/Cole; 1996.

33. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39:175–91.

34. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21:651–7.

35. Buckinx F, Landi F, Cesari M, et al. Pitfalls in the measurement of muscle mass: a need for a reference standard. J Cachexia Sarcopenia Muscle. 2018;9:269–78.

36. Reginster J-Y, Cooper C, Rizzoli R, et al. Recommendations for the conduct of clinical trials for drugs to treat or prevent sarcopenia. Aging Clin Exp Res. 2016;28:47–58.

Measurement error of the SarQoL questionnaire

PLOS ONE

RESEARCH ARTICLE

Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: An analysis of subjects from 9 validation studies

Anton Geerinck[1]*, Vidmantas Alekna[2], Charlotte Beaudart[1], Ivan Bautmans[3], Cyrus Cooper[4], Fabiana De Souza Orlandi[5], Jerzy Konstantynowicz[6], Beatriz Montero-Errasquín[7], Eva Topinková[8], Maria Tsekoura[9], Jean-Yves Reginster[1], Olivier Bruyère[1]

1 World Health Organization Collaborating Center for Public Health Aspects of Musculoskeletal Health and Ageing, Department of Public Health, Epidemiology and Health Economics, University of Liège, Liège, Belgium, **2** Faculty of Medicine, Vilnius University, Vilnius, Lithuania, **3** Frailty in Aging Research Group, Vrije Universiteit Brussel, Brussels, Belgium, **4** MRC Environmental Epidemiology Unit, Southampton General Hospital, Southampton, United Kingdom, **5** Department of Gerontology, Federal University of São Carlos, São Carlos, SP, Brazil, **6** Department of Pediatric Rheumatology, Immunology and Metabolic Bone Diseases, Medical University of Bialystok, Bialystok, Poland, **7** Department of Geriatrics, University Hospital Ramón y Cajal, Madrid, Spain, **8** Department of Geriatrics and Gerontology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic, **9** Department of Physical Therapy, School of Health and Welfare, Technological Institute of Western Greece, Aigio, Greece

* anton.geerinck@uliege.be

Abstract

Objectives

The Sarcopenia Quality of Life (SarQoL) questionnaire, a sarcopenia-specific patient-reported outcome measure, evaluates quality of life with 55 items. It produces 7 domain scores and 1 overall quality of life score, all between 0 and 100 points. This study aims to contribute to the interpretation of the SarQoL scores by calculating the standard error of measurement (SEM) and smallest detectable change (SDC) in a sample of subjects from 9 studies.

Methods

Subjects from 9 studies (conducted in Belgium, Brazil, Czech Republic, England, Greece, Lithuania, Poland and Spain) were included. The SEM, a measure of the error in the scores that is not due to true changes, was calculated by dividing the standard deviation of the difference between test and retest scores (SDdiff) by $\sqrt{2}$. The SDC, defined as change beyond measurement error, was calculated by multiplying SDdiff by 1.96. Bland-Altman plots were assessed for the presence of systematic errors.

Results

A total of 278 sarcopenic subjects, aged 77.67 ± 7.64 years and 61.5% women, were included. The SEM for the overall SarQoL score ranged from 0.18 to 4.20 points for the individual studies, and was 2.65 points when all subjects were analyzed together. The SDC for

OPEN ACCESS

Citation: Geerinck A, Alekna V, Beaudart C, Bautmans I, Cooper C, De Souza Orlandi F, et al. (2019) Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: An analysis of subjects from 9 validation studies. PLoS ONE 14 (4): e0216065. https://doi.org/10.1371/journal.pone.0216065

Editor: Ali Montazeri, Iranian Institute for Health Sciences Research, ISLAMIC REPUBLIC OF IRAN

Received: December 6, 2018

Accepted: April 12, 2019

Published: April 29, 2019

Data Availability Statement: The dataset for this study has been deposited on the Open Science Framework (www.osf.io) at DOI 10.17605/OSF.IO/GD2U8.

Funding: The authors received no specific funding for this work.

Competing interests: AG is supported by a FRIA fellowship grant from the F.R.S.-FNRS (Belgian Fund for Scientific Research). CB, OB, J-YR, IB &

CC are shareholders of SarQoL sprl. J-YR is the president of the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO), which has endorsed the SarQoL® questionnaire. CC reports personal fees from Alliance for Better Bone Health, Amgen, Eli Lilly, GSK, Medtronic, Merck, Novartis, Pfizer, Roche, Servier, Takeda and UCB. All other authors have declared that no competing interests exist. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

the overall score ranged from 0.49 to 11.65 points for the individual studies, and was 7.35 points for all subjects. The Bland-Altman plots revealed no systematic errors in the questionnaire.

Conclusion

This study shows that, for individual subjects, a change in overall quality of life of at least 7.35 points (on a scale from 0 to 100) would have to be observed to confirm that a true change, beyond measurement error, has occurred. It also demonstrated that the SarQoL questionnaire is a precise instrument, with the observed scores within less than 3 points of the theoretical "true score".

Introduction

Sarcopenia, often described as the age-related loss of muscle mass and strength, and defined by the European Working Group on Sarcopenia in Older People (EWGSOP2) as *"a progressive and generalised skeletal muscle disorder that is associated with increased likelihood of adverse outcomes including falls, fractures, physical disability and mortality"*, has been the subject of increased scientific attention as its prevalence and consequences have become more known [1]. Sarcopenia is confirmed to be present when a patient is diagnosed with low muscle strength and low muscle mass. When low physical performance is also established, that person is diagnosed with severe sarcopenia [1]

A systematic review conducted in 2014 which estimated the prevalence of sarcopenia diagnosed with the EWGSOP-algorithm in older community-dwelling adults found a range of 1 to 29% (up to 30% in women), while a recent meta-analysis which included 35 articles and a total of 58404 healthy subjects aged 60 years and older found an overall prevalence of sarcopenia of 10% (95% CI: 8–12%) in men and 10% (95% CI: 8–13%) in women diagnosed with the EWGSOP, the International Working Group on Sarcopenia (IWGS) or the Asian Working Group for Sarcopenia (AWGS) definitions [2,3]. It should be mentioned that the prevalence of sarcopenia varies greatly depending on the definition used, as demonstrated by Beaudart et al., who applied 6 different diagnostic criteria for sarcopenia to a single cohort of subjects and found a prevalence rate from 4.39% to 32.8% [4].

Projections about the future prevalence of sarcopenia (as diagnosed by the EWGSOP-criteria) in the European Union (EU28) predict a rise from 10.9 million people in 2016 to 18.7 million in 2045 on the low end and from 19.7 million to 32.3 million people on the high end [5]. Sarcopenia is a major public health problem and its impact will continue to grow, which should incite policy makers to act.

The available evidence concerning the impact and association of sarcopenia with several health outcomes has been steadily growing during the last decade. A systematic review and meta-analysis published in 2017 provided a comprehensive summary of what is currently known on the subject. This review included 17 prospective studies in which sarcopenia was diagnosed according to the EWGSOP guidelines. The authors found a higher risk for mortality (OR = 3.596; 95% CI = 2.96–4.37) and functional decline (OR = 3.03; 95% CI = 1.80–5.12) as well as a higher rate of falls and a higher incidence of hospitalization. The evidence on the incidence of fractures and the length of hospital stay was inconclusive [6].

The subject of quality of life in sarcopenia has mostly been examined using generic questionnaires such as the Short-Form 36-Item (SF-36) and the EuroQoL 5-Dimension (EQ-5D)

[7]. Recently, a new instrument, the Sarcopenia Quality of Life (SarQoL) questionnaire has become available. It is specifically designed to measure quality of life in sarcopenic, community-dwelling individuals aged 65 years or older and was developed in 2013–2015 by Beaudart et al. [8]. It has, to date, been translated into more than 20 languages [8].

The psychometric properties of the SarQoL questionnaire have been evaluated and published for 6 language-versions: the original questionnaire in French, and the English, Dutch, Polish, Romanian and Greek translations [9–14]. These examined the discriminative power, internal consistency, construct validity, test-retest reliability and the presence of floor or ceiling effects. These 6 studies found that the questionnaire can discriminate between sarcopenic and non-sarcopenic participants, with the former having significantly lower scores for the 7 domains and the overall score, and that the questionnaire possesses good internal consistency (Cronbach's alpha of 0.87, 0.88, 0.95, 0.92, 0.88 and 0.96). These studies also confirmed the construct validity of the SarQoL questionnaire with the help of hypotheses on correlations between the questionnaire and the SF-36 and EQ-5D, and demonstrated that the SarQoL questionnaire has an excellent test-retest reliability (intraclass correlation coefficient/ICC = 0.91, 0.95, 0.99, 0.98 and 0.96) [9–14]. Lastly, floor and ceiling effects were absent from all 6 published validation studies [9–14]. These results provide convincing evidence for the validity and reliability of the SarQoL questionnaire for the evaluation of quality of life in sarcopenic, community-dwelling older people.

However, until now, the standard error of measurement (SEM) and the smallest detectable change (SDC) of the SarQoL questionnaire have not yet been calculated. These parameters supply important information on the reliability of the instrument in question by indicating the range in which the theoretical "true" score lies; and supply context when interpreting data from longitudinal measurements by indicating by how much the score needs to change before one can be reasonably certain that a true change has occurred. Clinicians and researchers could use the values for SEM and SDC as a yardstick in the interpretation of the SarQoL scores, whether obtained in clinical practice or as part of a research project. The results of this study should prove particularly valuable in the interpretation of data from interventional clinical trials, and will hopefully expedite the adoption of this PROM in clinical trials [15].

The primary objective of this study is to determine the SEM and SDC of the SarQoL questionnaire in a sample of subjects from 9 international validation studies. The secondary objectives are to examine the measurement error of the questionnaire with the help of a Bland-Altman analysis, and to update the results previously obtained for the test-retest reliability of the SarQoL questionnaire in the complete sample.

Material and methods

This study combined data from 9 cohorts in 8 different countries that were established to test the psychometric properties of the SarQoL questionnaire after translation into the local language. The team behind the SarQoL questionnaire have made a concerted effort to widen the reach of the questionnaire by having it translated into a multitude of languages. To accomplish this, they have partnered with researchers from a host of countries and language groups, who were able and willing to undertake a translation of the questionnaire. The local teams responsible for the translations were also encouraged to carry out a validation study of the translation they produced, if feasible. A considerable number of them undertook this effort, although not all validations have been published. The researchers from 9 validation studies that had the necessary data for the current analysis were contacted and agreed to share their data. All the included studies obtained approval from their local ethics committees, and written informed consent from their participants.

Population

Subjects were included in the 9 validation studies if they were 60 years of age or older and community-dwelling. For this analysis, we included all subjects who were diagnosed as being sarcopenic, who completed the SarQoL questionnaire twice and reported that their health had been stable in the interval between the two administrations.

The SarQoL questionnaire

The analyses in this article center around the test-retest data for the SarQoL questionnaire collected by the 9 included studies. The SarQoL questionnaire is a patient-reported outcome measure (PROM) designed specifically for use with sarcopenic, community-dwelling subjects 65 years of age or older. The questionnaire consists of 55 items distributed over 22 questions, with the items categorized into 7 domains of health-related quality of life (HRQoL). These domains are: "Physical and Mental Health" (D1), "Locomotion" (D2), "Body Composition" (D3), "Functionality" (D4), "Activities of Daily Living" (D5), "Leisure activities" (D6), and "Fears" (D7). Apart from the domain scores, an Overall score for quality of life is also calculated. All scores are situated on a scale from 0 to 100, with 0 being the worst possible quality of life, and 100 the best possible. The questionnaire is auto-administered and takes about 10 minutes to complete [9]. More information on the SarQoL questionnaire and the different language-specific versions can be found on www.sarqol.org.

Test-retest reliability

The test-retest reliability of a questionnaire quantifies the extent to which a questionnaire produces the same scores during repeated measurements, provided that the participants' health remains stable. It is measured by the intraclass correlation coefficient (ICC) under a 2-way mixed model with absolute agreement specified, and its associated 95% confidence interval. A questionnaire is considered reliable if the obtained ICC values are greater than 0.70 [16].

Standard error of measurement

The standard error of measurement has been defined as *"the determination of the amount of variation or spread in the measurement errors for a test"* [17]. The SEM is considered to be a parameter for the amount of measurement error present in an instrument, and is subsequently an indicator of the reliability of said instrument. Much like the interpretation of the standard deviation around the mean value, the SEM can be used to provide a range around the observed value within which the theoretical "true" value lies. The interval between plus and minus 1 SEM provides a probability of 68% of containing the true value. For ± 2 SEM the probability becomes 95% and for ± 3 SEM we end up with 99% probability.

Smallest detectable change

The smallest detectable change is defined as the change in the instrument's score beyond measurement error [18]. This means that the SDC provides a value for the minimum change that needs to be observed in order to be confident that the observed change is real and not, potentially, a product of measurement error in the instrument. The SDC can be calculated for individual subjects (SDC_{ind}) as well as for comparisons of mean scores between groups (SDC_{group}) [18]. Both provide utility: The SDC_{ind} can be used in clinical practice or to label individual subjects in a study sample as either changed or unchanged. The SDC_{group} provides an aid to the interpretation of mean scores of groups. This can lend greater credibility to the results of

interventional trials that use the SarQoL questionnaires, and that want to know whether quality of life has changed in the intervention and control group as a whole.

Bland-Altman analysis

The Bland-Altman plot provides a visual representation of the presence of systematic errors in an instrument. The Bland-Altman plot is based around three variables: the mean systematic difference between test and retest scores ($\bar{d}$), and the upper and lower limit of agreement, which span 95% of observations, assuming that the values for the difference between test and retest scores are distributed normally [18,19]. These variables are integrated into a scatter plot where the difference between test and retest values is put on the Y-axis and the average of the test and retest values is put on the X-axis.

Statistical analysis

Data were analyzed using IBM SPSS Statistics, version 24.0.0.0 for Windows (Armonk, NY: IBM Corp). The distribution of the variables was determined by examining the histogram, the quantile-quantile-plot, the Shapiro-Wilk test and the difference between mean and median. Variables that are normally distributed are reported as mean ± standard deviation and non-normal variables as median (25^{th} percentile– 75^{th} percentile). Nominal variables are reported as absolute (n) and relative frequencies (%).

Differences between groups with regards to clinical characteristics were examined with one-way anova analysis for continuous variables and chi-squared test for nominal variables.

The SEM was calculated by first creating a variable for the difference between the score obtained during the first and the second administration (test score—retest score = *Difference*). Next, we calculated the standard deviation of *Difference* in our sample ($SD_{difference}$) and divided the obtained value by the square root of 2 ($SEM = \frac{SDdifference}{\sqrt{2}}$) [18,20].

The SDC_{ind} was calculated with the formula [$SDC_{ind} = 1.96 * \sqrt{2} * SEM$], and the SDC_{group} was calculated by dividing the SDC_{ind} by the square root of the number of subjects in the sample ($\frac{SDCind}{\sqrt{n}}$) [18].

The ICC was calculated with a 2-way mixed model and absolute agreement specified.

The mean difference score ($\bar{d}$) was calculated by calculating the mean of the differences between test and retest scores for all subjects [Mean(test score—retest score)]. The 95% limits of agreement were calculated with the formula [$\bar{d} \pm (1.96 * SD_{difference})$] [18,21]. Bland-Altman plots were created in SPSS following the instructions given in IBM tech-note n° 19420 [22]

Results were considered significant at $p \leq 0.05$.

Results

Characteristics of included studies

Information on the diagnosis of sarcopenia and the characteristics of the test-retest administration are given in Table 1.

Clinical characteristics

The 278 participants included in the analysis had a mean age of 77.67 ± 7.64 years, ranging from 60 to 98 years old. The majority of subjects were women, namely 171 participants or 61.5% of the complete sample. The participants had a mean body mass index of 25.57 ± 4.40 kg/m^2, spanning the whole gambit from underweight to morbidly obese with a minimum value of 17.42 kg/m^2 and a maximum value of 46.10 kg/m^2. In terms of prescription drug use,

Table 1. Characteristics of included studies.

	Sarcopenia diagnosis				Time between test and retest administration	Mode of administration	
	Sarcopenia definition	Muscle mass assessment	Muscle strength assessment	Physical performance assessment		Test	Retest
Belgium (Dutch) [12]	EWGSOP	BIA	Martin-Vigorimeter	Gait speed	2 weeks	At study center	At home
Belgium (French) [9]	EWGSOP	DXA	Hand dynamometer	SPPB	2 weeks	At study center	At home
Brazil	EWGSOP	DXA	Hand dynamometer	Gait speed	2 weeks	At home	At home
Czech Republic [23]	FNIH	DXA	Hand dynamometer	SPPB	2 weeks	At home or at study center without staff present	At home or at study center without staff present
England [10]	EWGSOP	DXA	Hand dynamometer	Gait speed	2 weeks	At home	At home
Greece [14]	EWGSOP	BIA	Hand dynamometer	Gait speed	2 weeks	At study center	At study center
Lithuania	EWGSOP	DXA	Hand dynamometer	SPPB	2 weeks	At study center	At study center
Poland [13]	EWGSOP	Lee equation [24]	Hand Dynamometer	Not performed	2 weeks	At study center	At study center
Spain	FNIH	DXA	Hand dynamometer	SPPB	2 weeks	At study center	At home

EWGSOP: European Working Group on Sarcopenia in Older People; BIA: bioelectrical impedance analysis; DXA: dual-energy x-ray absorptiometry; FNIH: Foundation for the National Institutes of Health

https://doi.org/10.1371/journal.pone.0216065.t001

the subjects took on average 4.78 ± 2.71 drugs (range: 0–13), linked to the number of comorbidities which was 3.59 ± 2.01 (range: 0–11). Clinical characteristics are reported in Table 2.

As expected, one-way anova analyses and chi-squared test revealed that the 9 studies differed significantly in terms of clinical characteristics. The results from these post-hoc analyses can be found in S1–S5 Tables.

The test-retest reliability of the SarQoL questionnaire in the complete sample resulted in an ICC of 0.969 (95% CI = 0.961–0.975) for the Overall score. Of the individual domains, 4 obtained an ICC higher than 0.9, namely domain 1, 2, 4 and 5, and all obtained ICC's higher than 0.7. The detailed results for the test-retest reliability can be found in Table 3.

Table 2. Clinical characteristics for individual studies–mean ± SD or n(%).

	All	Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain
n	278	26	29	12	48	10	58	50	30	15
Age (years)	77.67 ± 7.64	81.00 ± 5.88	77.03 ± 6.58	70.75 ± 6.57	82.96 ± 6.05	78.90 ± 2.56	80.18 ± 6.42	72.10 ± 7.71	73.82 ± 7.06	77.60 ± 6.27
Gender										
Female	171 (61.5)	12 (46.2)	19 (65.5)	6 (50.0)	37 (77.1)	3 (30.0)	28 (48.3)	37 (74.0)	19 (63.3)	10 (66.7)
Body mass index (kg/m^2)	25.57 ± 4.40	26.71 ± 4.75	23.16 ± 3.19	24.84 ± 4.32	29.16 ± 5.78	24.00 ± 2.73	24.62 ± 2.54	24.05 ± 3.39	27.01 ± 4.46	24.17 ± 1.99
Drugs (n)	4.78 ± 2.71	3.81 ± 2.62	6.72 ± 2.76	7.25 ± 1.55	6.27 ± 3.30	6.00 ± 2.45	4.36 ± 1.25	3.50 ± 1.28	2.70 ± 2.84	5.13 ± 2.75
Concomitant illnesses (n)	3.59 ± 2.01	2.48 ± 1.64	4.93 ± 2.36	4.17 ± 1.59	5.79 ± 1.47	NA	2.98 ± 0.78	2.96 ± 1.01	1.60 ± 1.85	3.80 ± 2.04

https://doi.org/10.1371/journal.pone.0216065.t002

Table 3. Results for complete analysis (n = 278).

	Test scores	Retest scores	ICC (95% CI)	$\bar{d}$ (95% CI)	SD_{diff}	SEM	SDC_{ind}	SDC_{group}	95% LoA
D1: Physical & mental health	56.56 ± 17.00	57.42 ± 17.12	0.915 (0.894; 0.933)	0.86 (0.04; 1.68)	6.98	4.94	13.68	0.82	-12.82; 14.54
D2: Locomotion	54.95 ± 21.40	54.88 ± 21.54	0.944 (0.929; 0.955)	-0.07 (-0.93; 0.78)	7.23	5.11	14.17	0.85	-14.24; 14.1
D3: Body composition	55.36 ± 16.91	56.10 ± 17.18	0.836 (0.797; 0.869)	0.74 (-0.41; 1.89)	9.74	6.89	19.09	1.14	-18.35; 19.83
D4: Functionality	62.31 ± 17.08	62.70 ± 16.61	(0.952 (0.939; 0.962)	0.39 (-0.23; 1.01)	5.24	3.71	10.27	0.62	-9.88; 10.66
D5: Activities of daily living	55.55 ± 17.33	55.40 ± 17.73	0.915 (0.894; 0.933)	-0.15 (-1.00; 0.70)	7.23	5.11	14.17	0.85	-14.32; 14.02
D6: Leisure activities	37.61 ± 17.83	37.00 ± 19.23	0.754 (0.698; 0.800)	-0.59 (-2.13; 0.94)	13.04	9.22	25.56	1.53	-26.15; 24.97
D7: Fears	78.98 ± 17.47	78.96 ± 17.13	0.783 (0.733; 0.825)	-0.02 (-1.37; 1.33)	11.42	8.08	22.38	1.34	-22.4; 22.36
Overall score	57.71 ± 14.97	57.89 ± 15.03	0.969 (0.961; 0.975)	0.18 (-0.26; 0.63)	3.75	2.65	7.35	0.44	-7.17; 7.53

ICC = intraclass correlation coefficient; $\bar{d}$ = mean difference score; CI = confidence interval; SD_{diff} = standard deviation of difference scores; SEM = standard error of measurement; SDC_{ind} = smallest detectable change for individual subject; SDC_{group} = smallest detectable change for group; LoA = limits of agreement

https://doi.org/10.1371/journal.pone.0216065.t003

Standard error of measurement

The SEM for the Overall score of the SarQoL questionnaire in the complete sample is 2.65 points. This means that one can be 68% confident (± 1 SEM) that the 'true' score of a subject can be found between -2.65 and +2.65 points from the observed score, and 95% confident (± 2 SEM) that the 'true' score is situated between -5.3 and +5.3 points of the observed score. The SEM for the different domains of the SarQoL questionnaire in the complete sample varied between 3.71 for domain 4 and 9.22 points for domain 6. The SEM-values for the complete sample can be found in Table 3, while the SEM-values for the individual included studies are available in Table 4.

Smallest detectable change

The SDC_{ind} for the Overall score of the SarQoL questionnaire in the complete sample is 7.35 points. This means that the Overall quality of life score of an individual would have to change with at least 7.35 points (on a scale of 0 to 100) before the observed change can be considered to be a true change in the quality of life of a subject, and not potentially a result of measurement error. The SDC_{ind} for the 7 domains of the SarQoL questionnaire goes from a minimum value of 10.27 points for domain 4 to a maximum value of 25.56 points for domain 6. The SDC_{group} for the Overall score in the complete sample is 0.44 points. The SDC-values for the complete sample can be found in Table 3. The SDC-values for the individual included studies are available in Table 4.

Bland-Altman analysis

The mean difference score in the complete sample for the Overall score of the SarQoL questionnaire is 0.18 points (95% CI = -0.26; 0.63) which shows that there is no systematic bias between the two administrations of the questionnaire because the confidence interval contains zero. The mean difference scores in the complete sample for the 7 domains are not significant (95% CI contains zero) for domains 2, 3, 4, 5, 6 and 7, once again indicating the absence of systematic bias. One domain in the complete sample does have a small but significant mean difference score, namely domain 1 [0.86 points (0.04; 1.68)], indicating the presence of a very

Table 4. SEM and SDC for individual studies.

		Belgium (Dutch)	Belgium (French)	Brazil	Czech Republic	England	Lithuania	Greece	Poland	Spain
SEM	D1	6.57	6.50	3.69	7.02	9.48	0.54	3.04	2.61	3.08
	D2	6.13	8.26	3.63	6.91	2.89	0.68	4.41	1.19	5.14
	D3	7.81	10.59	1.70	10.05	6.37	1.57	7.09	1.69	4.14
	D4	3.75	5.75	3.16	4.65	4.77	0.53	3.82	1.97	3.28
	D5	7.38	8.07	2.45	4.65	6.30	0.54	6.36	2.92	2.51
	D6	14.70	12.98	7.09	10.68	12.14	0	7.29	0.00	7.52
	D7	16.26	20.85	0.00	5.72	7.74	2.50	10.12	3.23	4.05
	Overall	2.54	4.06	2.17	2.86	4.20	0.18	3.34	1.07	1.73
SDC ind	D1	18.21	18.30	10.24	19.45	26.28	1.49	8.41	7.23	8.54
	D2	16.99	22.79	10.07	19.15	8.00	1.89	12.22	4.67	14.26
	D3	21.71	29.21	4.71	27.86	17.65	4.35	19.65	7.43	11.46
	D4	16.15	10.40	8.75	12.89	13.23	1.47	10.60	5.46	9.10
	D5	20.46	22.27	6.79	12.89	17.47	1.51	17.62	8.11	6.96
	D6	40.76	35.98	19.64	29.61	33.65	0	20.22	0.00	20.85
	D7	45.07	29.43	0.00	15.85	21.45	6.94	28.05	8.95	11.21
	Overall	7.05	11.34	6.00	7.92	11.65	0.49	9.24	2.96	4.81
SDC group	D1	3.57	3.40	2.95	2.81	8.31	0.20	1.19	1.32	2.21
	D2	3.33	4.23	2.91	2.76	2.53	0.25	1.73	0.85	3.68
	D3	4.26	5.42	1.36	4.02	5.58	0.57	2.78	1.36	2.96
	D4	3.17	1.93	2.53	1.86	4.18	0.19	1.50	1.00	2.35
	D5	4.01	4.14	1.96	1.86	5.53	0.20	2.49	1.48	1.80
	D6	7.99	6.68	5.67	4.27	10.64	0	2.86	0.00	5.38
	D7	8.84	5.47	0.00	2.29	6.78	0.91	3.97	1.63	2.90
	Overall	1.38	2.11	1.73	1.14	3.68	0.06	1.31	0.54	1.24

SEM: standard error of measurement; SDC_{ind}: smallest detectable change for individual subjects; SDC_{group}: smallest detectable change for groups

https://doi.org/10.1371/journal.pone.0216065.t004

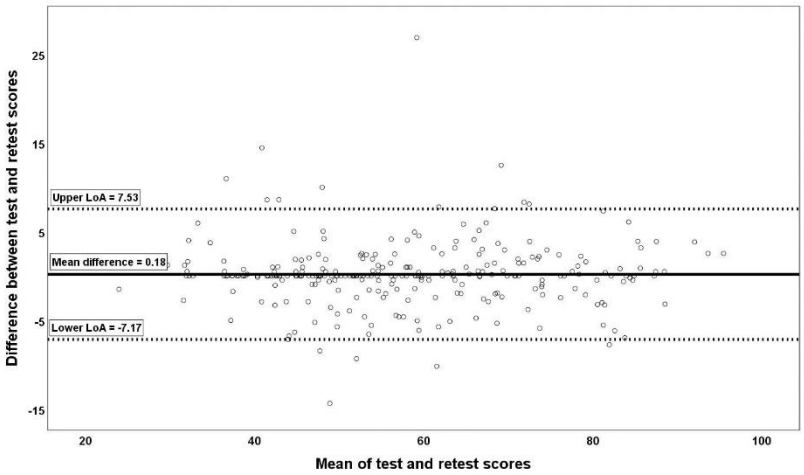


Fig 1. Bland Altman plot for the SarQoL overall score in the complete sample.

https://doi.org/10.1371/journal.pone.0216065.g001

slight systematic error. The full results of the Bland-Altman analysis are detailed in Table 3. A Bland-Altman plot for the Overall score in the complete sample is provided as Fig 1.

Discussion

In this study, values were obtained for the standard error of measurement and the smallest detectable change of the SarQoL questionnaire in a sample of 278 sarcopenic subjects hailing from 8 different countries and 9 different language-groups. The measurement error inherent to the questionnaire was found to be 2.65 points, and the minimum change needed to be confident that a real change in overall quality of life has occurred for an individual patient was 7.35 points. Systematic bias was further investigated with the method of Bland & Altman, and showed that there is no systematic bias for almost all domains (with domain 1 as the exception) and the overall score of the SarQoL questionnaire.

The SEM for the Overall score of the SarQoL questionnaire of 2.65 points represents 2.65% of the possible range of the Overall score (0–100) and 3.81% of the observed range of the SarQoL scores in the complete sample (min = 24.74; max = 94.22; range = 69.48).

This value for the standard error of measurement compares favorably with SEMs for the SF-36, the most frequently used quality of life questionnaire in sarcopenic populations. Hart found a SEM of 4 points for the Physical Component Summary (PCS–range: 0–100 points) and the Mental Component Summary (MCS–range: 0–100 points) of the SF-36 in a population of 68 subjects with a variety of orthopedic impairments [25] and Palmer calculated a SEM of 3.09 points for the PCS and 5.57 points for the MCS in a population of 233 subjects with joint hypermobility [26]. Other studies looked at the SEM for the 8 domains of the SF-36 (all range between 0–100 points), and found SEMs between 8.82 and 34.52 points in 106 women undergoing surgery for breast cancer [27], between 13.2 and 44.7 points in 92 subjects with neck pain [28], between 6.82 and 11.22 points for 628 subjects undergoing foot or ankle surgery [29], and between 11 and 32 points for 515 subjects undergoing orthopedic surgery [30]. While these have been calculated in populations that differ from ours, they show a trend for higher standard errors of measurement compared to the SarQoL questionnaire.

The SDC of the Overall score (7.35 points) of the SarQoL questionnaire is similar to the SDC found for the PCS and MCS of the SF-36. Palmer obtained SDCs of 8.56 points for the PCS and 15.44 points for the MCS, while Hart found SDCs of 9 points both for the PCS and MCS [25,26].

The results for the 7 domains of the SarQoL questionnaire in the complete sample show considerably higher SEM and SDC values compared to the Overall score. These values seem to correspond roughly to the number of items in each domain. When looking at the 3 domains with the least number of items (D6: 2 items; D3: 3 items; D7: 4 items), the largest SEM and SDC values are found, between 6.89 and 9.22 points for the SEM and between 19.09 and 25.51 points for the SDC. This contrasts with the 4 domains with larger numbers of items (D1: 8 items, D2: 9 items; D4: 14 items; D5: 15 items) which have SEM-values between 3.71 and 5.11 points and SDC-values between 10.27 and 14.17 points. It is not surprising that a domain score based on a larger number of items has greater precision and lower variability, represented by the standard deviation of the difference between test and retest scores.

The detailed breakdown of the SEM and SDC values obtained for the individual studies included in the analysis demonstrates the fact that the SEM and SDC depend on the population in which they are calculated. There is considerable variability between the studies, but not within the studies (i.e. studies with lower or higher SEM and SDC values are so for all the domains and the Overall score, and do not report low values for one domain and high for another). On the lower end are found the studies carried out in Lithuania, Poland and Spain,

in the middle those carried out in Belgium (Dutch), Brazil and the Czech Republic and on the higher end those carried out in Greece, England and Belgium (French). We were unable to formulate convincing hypotheses that could begin to explain why certain studies reported lower or higher values for SEM and SDC based on the clinical or study characteristics. It is likely that the observed variation is just the manifestation of the fact that the SEM and SDC are specific to the population in which they have been measured.

The Bland-Altman analysis, detailed in Table 3 and visually represented for the Overall score in Fig 1, shows that a very small systematic bias exists in only one domain. It is unlikely that this systematic bias is clinically relevant because of its small confidence interval and the fact that the lower end of the interval is extremely close to zero (95% CI = 0.04; 1.68). These results mean that clinicians and researchers can have confidence when administering the questionnaire that the results will not be distorted by systematic bias.

The analysis of the test-retest reliability in the complete sample confirmed the results from previous validation studies. The significantly larger sample in the combined analysis means that the confidence intervals found are much narrower than has been obtained previously. These results should inspire confidence that the SarQoL questionnaire is a reliable instrument.

Strengths and limitations

The main strength of this study is the fact that we were able to assemble a relatively large and heterogeneous sample (n = 278) of sarcopenic participants. This has the important advantage that the values calculated for the SEM and SDC are not dependent on a particular population, and could thus be more confidently used as a benchmark in future studies. The studies included in the analysis used different diagnostic criteria and instruments to establish sarcopenia. This is an advantage in this particular situation because the SEM and SDC values found in this study are not specific to a single definition of sarcopenia, but should be valid for different diagnostic criteria for sarcopenia, measured with different instruments. By combining multiple samples that differ with regards to clinical characteristics, we were able to find a middle ground and values for the SEM and SDC that are not highly specific to a single population. The sample size, which would be very difficult to gather in a single study, increased the accuracy of the standard deviation of the difference between test and retest score. Given that this parameter is key in the calculation of the SEM and SDC, the accuracy of these two parameters was enhanced by the large sample size. Because the SarQoL questionnaire has undergone validation in multiple languages, we were able to use test-retest data to calculate the SEM and the SDC, which is the preferred method because it takes into account biological variation, change of mood or concentration and other circumstances [18]. Since the data on which this study was based incorporates these elements and their subsequent influence on the SarQoL score, they have greater credibility than if other methods for calculating the SEM and SDC were to have been used.

There are, however, also limitations to this study. Although the researchers who carried out the individual translation and validation studies received the same guidance on the preferred design and conduct of these studies, local circumstances sometimes led them to deviate with regards to measurement of sarcopenia components (muscle mass, muscle strength and physical performance). Therefore, the methods for establishing the presence of sarcopenia are not standardized. This could, however, also be regarded as an opportunity in that we have a mix of subjects in the combined sample that represent a spectrum of methods and instruments. Secondly, because of the original purpose of the included studies, only the SarQoL questionnaire was administered twice, to calculate the test-retest reliability. It would have been preferable to compare the SEM and SDC of the SarQoL questionnaire to values for the SF-36 and the EQ-

5D measured in the same populations. But, since this data does not exist, we feel that a comparison to data from the literature was the second-best option and does provide a valid frame of reference.

Conclusion

The current study, which analyzed a sample of 278 subjects from 9 validation studies, obtained a standard error of measurement of 2.65 points and a smallest detectable change of 7.35 points for the Overall score of the SarQoL questionnaire. These values can be applied in future longitudinal research to evaluate the veracity of measured changes.

Supporting information

S1 Table. One-way Anova (Tukey) for age.
(PDF)

S2 Table. One-way Anova (Tukey) for BMI.
(PDF)

S3 Table. One-way Anova (Tukey) for number of drugs.
(PDF)

S4 Table. One-way Anova (Tukey) for number of concomitant illnesses.
(PDF)

S5 Table. Chi-squared test for gender [n(%)].
(PDF)

Author Contributions

Conceptualization: Charlotte Beaudart, Jean-Yves Reginster, Olivier Bruyère.

Data curation: Anton Geerinck.

Formal analysis: Anton Geerinck.

Investigation: Vidmantas Alekna, Charlotte Beaudart, Ivan Bautmans, Cyrus Cooper, Fabiana De Souza Orlandi, Jerzy Konstantynowicz, Beatriz Montero-Errasquín, Eva Topinková, Maria Tsekoura.

Methodology: Anton Geerinck, Charlotte Beaudart, Jean-Yves Reginster, Olivier Bruyère.

Writing – original draft: Anton Geerinck.

Writing – review & editing: Anton Geerinck, Vidmantas Alekna, Charlotte Beaudart, Ivan Bautmans, Cyrus Cooper, Fabiana De Souza Orlandi, Jerzy Konstantynowicz, Beatriz Montero-Errasquín, Eva Topinková, Maria Tsekoura, Jean-Yves Reginster, Olivier Bruyère.

References

1. Cruz-Jentoft A, Bahat G, Bauer J, Boire Y, Bruyère O, Cederholm T, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing. 2018;0: 1–16. https://doi.org/10.1093/cercor/bhw393
2. Cruz-Jentoft AJ, Landi F, Schneider SM, Zuniga C, Arai H, Boirie Y, et al. Prevalence of and interventions for sarcopenia in ageing adults: A systematic review. Report of the International Sarcopenia Initiative (EWGSOP and IWGS). Age Ageing. 2014; 43: 48–759. https://doi.org/10.1093/ageing/afu115 PMID: 25241753

3. Shafiee G, Keshtkar A, Soltani A, Ahadi Z, Larijani B, Heshmat R. Prevalence of sarcopenia in the world: a systematic review and meta- analysis of general population studies. J Diabetes Metab Disord. Journal of Diabetes & Metabolic Disorders; 2017; 16: 21.
4. Beaudart C, Locquet M, Reginster JY, Delandsheere L, Petermans J, Bruyère O. Quality of life in sarcopenia measured with the SarQoL: impact of the use of different diagnosis definitions. Aging Clin Exp Res. Springer International Publishing; 2018; 30: 307–313. https://doi.org/10.1007/s40520-017-0866-9 PMID: 29197020
5. Ethgen O, Beaudart C, Buckinx F, Bruyère O, Reginster JY. The Future Prevalence of Sarcopenia in Europe: A Claim for Public Health Action. Calcif Tissue Int. 2017; 229–234.
6. Beaudart C, Zaaria M, Pasleau F, Reginster J-Y, Bruyère O. Health outcomes of sarcopenia: a systematic seview and meta-analysis. PLoS One. 2017; 12: e0169548. https://doi.org/10.1371/journal.pone.0169548 PMID: 28095426
7. Beaudart C, Biver E, Bruyère O, Cooper C, Al-Daghri N, Reginster JY, et al. Quality of life assessment in musculo-skeletal health. Aging Clin Exp Res. Springer International Publishing; 2017;0: 1–6. https://doi.org/10.1007/s40520-017-0794-8 PMID: 28664458
8. Beaudart C, Biver E, Reginster J-Y, Rizzoli R, Rolland Y, Bautmans I, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. Age Ageing. England; 2015; 44: 960–966. https://doi.org/10.1093/ageing/afv133 PMID: 26433796
9. Beaudart C, Biver E, Reginster J-Y, Rizzoli R, Rolland Y, Bautmans I, et al. Validation of the SarQoL, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle. 2017; 8: 238–244. https://doi.org/10.1002/jcsm.12149 PMID: 27897430
10. Beaudart C, Edwards M, Moss C, Reginster JY, Moon R, Parsons C, et al. English translation and validation of the SarQoL, a quality of life questionnaire specific for sarcopenia. Age Ageing. 2017; 46: 271–277. https://doi.org/10.1093/ageing/afw192 PMID: 27789428
11. Ildiko GA, Mihai G, Beaudart C, Bruyère O, Pop R-M, Reginster J-Y, et al. Psychometric performance of the Romanian version of the SarQoL, a health-related quality of life questionnaire for sarcopenia. Arch Osteoporos. Archives of Osteoporosis; 2017; 12: 103. https://doi.org/10.1007/s11657-017-0397-1 PMID: 29139012
12. Geerinck A, Scheppers A, Beaudart C, Bruyère O, Vandenbussche W, Bautmans R, et al. Translation and validation of the Dutch SarQoL, a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact. 2018; 18: 463–472. PMID: 30511950
13. Konstantynowicz J, Abramowicz P, Glinkowski W, Taranta E, Marcinowicz L, Dymitrowicz M, et al. Polish Validation of the SarQoL, a Quality of Life Questionnaire Specific to Sarcopenia. J Clin Med. 2018; 7: 323. https://doi.org/10.3390/jcm7100323 PMID: 30287728
14. Tsekoura M, Billis E, Gliatis J, Tsepis E, Matzaroglou C, Sakkas GK, et al. Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. Disabil Rehabil. Taylor & Francis; 2018;0: 1–7. https://doi.org/10.1080/09638288.2018.1514076 PMID: 30453790
15. Reginster JY, Cooper C, Rizzoli R, Kanis JA, Appelboom G, Bautmans I, et al. Recommendations for the conduct of clinical trials for drugs to treat or prevent sarcopenia. Aging Clin Exp Res. 2016; 28: 47–58. https://doi.org/10.1007/s40520-015-0517-y PMID: 26717937
16. Terwee C, Bot S, de Boer M, van der Windt D, Knol D, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007; 60: 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012 PMID: 17161752
17. Harvill LM, Tennessee E, Uniuersit S. Standard Error of Measurement. Educ Meas Issues Pract. 1991; 10: 33–41. https://doi.org/10.1177/0272989X10380925
18. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurements in Medicine. Cambridge University Press; 2011. https://doi.org/10.1186/1741-7015-9-29
19. Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. Lancet. 1986; 327: 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8
20. Polit D. Getting serious about test-retest reliability: A critique of retest research and some recommendations. Qual Life Res. 2014; 23: 1713–1720. https://doi.org/10.1007/s11136-014-0632-9 PMID: 24504622
21. Bland J.M. & Altman DG. Measuring agreement in method comparison studies. Stat. Meth. 1999; 2802: 135–160.
22. IBM Support. The Bland-Altman Plot [Internet]. [cited 15 Mar 2019]. Available: https://www-01.ibm.com/support/docview.wss?uid=swg21476730
23. Michalkova H, Ryznarova I, Seget J, Topinková E. Health-related quality of life in sarcopenia: translation and validation of the Czech version of the SarQoL instrument. European Geriatric Medicine. 2017. p. s104. https://doi.org/10.1128/AEM.68.2.952–956.2002

24. Lee RC, Wang Z, Heo M, Ross R, Jannsen I, Heymsfield SB. Total-body skeletal muscle mass: Development and cross-validation of anthropometric prediction models. Am J Clin Nutr. 2000; 72: 796–803. https://doi.org/10.1093/ajcn/72.3.796 PMID: 10966902

25. Hart DL. Test-Retest Reliability of an Abbreviated Self-Report Overall Health Status Measure. J Orthop Sport Phys Ther. 2003; 33: 734–744. https://doi.org/10.2519/jospt.2003.33.12.734 PMID: 14743987

26. Palmer S, Manns S, Cramp F, Lewis R, Clark EM. Test-retest reliability and smallest detectable change of the Bristol Impact of Hypermobility (BIoH) questionnaire. Musculoskelet Sci Pract. Elsevier Ltd; 2017; 32: 64–69. https://doi.org/10.1016/j.msksp.2017.08.007 PMID: 28881227

27. Oliveira IS, Costa LCM, Manzoni ACT, Cabral CMN. Assessment of the measurement properties of quality of life questionnaires in Brazilian women with breast cancer. Brazilian J Phys Ther. 2014; 18: 372–383. https://doi.org/10.1590/bjpt-rbf.2014.0045 PMID: 25075998

28. Juul T, Søgaard K, Davis AM, Roos EM. Psychometric properties of the Neck OutcOme Score, Neck Disability Index, and Short Form–36 were evaluated in patients with neck pain. J Clin Epidemiol. Elsevier Inc; 2016; 79: 31–40. https://doi.org/10.1016/j.jclinepi.2016.03.015 PMID: 27038852

29. Dawson J, Boller I, Doll H, Lavis G, Sharp R, Cooke P, et al. Minimally important change was estimated for the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. J Clin Epidemiol. Elsevier Inc; 2014; 67: 697–705. https://doi.org/10.1016/j.jclinepi.2014.01.003 PMID: 24709030

30. Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. Health Qual Life Outcomes. 2008; 6: 1–12. https://doi.org/10.1186/1477-7525-6-1

Discriminative power of the SarQoL with the EWGSOP2 sarcopenia criteria

The Journal of Frailty & Aging

J Frailty Aging 2020;
Published online September 15, 2020, http://dx.doi.org/10.14283/jfa.2020.47

LETTER TO THE EDITOR

DISCRIMINATIVE POWER OF THE SARCOPENIA QUALITY OF LIFE (SARQOL®) QUESTIONNAIRE WITH THE EWGSOP2 CRITERIA

Dear Editor,

The Sarcopenia Quality of Life (SarQoL®) questionnaire was developed in 2015 to fill the need for a specific instrument to measure quality of life in sarcopenia. Since then, its validity and reliability have been evaluated in multiple languages, and it is now available in 30 language-specific versions. In multiple validation studies, the SarQoL® has demonstrated its ability to discriminate between sarcopenic and non-sarcopenic subjects when diagnosed according to the EWGSOP criteria (1). However, these criteria have now been updated, and the discriminative power of the SarQoL® questionnaire should be reaffirmed using the EWGSOP2 criteria (2). The analysis presented below aims to establish whether the SarQoL® questionnaire can discriminate between sarcopenic, probably sarcopenic (low grip strength in the EWGSOP2 algorithm) and non-sarcopenic participants.

This study used data gathered from older, community-dwelling volunteers recruited within the framework of the Sarcopenia and Physical Impairment with advancing Age (SarcoPhAge) cohort (3). The same data was used in the original validation study of the SarQoL® questionnaire (4). The sarcopenia components of muscle mass and muscle strength were measured with, respectively, dual-energy x-ray absorptiometry and a hydraulic hand dynamometer. We applied the thresholds specified by the EWGSOP2 for appendicular lean mass divided by height-squared (ALM/Ht²: less than 5.5 kg/m² for women and 7 kg/m² for men) and handgrip strength (less than 16 kg for women and 27 kg for men) (2). Quality of life was measured with the SarQoL® questionnaire, which provides an overall QoL score and 7 domain scores for specific aspects of QoL, all between zero (worst QoL) and 100 (best QoL). In line with the case-finding algorithm elaborated by the EWGSOP2, we considered participants to have "probable sarcopenia" when they demonstrated low grip strength, and sarcopenia when both low grip strength and low muscle mass were present (2).

In total, 296 participants, with a median age of 73.3 (68.9-78.6) years, were included in this analysis. In a previous analysis, 43 subjects were diagnosed as sarcopenic with the EWGSOP criteria (4). As expected, we found a lower prevalence of sarcopenia when applying the EWGSOP2 criteria, with 38 participants displaying low grip strength, of which 13 were ultimately considered sarcopenic.

Sarcopenic participants, as diagnosed with EWGSOP2 criteria, had significantly lower scores for all 7 SarQoL® QoL domains (all $p<0.05$) and the overall QoL score of the SarQoL® questionnaire [45.83 (38.62-60.26) versus 66.43 (56.10-78.26); $p<0.001$], indicating that the SarQoL® questionnaire can discriminate between sarcopenic and non-sarcopenic individuals. When the sample was categorized in probably sarcopenic (n=38) and non-sarcopenic (n= 258), similar results were obtained. All 7 domain scores of the SarQoL® questionnaire were significantly lower (all $p<0.05$) for probably sarcopenic participants, as well as the Overall QoL score [53.24 (41.18-63.24) versus 67.74 (57.35-79.02); $p<0.001$]. Detailed results are presented in table 1.

We investigated the robustness of these results by carrying out binary logistic regression analyses including age, gender, body mass index, n° of comorbidities and n° of medications as covariates. We found that for every one-unit increase in Overall QoL, we expect to see a 10% decrease in the odds of belonging to the EWGSOP2 sarcopenic group (OR: 0.90; 95% CI: 0.85-0.95), and a 6% decrease in the odds of belonging to the EWGSOP2 probable sarcopenia group (OR: 0.94; 95% CI: 0.90-0.97).

Table 1
Discriminative power of the SarQoL® questionnaire using the EWGSOP2 criteria for sarcopenia

	EWGSOP2 sarcopenia				EWGSOP2 probable sarcopenia			
	Not sarcopenic (n=283)	Sarcopenic (n=13)	p [a]	OR (95% CI) [b]	Not sarcopenic (n= 258)	Probably sarcopenic (n=38)	p [a]	OR (95% CI) [b]
(1) Physical and mental health	63.33 (54.43 - 76.67)	55.57 (38.33 - 60.55)	0.006	0.94 (0.89 - 0.98)	63.33 (55.57 - 76.67)	55.57 (45.57 - 63.33)	<0.001	0.96 (0.93 - 0.99)
(2) Locomotion	61.11 (50.00 - 83.33)	30.56 (25.00 - 62.50)	0.004	0.96 (0.93 - 0.99)	61.11 (50.00 - 86.11)	50.00 (27.78 - 61.81)	<0.001	0.97 (0.95 - 0.99)
(3) Body composition	60.00 (50.00 - 70.00)	50.00 (29.58 - 54.17)	0.001	0.92 (0.88 - 0.97)	60.00 (50.00 - 70.83)	50.00 (40.00 - 60.00)	0.003	0.98 (0.95 - 1.00)
(4) Functionality	73.21 (60.71 - 84.62)	53.85 (42.58 - 75.21)	0.002	0.95 (0.91 - 0.98)	75.00 (62.50 - 85.71)	57.69 (45.67 - 72.32)	<0.001	0.96 (0.93 - 0.99)
(5) Activities of daily living	63.33 (51.67 - 80.00)	40.00 (30.51 - 50.00)	0.001	0.92 (0.86 - 0.96)	65.00 (52.98 - 80.00)	48.03 (34.96 - 58.75)	<0.001	0.95 (0.93 - 0.98)
(6) Leisure activities	58.31 (33.25 - 66.75)	33.25 (33.25 - 50.00)	0.008	0.95 (0.91 - 0.99)	66.69 (33.25 - 66.75)	33.25 (33.25 - 50.00)	<0.001	0.97 (0.94 - 0.99)
(7) Fears	87.50 (87.50 - 100.00)	75.00 (75.00 - 100.00)	0.045	0.94 (0.89 - 0.99)	87.50 (87.50 - 100.00)	87.50 (75.00 - 100.00)	0.044	0.98 (0.95 - 1.01)
Overall score	66.43 (56.10 - 78.26)	45.83 (38.62 - 60.26)	<0.001	0.90 (0.85 - 0.95)	67.74 (57.35 - 79.02)	53.24 (41.18 - 63.24)	<0.001	0.94 (0.90 - 0.97)

a. p-value calculated with Mann-Whitney U-test; b. Binary logistic regression adjusted for age, gender, BMI, n° of comorbidities and n° of medications

Received July 10, 2020
Accepted for publication August 25, 2020

The current analysis shows that the SarQoL® questionnaire retains its capacity to discriminate between sarcopenic and non-sarcopenic persons when using the EWGSOP2 criteria for sarcopenia, despite the reduced prevalence of sarcopenic individuals. These results reinforce the results found during the validation of the Lithuanian version of the SarQoL® questionnaire, which also found significantly lower QoL scores for all 7 domains and the overall QoL score between non-sarcopenic and EWGSOP2 sarcopenic participants. The odds ratio found for the overall QoL score in this study is nearly identical to our own, at 0.913 (95% CI 0.876-0.951) (5).

We also found that participants with low grip strength, categorized as probably sarcopenic in the EWGSOP2 algorithm, had significantly lower QoL scores for all 7 domains and the overall QoL score. This is an important finding because it shows that, when an older person is found to have low muscle strength, his or her quality of life is likely to already have been impacted. This adds strength to the EWGSOP argument that the observation of low grip strength in clinical practice could be a sufficient indication to put in place interventions to mitigate and improve a patient's musculoskeletal health.

The SarQoL® questionnaire is currently the only sarcopenia-specific QoL questionnaire, and has demonstrated to be able to discriminate between sarcopenic, probably sarcopenic and non-sarcopenic groups. Its use, in combination with the EWGOSP2 criteria, could provide greater detail and precision on the impact of sarcopenia on QoL.

Conflict of interest: J-YR, CB and OB are shareholders of SarQoL sprl. AG and ML have no conflicts of interest to declare.

A. Geerinck[1], M. Locquet[1], J.-Y. Reginster[1,2], O. Bruyère[1], C. Beaudart[1]

1. Division of Public Health, Epidemiology and Health Economics, World Health Organization Collaborating Centre for Public Health aspects of musculoskeletal health and ageing, University of Liège, Belgium; 2. Chair for Biomarkers of Chronic Diseases, Biochemistry Department, College of Science, King Saud University, Riyadh, KSA, Saudi Arabia.

Corresponding Author: A. Geerinck, Division of Public Health, Epidemiology and Health Economics, World Health Organization Collaborating Center for Public Health aspects of musculoskeletal health and ageing, University of Liège, Belgium, anton.geerinck@uliege.be

References

1. SarQoL publications. www.sarqol.org/en/publications.
2. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019;48:16–31
3. Beaudart C, Locquet M, Reginster JY, Delandsheere L, Petermans J, Bruyère O. Quality of life in sarcopenia measured with the SarQoL®: impact of the use of different diagnosis definitions. Aging Clin Exp Res 2018;30:307–313
4. Beaudart C, Biver E, Reginster J-Y, et al. Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 2017;8
5. Alekna V, Kilaite J, Tamulaitiene M, et al. Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). Eur Geriatr Med 2019

SarQoL and physical frailty

ORIGINAL ARTICLE

Journal of Cachexia, Sarcopenia and Muscle 2021; **12**: 319–330
Published online 28 February 2021 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/jcsm.12687

Evaluating quality of life in frailty: applicability and clinimetric properties of the SarQoL® questionnaire

Anton Geerinck[1]*, Médéa Locquet[1], Olivier Bruyère[1], Jean-Yves Reginster[1,2] & Charlotte Beaudart[1]

[1]*Division of Public Health, Epidemiology and Health Economics, University of Liège, World Health Organization Collaborating Centre for Public Health aspects of musculo-skeletal health and ageing, Liège, Belgium;* [2]*Biochemistry Department, College of Science, King Saud University, Riyadh, Saudi Arabia*

Abstract

Background The SarQoL® questionnaire was specifically designed to measure quality of life (QoL) in sarcopenia. Frailty and sarcopenia have areas of overlap, notably weak muscle strength and slow gait speed, which may mean that the SarQoL could provide a measure of QoL in frailty. This study aimed to evaluate the clinimetric properties of the SarQoL questionnaire in physical frailty using the Fried criteria.

Methods Analyses were carried out on data from the Sarcopenia and Physical impairment with advancing Age study. Frailty was assessed with the Fried criteria and QoL with the SarQoL, the Short-Form 36-Item, and the EuroQoL 5-Dimension (EQ-5D) questionnaires. We evaluated discriminative power (with the Kruskal–Wallis analysis of variance test), internal consistency (with Cronbach's alpha), construct validity (through hypotheses testing), test–retest reliability (with the intraclass correlation coefficient), measurement error (calculating standard error of measurement and smallest detectable change), and responsiveness (through hypotheses testing and standardized response mean).

Results In total, 382 participants were included for the validation and 117 for the responsiveness evaluation. They had a median age of 73 (69–79) years, took 5 (3–8) drugs, and had 4 (3–5) co-morbidities. There were more women (n = 223; 58.4%) than men and, in total, 172 (45%) robust, 167 (44%) pre-frail, and 43 (11%) frail participants. Discriminative power was confirmed when significantly lower ($P < 0.001$) overall SarQoL scores, and thus also worse QoL, were observed between robust [77.1 (64.35–85.90)], pre-frail [62.54 (53.33–69.57)], and frail [49.99 (40.45–56.06)] participants. Six of the SarQoL domains performed likewise, with significantly lower scores according to frailty status with Domain 7 (fears) being the exception. Internal consistency was good (α = 0.866). Convergent (using Short-Form 36-Item and EQ-5D) and divergent construct validity (using EQ-5D) was confirmed. Test–retest reliability was excellent [intraclass correlation coefficient = 0.918 (0.834–0.961)], with a standard error of measurement of 3.88 and a smallest detectable change of 10.76 points. We found moderate responsiveness when five of the nine hypotheses were confirmed, coupled with a large effect size for the overall SarQoL score (corrected standardized response mean of −1.44).

Conclusions The SarQoL questionnaire has adequate clinimetric properties for use with frail patients in clinical practice and trials and could provide data that are more appropriate and detailed than the generic questionnaires currently used.

Keywords Frailty; Quality of life; Clinimetrics; Psychometrics; Patient-reported outcome measure

Received: 16 June 2020; Revised: 18 December 2020; Accepted: 26 January 2021
**Correspondence to: Anton Geerinck, CHU—Sart Tilman—Quartier Hôpital, Avenue Hippocrate 13 (Bât. B23), 4000 Liège, Belgium. Phone: +32 43 66 20 88, Email: anton.geerinck@uliege.be*

Introduction

The World Health Organization declared the period from 2020 to 2030 to be the decade of healthy ageing, which they define as 'the process of developing and maintaining the functional ability that enables wellbeing in older age'.[1] This concept is closely linked to the syndrome of frailty, a clinically recognizable state of increased vulnerability in older people, caused by age-related losses in physiological reserves and function across multiple organ systems, such that the ability to cope with everyday or acute stressors is compromised.[2]

This state of increased vulnerability is associated with negative health outcomes, as evidenced by a recent meta-analysis, which found an increased likelihood of premature mortality, hospitalization, and institutionalization.[3] Frailty was also associated with an increased risk for developing disability in both basic and instrumental activities of daily living, an increased risk for physical limitations, dependency, falling, fractures, cognitive decline, decline in lean body mass, and lower life satisfaction.[3] These outcomes, in combination with an estimated prevalence of 10.7–18%, mean that frailty represents an important burden on public health.[4–6]

While hard outcomes such as mortality and hospitalizations remain the primary indicators in research settings, outcomes measuring the subjective experience of patients are becoming as essential part of the arsenal. Health-related quality of life is one of the main patient-reported outcome measures used in research, and several studies have already focused on quality of life (QoL) in frailty in the last decade. A 2019 systematic review listed 22 studies that assessed QoL in frailty and which demonstrated that frail participants had worse QoL than robust participants. However, these differences between frail and robust people were only clear for the sub-concepts of physical functioning and satisfaction with life. For social and environment scales, results were inconsistent between the different questionnaires used, limiting their usefulness in assessing the psychosocial well-being pre-frail and frail individuals. In this systematic review, the Short-Form 36-Item (SF-36) was the most frequently used instrument out of the 14 instruments included, followed by the WHOQOL-BREF, the CASP-19, and the EUROHIS-QOL.[7] Several observations can be made from the results of this systematic review. First, the SF-36, which was the most frequently used instrument to measure QoL in frailty, is a generic instrument and not adapted to specific populations or diseases.[8] While generic instruments allow QoL to be compared between a range of populations, specific instruments often possess better construct validity and are more sensitive to changes in QoL over time.[9] Secondly, the concept of QoL and the components needed to provide a holistic assessment were interpreted differently between each of the QoL questionnaires. While some concepts from the generic QoL questionnaires mentioned previously are shared with the sarcopenia quality of life (SarQoL®) questionnaire (i.e. physical and mental health and activities of daily living), others such as 'body composition', 'leisure activities', and 'fears' are unique.

The systematic review did not include frailty-specific QoL instruments. A QoL instrument specific to the frailty syndrome might improve sensitivity to change in disease-specific QoL over time in this group.[10]

One such specific questionnaire is the SarQoL questionnaire, developed in 2015 with the aim of measuring health-related QoL in sarcopenic persons.[11] The questionnaire was constructed using input from experts, literature review, and crucially, interviews with older, sarcopenic individuals. It has been validated for use with sarcopenic, older, community-dwelling participants in multiple languages and has consistently been shown to be a valid and reliable instrument, as well as responsive to changes in QoL.[12–21]

Multiple authors have argued that the conceptual frameworks of frailty and age-related sarcopenia overlap substantially, notably on the similar clinical manifestations used to diagnose the two conditions. The slowness indicator in the Fried criteria for frailty and the low gait speed indicator used to characterize muscle function in sarcopenia are one area of overlap between the two conditions. Partial overlap exists between weight loss in frailty and muscle loss in sarcopenia, and fatigue/exhaustion in frailty and grip strength in sarcopenia. Some have argued that sarcopenia is equivalent to the physical component of frailty, separate from the cognitive, psychological, sociological, and spiritual components of frailty.[22–24]

Because of the overlap between sarcopenia and physical frailty, we considered it worthwhile to explore whether the SarQoL questionnaire could be used in the assessment of QoL in frail and pre-frail individuals, as diagnosed with the Fried criteria. This study aims to examine the clinimetric properties of the SarQoL questionnaire in robust, pre-frail, and frail participants of the Sarcopenia and Physical impairment with advancing Age (SarcoPhAge) study.

Methods

Population

The analyses described in this manuscript have been carried out using the data collected during the SarcoPhage study. This cohort study followed a sample of community-dwelling older people for 5 years and has been described in multiple publications [25–29]. In brief, the SarcoPhAge study recruited a convenience sample of volunteers aged 65 years or older living in the Liège province of Belgium. Participants were recruited from different departments of an outpatient clinic in Liège, as well as through advertisement in the local press. Candidates were not eligible for inclusion in the cohort if they presented with a body mass index (BMI) >50 kg/m^2 or if they

had one or more amputated limbs. No other exclusion criteria were applied. Participants were invited to the research centre once yearly, where they performed physical tests and completed questionnaires.[25] For the analyses presented here, we used data from Year 1 of follow-up, except for the evaluation of the responsiveness of the questionnaire, where we used data from the visits carried out at 1 and 5 years into the study.

Frailty evaluation

In the SarcoPhAge sample, physical frailty was evaluated with the criteria described by Fried *et al*.[30] The Fried diagnostic criteria evaluate five items to determine whether a person is considered to be robust, pre-frail, or frail. In this study, the five criteria were measured with the following instruments: weakness was present if handgrip strength measured with hydraulic dynamometer was below the cut-offs based on gender and BMI, low gait speed was detected by evaluating usual walking speed on a 4 m track (results corrected to 4.5 m track) with cut-offs based on gender and height, low physical activity was measured with the Minnesota Leisure Time Activity Questionnaire[31] using gender-specific cut-offs for kilocalories used in physical activity in the preceding week, exhaustion was established using two items from the Center for Epidemiological Studies Depression scale,[32] and weight loss was detected through a self-reported question on unintentional weight loss of more than 4.5 kg in the past year.[25,30] For each item, participants were given 1 point if below the cut-off, and 0 if not, and these item scores were summed for a frailty score between 0 and 5. Participants with zero points were considered robust, a score of 1 or 2 points indicated a pre-frail state, and subjects with a score of 3 or more points were considered to be frail. A detailed description of the criteria, instruments, and cut-off values is provided in Supporting Information, *Table* S1.

Quality of life measurement

The SarQoL questionnaire, the focus of this validation study, is a patient-reported outcome measure specifically designed to evaluate QoL in older, sarcopenic, community-dwelling people. There are 55 items in the questionnaire, categorized into seven domains of health-related dysfunction: (i) physical and mental health, (ii) locomotion, (iii) body composition, (iv) functionality, (v) activities of daily living, (vi) leisure activities, and (vii) fears. A score between 0 (worst QoL) and 100 (best QoL) is provided for each domain, and an overall QoL score (range: 0–100 points) is calculated on the entirety of the questionnaire.[11] The scoring algorithm is not publicly available, but tools to calculate the scores are available upon request via info@sarqol.org or via the website www.sarqol.org and free for non-sponsored research. The questionnaire is self-reported and takes about 15 min to complete. The SarQoL questionnaire has been validated in multiple languages and has been shown to be a valid and reliable instrument.[13,15–19,33] The questionnaire was shown to be responsive to changes in QoL in a sample of 42 sarcopenic subjects followed over 3 years, and its standard error of measurement (SEM) and smallest detectable change have been calculated in different European populations as well as pooled.[20,21]

Complementary to the SarQoL questionnaire, two generic QoL questionnaires were also completed by each participant to allow the evaluation of the construct validity of the SarQoL questionnaire. The first of these, the SF-36 questionnaire, measures functional health and well-being from the patient's perspective, providing eight domain scores (physical functioning, social functioning, role functioning physical, role functioning emotional, vitality, bodily pain, mental health, and general health) and two summary scores (physical and mental), all scored from 0 (worst QoL) to 100 (best QoL) points.[34] Secondly, the EuroQoL 5-Dimension 3-Level (EQ-5D-3L) and the associated visual analogue scale (EQ-VAS) were administered. The EQ-5D is a generic measure of health status, which records self-reported problems (none, some, and extreme) on five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression).[35] Results are reported as an index score (between 0 and 1, with 0 indicating death and 1 indicating perfect health) and a self-rated health evaluation (between 0, worst imaginable health, and 100, best imaginable health).[36]

Clinimetric properties

The measurement properties to be included in this validation were selected based on the COSMIN taxonomy and its related documentation.[37,38] These include known-groups validity (also known as discriminative power), internal consistency, construct validity (through hypotheses testing), reliability (test–retest), measurement error, and responsiveness. We also looked at the presence of floor and/or ceiling effects and provided the smallest detectable change to aid in the interpretation of the evolution of the SarQoL scores over time.

i Known-groups validity is based on the hypothesis that two or more groups with distinctive characteristics should logically differ in the construct that is measured.[39] In the context of this study, the hypothesis is that robust participants should have higher QoL scores than pre-frail and frail participants, which would mean that the SarQoL questionnaire can discriminate between the three frailty profiles.

ii Internal consistency quantifies the degree of interrelatedness between the items in the questionnaire, that is,

whether all items in the SarQoL measure the same underlying construct (QoL).[37,38]

iii Construct validity is used to assess whether the questionnaire under investigation actually measures what it theoretically aims to measure. This is performed by comparing the questionnaire with other questionnaires (or subscales of) that should, in theory, measure the same construct (convergent validity) or a different construct (divergent validity).[37,38] In this study, we utilized the same eight hypotheses on the strength of association between the overall QoL score of the SarQoL questionnaire and domains of the SF-36 and EQ-5D that were used in previous validations.[13–19,33]

iv The test–retest reliability of the questionnaire shows whether the scores measured by the SarQoL questionnaire remain stable between multiple administrations, on the condition that the participants' health state also remains stable.[37,38] To measure this, the SarQoL questionnaire was administered twice, with an approximate interval of 2 weeks in between, and participants provided information on the stability of their health. Because of the different objectives of the study that collected the data analysed in this article, only the 43 participants who were diagnosed as sarcopenic with the European Working Group on Sarcopenia in Older People criteria were invited, at the time of the original validation study, to participate in the retest part, with 30 providing usable data.[13] ReliabilityJCSM_12687 is also demonstrated by the SEM, which provides a measure of the dispersion of observed scores around the 'true' score from repeated measurements. The smallest detectable change provides the value for the minimum change in QoL scores that needs to be observed to be certain that the measured change in QoL is real and not possibly due to measurement error.[38]

v Floor and ceiling effects indicate that the range of the scale is too narrow and that extreme profiles cannot be accurately measured. They are present when $>$15% of the participants obtain either the highest or lowest score.

vi The last clinimetric property investigated was the responsiveness of the questionnaire, that is, its capacity to detect change over time, between the first and fifth years of the SarcoPhAge study.[21] We used the same methodology as in a previous evaluation of the responsiveness of the SarQoL, which was combination of hypothesis testing and effect size evaluation.[21] In short, we evaluated nine hypotheses (see *Table* 5) on the theorized strength of correlation between the changes observed with the SarQoL questionnaire between Year 1 and Year 5 and the changes observed with (domains of) the SF-36, EQ-5D, and EQ-VAS. The results were interpreted with the criteria from de Boer *et al.*, which indicate that a questionnaire has high responsiveness if at least 75% of hypotheses are confirmed, moderate when at least 50–75% are confirmed, and poor responsiveness when less than 50% are confirmed.[40] In this analysis, we included all participants for whom we had valid data at Year 1 and Year 5. For the second method, we calculated standardized response means (SRMs) (a measure of effect size), which reflect the magnitude of change measured by the SarQoL and by the other questionnaires used in this study. Larger effect sizes indicate that the questionnaire possesses better responsiveness.[41] Because this method is based on the assumption that a change in health status has occurred, we could only include those participants for whom we had valid data at Year 1 and Year 5 and whose frailty status changed in the years between evaluations. The change in frailty status is used here as a proxy measure of change in health status, and we hypothesize that a change in frailty status will be reflected in the observed change in QoL.

Statistical analysis

Normality of distribution for quantitative variables was tested with the Shapiro–Wilk test, by comparing mean and median and by evaluating the histogram and Q–Q plot. Continuous variables following a Gaussian distribution are reported as mean ± standard deviation, while those who do not are reported as median (25th–75th percentile). Nominal variables are reported as absolute (n) and relative (%) frequencies. The evaluation of differences between groups for nominal variables was carried out using Pearson's χ^2 test. All results were considered significant at 5% level ($P \leq 0.05$), except for pairwise comparisons between the robust, pre-frail, and frail groups, which were considered significant at $P \leq 0.017$ (P-value adjusted for the number of comparisons: α = 0.05/3). IBM SPSS Statistics Version 25 for Windows (IBM Corp., Armonk, NY) was used for all statistical manipulations.

i Analysis of continuous variables to determine the known-groups validity between the three frailty categories was carried out with the analysis of variance test if distributions in all groups were Gaussian and with the Kruskal–Wallis test if they were not. Paired comparisons were carried out with multinomial regression analysis, so as to obtain P-values for the differences between the robust, pre-frail, and frail groups.

ii Internal consistency was determined with Cronbach's alpha test, where a value between 0.70 and 0.95 indicates good internal consistency.[42]

iii Associations between two continuous variables (such as used in the hypotheses for evaluating construct validity and responsiveness) were examined with Pearson's or Spearman's correlations, depending on normality of distribution.

iv The test–retest reliability was quantified by calculating the intraclass correlation coefficients (ICCs) (two-way mixed model – absolute agreement type) between the scores from the first and the second administrations. ICCs greater than 0.7 indicate acceptable reliability.[38] The SEM was calculated by dividing the standard deviation of the difference between the scores from the first administration and those of the second administration by the square root of 2. This gives the following formula: SEM = ($SD_{(test\ score\ -\ retest\ score)}/\sqrt{2}$). The smallest detectable change is derived from the SEM value, by the following formula: 1.96 * √2 * SEM.

v Floor and ceiling effects were evaluated following inspection of the frequency tables.

vi Finally, SRMs, a measure of effect size and used to evaluate responsiveness, were calculated by dividing the mean difference between the SarQoL scores from the first year and the fifth year of the SarcoPhAge study by the standard deviation of the differences between these paired values. The SRM values were subsequently transformed with the formula SRM/ $\sqrt{2}/\sqrt{(1-r)}$, where '*r*' signifies the correlation between Year 1 and Year 5 scores.[43] The corrected SRM values can now be interpreted with the thresholds formulated by Cohen *et al.*, where SRM < 0.20 is trivial effect, $0.20 \leq$ SRM < 0.50 is a small effect, $0.50 \leq$ SRM < 0.80 is a moderate effect, and SRM ≥ 0.80 is considered a large effect.[44]

Results

Clinical characteristics

In total, 382 subjects were eligible for inclusion at the first follow-up visit of the SarcoPhAge study. These subjects had a median age of 73 (69–78) years old, were slightly overweight at a median BMI of 27 (24–30) kg/m^2, took a median of 5 (3–8) drugs, and had a median of 4 (3–5) co-morbidities. There were slightly more women (n = 223; 58.4%) than men in the sample. The median grip strength was 39 (33–45) kg for men and 21 (17.5–25) kg for women. Lastly, the median gait speed in the complete sample was 1.09 (0.91–1.28) m/s.

All 382 participants were evaluated for frailty with the Fried criteria, and we found 172 (45%) robust, 167 (44%) pre-frail, and 43 (11%) frail individuals. Clinical characteristics were significantly different between these three groups. Frail participants were older than pre-frail participants, who, in turn, were older than robust individuals ($P < 0.001$). The same dynamic was present for BMI (with frail participants having the highest BMI; $P < 0.001$), drug consumption (with frail participants taking the most drugs; $P < 0.001$), and co-morbidities (with frail participants having the highest number of co-morbidities; $P < 0.001$). As expected, frail participants had lower grip strength than pre-frail participants, who, in turn, had lower grip strength than robust people ($P < 0.001$ for men and women). The same was observed for gait speed (with frail participants having the lowest gait speed; $P < 0.001$). Detailed results and pairwise comparisons are available in *Table* 1.

Known-groups validity

The overall QoL score measured by the SarQoL questionnaire was significantly different ($P < 0.001$) between the three categories of frailty, following a downward trend with robust participants having the best QoL [77.10 (64.35–85.90)], followed by the pre-frail participants [62.54 (53.33–69.57)] and with the frail participants presenting with the worst QoL [49.99 (40.45–56.06)]. The differences between the overall QoL scores in these three groups were revealed to be significant in the paired comparisons (all $P < 0.001$).

The QoL scores for the seven domains of health-related quality of life in the SarQoL questionnaire were also shown to be highly significantly different ($P < 0.001$). An examination of the paired differences showed that only Domain 7 (fears) was not significantly different in the comparison between pre-frail and frail groups ($P = 0.119$).

The complete results of the known-groups validity are presented in *Table* 2.

Internal consistency

The homogeneity of the questionnaire was found to be excellent, with Cronbach's alpha of 0.866, at the upper end of the 0.70 to 0.95 range considered good. This shows that the questionnaire is consistent without showing the increased likelihood for redundancy associated with alpha values greater than 0.95. The influence of individual domains was tested by deleting a single domain at a time. The resulting alpha values ranged from 0.854 to 0.894, indicating that no domain unduly influences the internal consistency.

Construct validity

Two sets of hypotheses were examined: for the convergent construct validity, we theorized that the overall QoL score of the SarQoL questionnaire measures a construct related to the SF-36 physical functioning, role limitation due to physical problems, and vitality domains as well as to the EQ-5D utility score. We therefore expect to find moderate to strong correlations between the SarQoL and these domains. For the divergent construct validity, we theorized that the overall QoL score of the SarQoL questionnaire measures a different construct than the SF-36 role limitation due to emotional problems and mental health domains, as well as the self-care

Table 1 Clinical characteristics

	All (n = 382)	Robust (n = 172)	Pre-frail (n = 167)	Frail (n = 43)	P R-PF[a]	P R-F[b]	P PF-F[c]	P Overall[d]
Age (years)	73.18 (68.77–78.47)	70.76 (68.15–76.05)	73.79 (69.45–79.59)	76.52 (71.87–81.82)	<0.001	<0.001	0.031	<0.001
BMI (kg/m^2)	26.86 (23.74–30.12)	26.24 (23.4–28.68)	27 (23.82–30.43)	28.74 (24.67–34)	0.077	0.001	0.037	0.004
Drug consumption (n)	5 (3–8)	5 (3–6)	6 (4–8)	7 (5–10)	<0.001	<0.001	0.079	<0.001
Co-morbidities (n)	4 (3–5)	3 (2–5)	4 (3–6)	4 (3–7)	<0.001	<0.001	0.249	<0.001
Female, n (%)	223 (58.4%)	97 (56.4%)	98 (58.7%)	28 (65.1%)				0.580
Grip strength (kg)								
Men	39 (33–45)	40 (37–45)	38 (31–45)	26 (18–32)	0.012	<0.001	<0.001	<0.001
Women	21 (17.5–25)	24 (21–27.5)	19 (16–22)	13 (11.63–17.75)	<0.001	<0.001	<0.001	<0.001
Gait speed (m/s)	1.09 (0.91–1.28)	1.2 (1.05–1.35)	1.06 (0.86–1.23)	0.72 (0.53–0.87)	<0.001	<0.001	<0.001	<0.001

BMI, body mass index; QoL, quality of life.
All results reported as median (25th–75th percentile).
[a]P-values for pairwise comparison between robust and pre-frail groups (significant at $P \leq 0.017$).
[b]P-values for pairwise comparison between robust and frail groups (significant at $P \leq 0.017$).
[c]P-values for pairwise comparison between pre-frail and frail groups (significant at $P \leq 0.017$).
[d]P-values obtained through Kruskal–Wallis analysis of variance test with pairwise comparisons, except for gender (χ^2 test).

Table 2 Discriminative power of the SarQoL® questionnaire in frailty

	Robust (n = 172)	Pre-frail (n = 167)	Frail (n = 43)	P R-PF[a]	P R-F[b]	P PF-F[c]	P Overall[d]
Domain 1: physical and mental health	72.2 (59.5–86.63)	59.97 (52.2–68.87)	45.53 (38.87–55.53)	<0.001	<0.001	<0.001	<0.001
Domain 2: locomotion	77.78 (58.33–91.67)	55.56 (47.22–69.44)	41.67 (30.56–52.78)	<0.001	<0.001	0.001	<0.001
Domain 3: body composition	70.83 (55.21–80)	58.33 (50–66.67)	50 (41.67–58.33)	<0.001	<0.001	0.014	<0.001
Domain 4: functionality	82.69 (73.11–91.07)	69.23 (59.62–78.85)	55.36 (48.08–62.50)	<0.001	<0.001	<0.001	<0.001
Domain 5: activities of daily living	76.67 (64.47–85.32)	60.71 (48.33–72.50)	50 (33.33–58.33)	<0.001	<0.001	<0.001	<0.001
Domain 6: leisure activities	66.50 (49.88–66.50)	33.25 (33.25–49.88)	33.25 (16.62–33.25)	<0.001	<0.001	0.011	<0.001
Domain 7: fears	100.00 (87.50–100)	87.50 (87.50–100.00)	87.5 (75.00–87.50)	0.032	0.004	0.119	<0.001
Overall QoL score	77.10 (64.35–85.90)	62.54 (53.33–69.57)	49.99 (40.45–56.06)	<0.001	<0.001	<0.001	<0.001

BMI, body mass index; QoL, quality of life.
All results reported as median (25th–75th percentile).
[a]P-values for comparison between robust and pre-frail groups. P-values obtained through multinomial regression analysis adjusted on age, BMI, no. of drugs, no. of co-morbidities, and gender.
[b]P-values for comparison between robust and frail groups. P-values obtained through multinomial regression analysis adjusted on age, BMI, no. of drugs, no. of co-morbidities, and gender.
[c]P-values for comparison between pre-frail and frail groups. P-values obtained through multinomial regression analysis adjusted on age, BMI, no. of drugs, no. of co-morbidities, and gender.
[d]P-values obtained through Kruskal–Wallis analysis of variance test.

and anxiety/depression items of the EQ-5D. If this is correct, we expect to find weak or non-existent correlations between the SarQoL and these domains.

The convergent validity of the SarQoL questionnaire in the entire sample was excellent, as evidenced by the confirmation of the four pre-specified hypotheses and the strong correlations between the overall QoL score of the SarQoL questionnaire and the four domains theorized to measure similar constructs (correlation coefficients between 0.447 and 0.798). When isolating the three frailty groups, the results were largely similar, with the exception of the correlation between the SarQoL and the SF-36 role limitation due to physical problems domain in the frail group, which dropped from $r = 0.628$ ($P < 0.001$) to $r = 0.246$ ($P = 0.199$).

The results of the divergent construct validity were less straightforward: both in the complete sample and in the three frailty categories, we found moderate to strong correlations between the SarQoL questionnaire and the two domains of the SF-36 theorized to be measuring a different construct. The hypotheses with the EQ-5D self-care and anxiety/depression items were confirmed by weak correlations (respectively, $r = -0.273$; $P < 0.001$ and $r = -0.257$; $P < 0.001$).

The full results, shown in *Table* 3, demonstrate that six out of the eight pre-specified hypotheses were confirmed, fulfilling the criteria of 75%, which indicates acceptable construct validity.

Reliability

One of the 30 participants was not evaluated for physical frailty, which means that test–retest data were available for 29 participants, of which 4 (13.8%) were robust, 18 (62.1%) were pre-frail, and 7 (24.1%) were frail. An ICC of 0.918 [95% confidence interval (CI) = 0.834–0.961] was found for the overall QoL score of the SarQoL questionnaire, demonstrating excellent test–retest reliability. The ICCs for the individual domains showed acceptable (ICC > 0.7) reliability for all but two domains: Domain 6, leisure activities [ICC = 0.391 (95% CI = 0.029–0.660)], and Domain 7, fears [ICC = 0.318 (95% CI = −0.055 to 0.612)]. Detailed results for the test–retest reliability are reported in *Table* 4.

The SEM in this sample was calculated to be 3.88 points, leading to a smallest detectable change of 10.76 points. In practical terms, the overall QoL score of an individual participant would have to change by 10.76 points to be able to be sure that the observed change in QoL is due to a real change in QoL in the patient. SEM and smallest detectable change for the individual domains are reported in *Table* 4.

Floor and ceiling effects

None of the 382 participants obtained the lowest (0) or the highest (100) score possible for the overall QoL score of the SarQoL® questionnaire, showing the absence of floor and ceiling effects in the summary score.

Responsiveness

Out of the 382 participants who provided usable data at the first year of the SarcoPhAge study, 235 remained in the study at the fifth year of follow-up and were included in the responsiveness evaluation. Of these 235, a further 117 changed in terms of their frailty status between the first and fifth years of the study, and these were included in the analysis of responsiveness through the evaluation of effect size (SRMs).

We examined nine hypotheses used in an earlier study of the responsiveness of the SarQoL questionnaire, on the theorized correlation between changes measured by the SarQoL questionnaire and by other questionnaires. We were able to confirm five out of nine hypotheses but had to reject Hypothesis 1 (ΔSarQoL overall score and ΔSF-36 general health), Hypothesis 4 (ΔSarQoL overall score and ΔEQ-VAS), Hypothesis

Table 3 Construct validity of the SarQoL® questionnaire in frailty

	Robust ($n = 172$)		Pre-frail ($n = 167$)		Frail ($n = 43$)		All ($n = 382$)	
	r	*P*	*r*	*P*	*r*	*P*	*r*	*P*
Convergent validity								
SF-36 physical functioning	0.761	<0.001	0.693	<0.001	0.608	<0.001	0.798	<0.001
SF-36 role limitation physical	0.408	<0.001	0.611	<0.001	0.246	0.199	0.628	<0.001
SF-36 vitality	0.595	<0.001	0.499	<0.001	0.695	<0.001	0.678	<0.001
EQ-5D utility score	0.305	<0.001	0.311	<0.001	0.564	0.001	0.447	<0.001
Divergent validity								
SF-36 role limitation emotional	0.400	<0.001	0.473	<0.001	0.015	0.936	0.503	<0.001
SF-36 mental health	0.588	<0.001	0.489	<0.001	0.392	0.035	0.554	<0.001
EQ-5D self-care	[a]	[a]	−0.207	0.011	−0.278	0.145	−0.273	<0.001
EQ-5D anxiety/depression	−0.161	0.070	−0.222	0.010	−0.489	0.007	−0.257	<0.001

[a]All 172 robust subjects responded identically on the EQ-5D self-care question.

Table 4 Test–retest reliability of the SarQoL® questionnaire in frailty

	ICC	95% CI	SEM	SDC
Domain 1: physical and mental health	0.764	0.558–0.881	7.06	19.57
Domain 2: locomotion	0.850	0.706–0.926	7.92	21.94
Domain 3: body composition	0.700	0.454–0.847	8.81	24.41
Domain 4: functionality	0.879	0.759–0.941	5.50	15.25
Domain 5: activities of daily living	0.812	0.638–0.907	6.78	18.80
Domain 6: leisure activities	0.391	0.029–0.660	13.82	38.30
Domain 7: fears	0.318	−0.055 to 0.612	14.32	39.68
Overall QoL score	0.918	0.834–0.961	3.88	10.76

CI, confidence interval; ICC, intraclass correlation coefficient; QoL, quality of life; SDC, smallest detectable change; SEM, standard error of measurement.

Table 5 Evaluation of responsiveness with hypotheses

Hypothesis	Expected strength of correlation	Observed correlation		Confirmation/ rejection
		r	*P*-value	
1. ΔSarQoL® overall score and ΔSF-36 general health domain are correlated.	$r > 0.4$	0.389[a]	<0.001	Rejected
2. ΔSarQoL® overall score and ΔSF-36 vitality domain are correlated.	$r > 0.3$	0.460[b]	<0.001	Confirmed
3. ΔSarQoL® overall score and ΔSF-36 physical functioning domain are correlated.	$r > 0.5$	0.690[a]	<0.001	Confirmed
4. ΔSarQoL® overall score and ΔEQ-VAS are correlated.	$r > 0.4$	0.226[a]	<0.027	Rejected
5. ΔSarQoL® Domain 1 (physical and mental health) and ΔSF-36 general health domain are correlated.	$r > 0.3$	0.139[a]	0.176	Rejected
6. ΔSarQoL® Domain 1 (physical and mental health) and ΔEQ-VAS are correlated.	$r > 0.3$	0.142[a]	0.166	Rejected
7. ΔSarQoL® Domain 2 (locomotion) and ΔSF-36 physical functioning domain are correlated.	$r > 0.4$	0.539[a]	<0.001	Confirmed
8. ΔSarQoL® Domain 4 (functionality) and ΔSF-36 physical functioning domain are correlated.	$r > 0.5$	0.601[a]	<0.001	Confirmed
9. ΔSarQoL® Domain 5 (activities of daily living) and ΔSF-36 physical functioning domain are correlated.	$r > 0.5$	0.617[a]	<0.001	Confirmed

Δ = change over 4 years.
[a]Spearman correlation.
[b]Pearson correlation.

5 (ΔSarQoL Domain 1 and ΔSF-36 general health), and Hypothesis 6 (ΔSarQoL Domain 1 and ΔEQ-VAS). According to the criteria formulated by De Boer *et al.*, this indicated that the SarQoL questionnaire possesses moderate responsiveness because 45% of the hypotheses have been refuted. The details of the hypotheses and the observed correlations can be found in *Table* 5.

We also evaluated responsiveness with the metric of effect size. We calculated SRMs for all domains and summary scores of the SarQoL, SF-36, and EQ-5D questionnaires. The complete results are reported in *Table* 6. We can observe that the SRM of the SarQoL overall score (corrected SRM = −1.14) is much larger than the SF-36 PCS (corrected SRM = −0.634), the EQ-5D index (corrected SRM = 0.064), and the EQ-VAS (corrected SRM = −0.267). Globally, the SarQoL questionnaire had small effect sizes for three domain scores, moderate for 2 and large for 1. The SF-36 obtained small effect sizes for five domains and the MCS, and moderate effect sizes for three domains and the PCS.

Discussion

This study examined whether the SarQoL questionnaire could be used as a disease-specific instrument to measure health-related QoL in frailty. The psychometric results presented in this article indicate that it has adequate measurement properties when used with the Fried frailty criteria. This means that the SarQoL could be a new option for researchers seeking to evaluate QoL in populations characterized by the presence of pre-frailty and/or frailty.

This study demonstrated that the SarQoL questionnaire can discriminate between robust, pre-frail, and frail subjects, with declining QoL scores according to the category of frailty, and that it can do so over a wide range of concepts. The systematic review by Crocker *et al.* highlighted that (sub)scales measuring physical aspects of QoL were broadly able to discriminate between robust and frail people but reported inconsistent results for other aspects of QoL.[7] Therefore, it is encouraging to see that the SarQoL questionnaire

Table 6 Standardized response means

Domains	Corrected SRM	Interpretation[a]
1. ΔSarQoL® D1 physical and mental health	−0.383	Small
2. ΔSarQoL® D2 locomotion	−0.755	Moderate
3. ΔSarQoL® D3 body composition	−0.315	Small
4. ΔSarQoL® D4 functionality	−0.940	Large
5. ΔSarQoL® D5 activities of daily living	−0.883	Large
6. ΔSarQoL® D6 leisure activities	−0.255	Small
7. ΔSarQoL® D7 fears	−0.070	Trivial
8. ΔSarQoL® overall score	−1.144	Large
9. ΔSF-36 physical functioning	−0.749	Moderate
10. ΔSF-36 social functioning	−0.204	Small
11. ΔSF-36 role limitations due to physical health	−0.301	Small
12. ΔSF-36 role limitations due to emotional problems	−0.251	Small
13. ΔSF-36 mental health	−0.274	Small
14. ΔSF-36 vitality	−0.577	Moderate
15. ΔSF-36 bodily pain	−0.490	Small
16. ΔSF-36 general health	−0.693	Moderate
17. ΔSF-36 physical component summary	−0.634	Moderate
18. ΔSF-36 mental component summary	−0.224	Small
19. ΔEQ-5D utility index	0.064	Trivial
20. ΔEQ-VAS	−0.267	Small

SRM, standardized response mean.
SRMs are calculated by dividing the mean difference between scores from the first year and the first year by the standard deviation of the differences between these paired values. The SRM values were subsequently corrected with the formula SRM $\sqrt{2}\sqrt{(1-r)}$, where '*r*' signifies the correlation between Year 1 and Year 5 scores. Δ = change over 4 years.
[a]Interpretation of corrected SRMs: $0.20 \leq$ SRM < 0.49 = small change; $0.50 \leq$ SRM < 0.79 = moderate change; and SRM ≥ 0.80 = large change.

is able to discriminate on more than just the physical aspects of QoL and that it brings extra precision in being able to discriminate between robust, frail, and pre-frail individuals. A note of caution is warranted with regard to Domain 7, where only the comparison between robust and frail participants yielded significantly different QoL scores. This domain should not be interpreted in a vacuum but taking into account the other domain scores and the overall QoL score.

The internal consistency was shown to be high (α = 0.866), indicating that the domains in the questionnaire are highly interrelated and measure the same construct, QoL. Mixed results were obtained in the evaluation of the construct validity of the questionnaire. All four hypotheses on the convergent validity were confirmed, but two out of the four hypotheses for divergent validity were rejected. The two rejected hypotheses, where we found stronger correlations than expected, were between the overall QoL score of the SarQoL questionnaire and the mental health and role limitations due to emotional problems domains of the SF-36. It may be that our hypotheses are erroneous and that these two domains are conceptually closer to the SarQoL questionnaire than we theorized. One correlation of particular interest is between the SarQoL overall QoL score and the SF-36 role limitation due to physical limitations in the frail group (r = 0.246), because it is significantly lower than the correlation coefficients in the robust (r = 0.408) and pre-frail groups (r = 0.611). Upon further investigation, this discrepancy is linked to the significant floor effect in this SF-36 domain, where 16 of the 30 participants obtain the lowest score possible.

The test–retest reliability of the questionnaire was excellent, with an ICC of 0.918 (95% CI = 0.834–0.961) for the overall score. However, because the original study only contacted the participants diagnosed as sarcopenic with the European Working Group on Sarcopenia in Older People criteria to enter the evaluation of the test–retest reliability, there were only data available for 29 participants. So, while this is a result that indicates good test–retest reliability, with an elevated ICC and a relatively small CI, these results should be confirmed in a larger sample and in particular samples with sufficient pre-frail and frail participants to calculate ICC's for these particular groups. Because the SEM and the smallest detectable change are based on the test–retest data, this same remark also applies to these two indicators. It should also be noted that, in this study, Domain 6 (leisure activities) and Domain 7 (fears) did not demonstrate adequate reliability. We hypothesize that this may because of the low sample size in combination with the low number of items for these two domains (two items for Domain 6 and four items for Domain 7), which causes any difference between the responses between the first and second administration of the questionnaire to be exaggerated in the scores.

We examined the ability of the SarQoL questionnaire to detect a change in QoL. We found moderate responsiveness through the confirmation of five out of nine hypotheses on the correlation between changes in QoL observed by the SarQoL questionnaire and by other questionnaires. It is

possible that the rejection of several hypotheses is linked to the lower SRMs found between the different questionnaires. In fact, the SRM of the overall QoL score of the SarQoL questionnaire is markedly stronger at SRM = −1.144 compared with the strongest effect size of the SF-36, which was the physical functioning subscale at SRM = −0.749. It may be that the rejection of some hypotheses was thus caused not by poor responsiveness of the SarQoL questionnaire but by smaller effect sizes found by the SF-36. Similarly, for the EQ-VAS, we found a small effect at SRM = −0.267 and the rejection of two hypotheses associated with this instrument. Here also, this may be more linked to the performance of the EQ-VAS in combination with the 4 year interval between the assessments. It is highly likely that an instrument such as the EQ-VAS would be influenced by response shift, which is defined as a change in the self-evaluation of the meaning of a target construct caused by reconceptualization of the construct, a reprioritization of the participants' values, or a recalibration of the respondents' internal standards of measurement.[38,45] Overall assessments, such as the EQ-VAS, which asks the respondent to indicate on a scale from 0 to 100 'how good or bad your health state is today', are more vulnerable to response shift because they require careful consideration and interpretation of the question. The participants had to evaluate for themselves the meaning of the concept 'health state' and what is considered 'good' and 'bad' and assign a numerical value to this, leaving open the possibility of reconceptualization, reprioritization, or recalibration.[46] Researchers investigating changes in QoL over time or pre-intervention/post-intervention should make the overall QoL score of the SarQoL questionnaire their main outcome, given that it has the highest SRM and the smallest detectable change. If a significant change in overall QoL is found, further analyses of the individual domains could be useful in indicating on what domains a participant's QoL has changed.

Because this study used data collected during a previous study, we were unable to investigate and quantify the content validity of the SarQoL questionnaire in a population of frail, older, community-dwelling individuals. In the development of the questionnaire, content validity had been put at the heart of the process by soliciting, at each step of the item generation and selection process, input from multiple sarcopenic persons.[11] In this study, we were unable to provide this information from frail individuals. However, some authors have theorized that sarcopenia, the target condition for which the SarQoL questionnaire was developed, constitutes one of the main components of the clinical frailty syndrome, all the while recognizing that frailty should not be limited to physical manifestations but should also incorporate psychological, cognitive, emotional, social, and spiritual factors.[24,47] Currently, to our knowledge, the only questionnaire that measures QoL and that is specifically designed with and for older frail persons is the Geriatric Quality of Life Questionnaire.[48] However, the developers left the definition of what constitutes the 'frail elderly' up to the appreciation of the clinicians responsible for recruitment, instead of a recognized diagnostic tool. While the SarQoL questionnaire was not specifically developed for frailty, the shared characteristics between sarcopenia and frailty mean that it should be able to provide a precise measurement of the physical weakness aspect of frailty. Apart from the physical domains, the SarQoL has also incorporated items on mental health, body image, sexuality, activities of daily living, leisure activities, and fears, making for a multidimensional framework of QoL.

Healthy ageing is already high on the agenda for most health systems in both Western and Asian countries and will only gain in importance as the number of older people increases.[49] Concepts such as frailty, sarcopenia, or the construct recently proposed by the World Health Organization called Intrinsic Capacity, which is a composite of all the physical and mental capacities of an individual, may play an important role in any future medical approach.[50] Whatever approach is adopted, it must take in the perspective and priorities of the target population, and QoL can be an important metric for this. Having valid, reliable, and precise instruments to measure QoL that can pick up on the impact of a specific target condition is a prerequisite to be able to rely on QoL instruments to provide information on the patients' lived experience.

There are some limitations to this study. First off, we adopted the frailty criteria developed by Fried *et al.*, but other diagnostic approaches are available, such as the Rockwood Clinical Frailty Scale or the IF-VIG, among others.[51–53] Although all these approaches purportedly measure the same concept, frailty, we cannot be sure that the results on the validity and reliability of the SarQoL would have been the same if we had applied other diagnostic approaches. Secondly, our sample of robust, pre-frail, and frail participants is not necessarily representative of frailty in the wider community. Because these data were collected within a study that recruited volunteers, and which asked those volunteers to make several trips to the research centre, it is likely that the SarcoPhAge study recruited a sample that was in better overall condition, and that had better mobility, than a representative sample of pre-frail and frail participants. While this study has shown that the SarQoL questionnaire is a valid and reliable tool in frailty, additional investigations in samples with a different make-up need to confirm these results. Lastly, while the overall sample size was more than adequate for a psychometric study, the test–retest sample is relatively small with only 29 participants. This steep reduction from the overall sample size is a result of the fact that only a subset of participants was invited to complete the questionnaire a second time. However, because we have the 95% CI around the ICC, we can judge that most values have adequate precision, apart from Domains 6 and 7.

In conclusion, the study evaluated the validity and reliability of the SarQoL questionnaire in frailty and found that it is a valid and reliable tool for the assessment of QoL. Because of the shared mechanism of physical weakness between sarcopenia and frailty, the SarQoL questionnaire can provide more specific information on QoL in frailty than the generic questionnaires available.

Ethical standards

This study performed a secondary analysis of previously collected data. Therefore, no specific approval of a medical ethics committee was sought. The SarcoPhAge study was approved by the Medical Ethics Committee of the University Hospital of Liège, and all participants provided written informed consent, in compliance with the 1964 Declaration of Helsinki and its subsequent amendments.

Author contributions

All authors participated in the conception, design, and redaction of the manuscript, as well as read and approved the final version of the manuscript. The authors of this manuscript certify that they comply with the ethical guidelines for authorship and publishing in the *Journal of Cachexia, Sarcopenia and Muscle*.[54]

Funding

A.G. is supported by a FRIA grant from the Fonds De La Recherche Scientifique - FNRS.

Online supplementary material

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Frailty criteria and diagnosis

Conflict of interest

C.B., O.B., and J.-Y.R. are shareholders of SarQoL sprl. A.G. and M.L. declare that they have no conflicts of interest.

References

1. World Health Organisation. Decade of healthy ageing. 2020 https://www.who.int/docs/default-source/decade-of-healthy-ageing/final-decade-proposal/decade-proposal-final-apr2020-en.pdf?sfvrsn=b4b75ebc_3
2. WHO Clinical Consortium on Healthy Ageing. Report of consortium meeting 1–2 December 2016 in Geneva, Switzerland. Geneva; 2017.
3. Vermeiren S, Vella-Azzopardi R, Beckwée D, Habbig AK, Scafoglieri A, Jansen B, et al. Frailty and the prediction of negative health outcomes: a meta-analysis. *J Am Med Dir Assoc* 2016;**17**:1163.e1–1163.e17.
4. Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of frailty in community-dwelling older persons: a systematic review. *J Am Geriatr Soc* 2012;**60**:1487–1492.
5. O'Caoimh R, Galluzzo L, Rodríguez-Loso A, Van der Heyden J, Ranhoff A, Lemprini-Koula M, et al. Prevalence of frailty at population level in European ADVANTAGE Joint Action Member States: a systematic review and meta-analysis. *Ann Ist Super Sanita* 2018;**54**:226–238.
6. Buckinx F, Rolland Y, Reginster J-Y, Ricour C, Petermans J, Bruyère O. Burden of frailty in the elderly population: perspectives for a public health challenge. *Arch Public Health* 2015;**73**:19.
7. Crocker TF, Brown L, Clegg A, Farley K, Franklin M, Simpkins S, et al. Quality of life is substantially worse for community-dwelling older people living with frailty: systematic review and meta-analysis. *Qual Life Res* 2019;**28**:2041–2056.
8. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000;**17**:13–35.
9. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol* 2003;**56**:52–60.
10. Beaudart C, Biver E, Bruyère O, Cooper C, Al-Daghri N, Reginster JY, et al. Quality of life assessment in musculo-skeletal health. *Aging Clin Exp Res* 2018;**30**(5):413–418.
11. Beaudart C, Biver E, Reginster JY, Rizzoli R, Rolland Y, Bautmans I, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. *Age Ageing* 2015;**44**:960–966.
12. Beaudart C, Edwards M, Moss C, Reginster JY, Moon R, Parsons C, et al. English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. *Age Ageing* 2017;**46**:271–277.
13. Beaudart C, Biver E, Reginster JY, Rizzoli R, Rolland Y, Bautmans I, et al. Validation of the SarQoL®, a specific health-related quality of life questionnaire for sarcopenia. *J Cachexia Sarcopenia Muscle* 2017;**8**:238–244.
14. Geerinck A, Scheppers A, Beaudart C, Bruyère O, Vandenbussche W, Bautmans R, et al. Translation and validation of the Dutch SarQoL®, a quality of life questionnaire specific to sarcopenia. *J Musculoskelet Neuronal Interact* 2018;**18**:463–472.
15. Konstantynowicz J, Abramowicz P, Glinkowski W, Taranta E, Marcinowicz L, Dymitrowicz M, et al. Polish validation of the SarQoL®, a quality of life questionnaire specific to sarcopenia. *J Clin Med* 2018;**7**:323.
16. Gasparik AI, Mihai G, Beaudart C, Bruyere O, Pop R-M, Reginster J-Y, et al. Correction to: psychometric performance of the Romanian version of the SarQoL®, a health-related quality of life questionnaire for sarcopenia. *Arch Osteoporos* 2018;**13**:98.
17. Fábrega-Cuadros R, Martínez-Amat A, Cruz-Díaz D, Aibar-Almazán A, Hita-Contreras F. Psychometric properties of the Spanish version of the sarcopenia and quality of life, a quality of life questionnaire specific

for sarcopenia. *Calcif Tissue Int* 2020;**106**:274–282.
18. Alekna V, Kilaite J, Tamulaitiene M, Geerinck A, Mastaviciute A, Bruyère O, et al. Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). *Eur Geriatr Med* 2019;**10**:761–767.
19. Tsekoura M, Billis E, Gliatis J, Tsepis E, Matzaroglou C, Sakkas GK, et al. Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. *Disabil Rehabil* 2020;**42**(7):1006–1012.
20. Geerinck A, Alekna V, Beaudart C, Bautmans I, Cooper C, De Souza Orlandi F, et al. Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: an analysis of subjects from 9 validation studies. *PLoS One* 2019;**14**:e0216065.
21. Geerinck A, Bruyère O, Locquet M, Reginster JY, Beaudart C. Evaluation of the responsiveness of the SarQoL® questionnaire, a patient-reported outcome measure specific to sarcopenia. *Adv Ther* 2018;**35**:1842–1858.
22. Cederholm T. Overlaps between frailty and sarcopenia definitions. *Nestle Nutr Inst Workshop Ser* 2015;**83**:65–69.
23. Sieber CC. Frailty—from concept to clinical practice. *Exp Gerontol* 2017;**87**:160–167.
24. Cruz-Jentoft AJ, Michel JP. Sarcopenia: a useful paradigm for physical frailty. *Eur Geriatr Med* 2013;**4**:102–105.
25. Beaudart C, Reginster JY, Petermans J, Gillain S, Quabron A, Locquet M, et al. Quality of life and physical components linked to sarcopenia: the SarcoPhAge study. *Exp Gerontol* 2015;**69**:103–110.
26. Locquet M, Beaudart C, Hajaoui M, Petermans J, Reginster JY, Bruyère O. Three-year adverse health consequences of sarcopenia in community-dwelling older adults according to 5 diagnosis definitions. *J Am Med Dir Assoc* 2019;**20**:43–46, e2.
27. Beaudart C, Locquet M, Touvier M, Reginster J-Y, Bruyère O. Association between dietary nutrient intake and sarcopenia in the SarcoPhAge study. *Aging Clin Exp Res* 2019;**31**:815–824.
28. Beaudart C, Sanchez-Rodriguez D, Locquet M, Reginster JY, Lengelé L, Bruyère O. Malnutrition as a strong predictor of the onset of sarcopenia. *Nutrients* 2019;**11**.
29. Locquet M, Beaudart C, Reginster JY, Bruyère O. Association between the decline in muscle health and the decline in bone health in older individuals from the SarcoPhAge cohort. *Calcif Tissue Int* 2019;**104**:273–284.
30. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001;**56**: M146–M156.
31. Taylor HL, Jacobs DR, Schucker B, Knudsen J, Leon AS, Debacker G. A questionnaire for the assessment of leisure time physical activities. *J Chronic Dis* 1978;**31**:741–755.
32. Roadolff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Measur* 1977;**1**:385–401.
33. Safonova YA, Lesnyak OM, Baranova IA, Suleimanova AK, Zotkin EG. Russian translation and validation of SarQoL®—quality of life questionnaire for patients with sarcopenia. *Nauchno-Prakticheskaya Revmatol* 2019;**57**:38–45.
34. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–483.
35. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;**33**:337–343.
36. The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy (New York)* 1990;**16**:199–208.
37. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;**63**:737–745.
38. de Vet HCW, Terwee CB, Mokkink LB. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011.
39. Davidson M. Known-Groups Validity. In *Encyclopedia of Quality of Life and Well-being Research*. Dordrecht: Springer Netherlands; 2014. p 3481–3482.
40. de Boer MR, Moll AC, de Vet HCW, Terwee CB, Völker-Dieben HJM, van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt* 2004;**24**:257–273.
41. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;**53**:459–468.
42. Terwee C, Bot S, de Boer M, van der Windt D, Knol D, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;**60**:34–42.
43. Middel B, Van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *Int J Integr Care* 2002;**2**.
44. Cohen J. Quantitative methods in psychology. *Psychol Bull* 1992;**112**:155–159.
45. Sprangers M, Schwartz C. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med* 1999;**48**:1507–1515.
46. Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health Qual Life Outcomes* 2004;**2**:1–11.
47. Rodríguez-Mañas L, Féart C, Mann G, Viña J, Chatterji S, Chodzko-Zajko W, et al. Searching for an operational definition of frailty: a Delphi method based consensus statement. The Frailty Operative Definition-Consensus Conference Project. *J Gerontol - Ser A Biol Sci Med Sci* 2013;**68**:62–67.
48. Guyatt GH, Eagle J, Sacket B, Willan A, Griffith L, McIlroy W, et al. Measuring quality of life in the frail. *J Clin Epidemiol* 1993;**46**:1433–1444.
49. Xia W, Cooper C, Li M, Xu L, Rizzoli R, Zhu M, et al. East meets West: current practices and policies in the management of musculoskeletal aging. *Aging Clin Exp Res* 2019;**31**:1351–1373.
50. Cesari M, De Carvalho IA, Thiyagarajan JA, Cooper C, Martin FC, Reginster JY, et al. Evidence for the domains supporting the construct of intrinsic capacity. *J Gerontol - Ser A Biol Sci Med Sci* 2018;**73**:1653–1660.
51. Dent E, Kowal P, Hoogendijk EO. Frailty measurement in research and clinical practice: a review. *Eur J Intern Med* 2016;**31**:3–10.
52. Woolford SJ, Sohan O, Dennison EM, Cooper C, Patel HP. Approaches to the diagnosis and prevention of frailty. *Aging Clin Exp Res* 2020;**32**:1629–1637.
53. Moreno-Ariño M, Torrente Jiménez I, Cartanyà Gutiérrez A, Oliva Morera JC, Comet R. Assessing the strengths and weaknesses of the Clinical Frailty Scale through correlation with a frailty index. *Aging Clin Exp Res* 2020;**32**:2225–2232.
54. von Haehling S, Morley JE, Coats AJS, Anker SD. Ethical guidelines for publishing in the Journal of Cachexia, Sarcopenia and Muscle: update 2019. *J Cachexia Sarcopenia Muscle* 2019;**10**:1143–1145.

Screening for sarcopenia with the SarQoL questionnaire

Aging Clinical and Experimental Research (2021) 33:2149–2155
https://doi.org/10.1007/s40520-021-01913-z

ORIGINAL ARTICLE

Assessment of the performance of the SarQoL® questionnaire in screening for sarcopenia in older people

Anton Geerinck[1] · Bess Dawson-Hughes[2] · Charlotte Beaudart[1] · Médéa Locquet[1] · Jean-Yves Reginster[1,3] · Olivier Bruyère[1,4,5]

Received: 23 April 2021 / Accepted: 10 June 2021 / Published online: 1 July 2021

Abstract

Background Because of its low prevalence and the need for physical tests to establish a diagnosis, recruiting sarcopenic people for clinical studies can be a resource-intensive process.

Aims We investigated whether the SarQoL®, a 55-item questionnaire designed to measure quality of life in sarcopenia, could be used to identify older people with a high likelihood of being sarcopenic, and to compare its performance to the SARC-F tool.

Methods We performed a secondary analysis of data from older, community-dwelling participants of the SarcoPhAge study, evaluated for sarcopenia according to the EWGSOP2 criteria, and who completed the SarQoL® and SARC-F questionnaires. We determined the optimal threshold to distinguish between sarcopenic and non-sarcopenic people with the Youden index. Screening performance was evaluated with the area under the curve (AUC) and by calculating sensitivity and specificity.

Results The analysis of 309 participants provided an optimal threshold value of ≤ 52.4 points for identifying people with sarcopenia with the SarQoL® questionnaire, which resulted in a sensitivity of 64.7% (41.1–84.2%), a specificity of 80.5% (75.7–84.7%) and an AUC of 0.771 (0.652–0.889). Compared to the SARC-F, the SarQoL® has greater sensitivity (64.7% vs 52.39%), but slightly lower specificity (80.5% vs. 86.6%).

Discussion The SarQoL® questionnaire showed acceptable screening accuracy, on par with the SARC-F. The optimal threshold of ≤ 52.4 points should be confirmed in other cohorts of older people.

Conclusions This exploratory study showed that the SarQoL® could potentially be applied in a screening strategy, with the added benefit of providing a measure of QoL at the same time.

Keywords Sarcopenia · Screening · SarQoL · Sensitivity · Specificity

Introduction

Sarcopenia has been described by the 2nd European Working Group on Sarcopenia in Older People (EWGSOP2) as a "progressive and generalised skeletal muscle disorder that is associated with increased likelihood of adverse outcomes including falls, fractures, physical disability and mortality". In the same article, the EWGSOP2 also presented a revision of its diagnostic criteria for sarcopenia, presenting a new diagnostic algorithm and changing the threshold values for low muscle strength and low muscle mass [1]. This revision has increased the consistency between studies in the evaluation of sarcopenia, but some studies have observed that it lowers the prevalence of sarcopenia compared to the EWGSOP1 criteria [2, 3]. For clinical research and epidemiological studies, this means that more candidates need to

✉ Anton Geerinck
Anton.geerinck@uliege.be

1 Division of Public Health, Epidemiology and Health Economics, World Health Organization Collaborating Centre for Public Health Aspects of Musculoskeletal Health and Ageing, University of Liège, Liège, Belgium

2 Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA

3 Chair for Biomarkers of Chronic Diseases, Biochemistry Department, College of Science, King Saud University, Riyadh, Saudi Arabia

4 Department of Sport Rehabilitation Sciences, University of Liège, Liège, Belgium

5 Physical, Rehabilitation Medicine and Sports Traumatology, SportS2, University Hospital of Liège, Liège, Belgium

be evaluated to achieve a sufficient number of sarcopenic participants to obtain the desired statistical power.

To help researchers recruit sarcopenic individuals in an efficient and cost-effective manner, multiple screening tools have been developed to identify those candidates with the highest probability of having sarcopenia. These come in different forms: there are questionnaires such as the Mini Sarcopenia Risk Assessment (MSRA—both a 7 and 5-item version available) and the SARC-F questionnaire (a 5 and 3-item version exist, as well as a version with calf circumference and a version which takes into account age and body mass) [4]. Other screening instruments rely solely on physical characteristics, such as the score developed by Ishii et al (age, grip strength and calf circumference), muscle mass prediction formulas or the chair stand test [4, 5].

Clinical studies in sarcopenia require a substantial amount of time and effort, because of the need to include and evaluate a large number of candidates to find sufficient sarcopenic subjects to achieve the required level of statistical power. A full diagnostic evaluation where muscle mass is evaluated by dual-energy X-ray absorptiometry (DXA) and muscle strength by dynamometer, as recommended, necessitates the use of qualified personnel and expensive instruments. Given the cost per patient for these evaluations, screening instruments that can significantly increase the proportion of sarcopenic persons within the pool of candidates invited for a full body composition assessment could greatly help the financial feasibility of large-scale clinical studies in sarcopenia. With this in mind, the hypothesis was raised that an existing instrument, developed to measure quality of life in sarcopenia, could potentially be of use in screening candidates for referral to full body composition evaluation and/or physical function assessment.

The instrument investigated in this study is the Sarcopenia Quality of Life (SarQoL®) questionnaire. It evaluates quality of life in sarcopenia through 55 items categorized into 7 domains of health-related dysfunction [6]. It is an auto-administered instrument and takes about 15 min to complete. Its clinimetric properties as a QoL questionnaire have been demonstrated in multiple validation studies conducted in multiple languages [7–18]. Of particular interest in this context is the repeated observation that the SarQoL® questionnaire is able to discriminate between sarcopenic and non-sarcopenic groups, with the former scoring significantly lower on the overall QoL score of the questionnaire compared to the latter. Its focus on the impact of musculoskeletal health on quality of life contributes to our expectation that the overall QoL score produced by the SarQoL® questionnaire could be used to screen older people and identify those with a higher likelihood of sarcopenia.

The objective of this study is therefore to evaluate the capacity of the Overall QoL score of the SarQoL® questionnaire to detect individuals with sarcopenia according to the revised EWGSOP2 consensus criteria. The hypothesis linked to this objective is that the Overall QoL score of the SarQoL® questionnaire has an area under the ROC curve (AUC) greater than 0.7, indicating the test is useful in distinguishing between sarcopenic and non-sarcopenic people [19].

The secondary objective of this study is to compare the screening performance of the Overall QoL score of the SarQoL® questionnaire with the performance of the 5-item SARC-F questionnaire, the screening tool recommended by the EWGSOP2 [1]. The hypothesis linked to this objective is that the Overall QoL score is at least as accurate as the SARC-F, judged by AUC, sensitivity and specificity.

Material and methods

This study is a cross-sectional secondary evaluation of data collected at the third year of follow-up of the Sarcopenia and Physical Impairment with advancing Age (SarcoPhAge) prospective cohort study, carried out in the Liège province of Belgium [20]. The SarcoPhAge study was conducted in compliance with the principles outlined in the Declaration of Helsinki. The study protocol and its amendments received approval from the Ethics Committee of the University Teaching Hospital of Liège (n° 2012-277), and all participants provided written informed consent. This article was written to comply, as much as feasible, with the most recent version of the Standards for Reporting Diagnostic Accuracy (STARD) checklist [21].

Participants

The SarcoPhAge study enrolled a convenience sample of people who visited an outpatient clinic in Liège (Belgium) as well as people who responded to a press advertisement between June 2013 and July 2014. Participants in this study were 65 years of age or older, and, because of the limitations of the dual-energy X-ray absorptiometry (DXA) instrument, people with a BMI above 50 kg/m^2 or with amputated limbs were not eligible. There were no additional inclusion criteria beyond these [20]. The third year of follow-up (July 2015–2016) was selected for inclusion because this was the first year that both the SarQoL® questionnaire and the SARC-F questionnaire were administered to all participants.

Measurements

For each participant, muscle mass was measured with a dual-energy X-ray absorptiometry instrument (Hologic Discovery A, USA) and grip strength with the Saehan hydraulic hand dynamometer (Saehan Corp., Masan, South Korea). Both instruments were calibrated according to the respective

manufacturer's instructions at the recommended intervals. Appendicular skeletal muscle mass was calculated as the sum of all 4 limbs and divided by the squared height of the participant in question to obtain a skeletal muscle mass index ($SMI = ASM/Ht^2$). The grip strength of a person was defined as the highest value out of 6 measurements (3 for the dominant hand and 3 for the non-dominant hand). Detailed descriptions of both measurements are available in the article on the baseline results of the SarcoPhAge study [20]. These data allowed us to diagnose sarcopenia according to the EWGSOP2 criteria in participants with low muscle mass ($ASM/Ht^2 < 7.0$ kg/m^2 for men and < 5.5 kg/m^2 for women) and low muscle strength (grip strength < 27 kg for men and < 16 kg for women) [1]. Sarcopenia diagnosed with the EWGSOP2 criteria constitutes the reference standard in this study because of its status as the current consensus criteria and its applicability to samples recruited in Europe [1].

The index test in this study, the paper-based French-language SarQoL® questionnaire, was completed by the participants without assistance. An Overall QoL score (0–100 points) is calculated where lower scores indicate lower QoL, and thus also greater sarcopenia-related disability [6, 22]. The questionnaire is available in multiple languages from the website www.sarqol.org, and the Overall QoL score was calculated with an Access database developed for this purpose. Given the exploratory nature of this investigation, we did not pre-specify a test-positivity cut-off point.

We included a second index test in this analysis, so as to be able to compare the performance of the SarQoL® questionnaire against the current most widely used screening instrument in sarcopenia, the SARC-F [23]. It is composed of 5 questions on strength, locomotion, rising from a chair, climbing stairs and history of falls. A total score is calculated and ranges from 0 to 10 points, where higher scores are linked with a higher probability of being diagnosed with sarcopenia. A score of ≥ 4 points is used as a cut-off to identify individuals who require a full examination for sarcopenia in clinical practice [23]. The SARC-F was developed to be able to detect sarcopenia as diagnosed with the EWGSOP1 criteria, and a meta-analysis found a pooled sensitivity of 0.21 (0.13–0.31) combined with a specificity of 0.90 (0.83–0.94) [24]. With the publication of the revised EWGSOP2 criteria, several authors have looked again at the performance of the SARC-F, and a meta-analysis that pooled the results from 4 studies found an AUC of 0.75 (95% CI 0.71–0.78) with a sensitivity of 0.77 (95% CI 0.49–0.92) and a specificity of 0.63 (95% CI 0.43–0.79), while the same meta-analysis found a pooled sensitivity of 0.32 (95% CI 0.19–0.47) and specificity of 0.86 (95% CI 0.77–0.92) for the EWGSOP1 criteria in 13 studies [25].

To compare the performance of the SarQoL® questionnaire and the SARC-F instrument with a screening instrument based on physical indicators, we calculated the probability of sarcopenia according to the Ishii formula, which was the best-performing screening instrument in a comparison of 5 with the EWGSOP1 criteria [5]. For men, we used the formula $[0.62 \times (\text{age} - 64) - 3.09 \times (\text{grip strength} - 50) - 4.64 \times (\text{calf circumference} - 42)]$ to calculate the sum score and the formula $[1 / (1 + e^{-(\text{sum score} / 10 - 11.9)})]$ to calculate the probability of sarcopenia (expressed in percentage). For women, the formula $[0.80 \times (\text{age} - 64) - 5.09 \times (\text{grip strength} - 34) - 3.28 \times (\text{calf circumference} - 42)]$ provided the sum score and the formula $[1 / (1 + e^{-(\text{sum score} / 10 - 12.5)})]$ the probability of sarcopenia [26]. A sum score higher than 105 for men and 120 for women was used as the cut-off for a high probability of sarcopenia [26]. To the best of our knowledge, its performance when used to screen patients for sarcopenia with the EWGSOP2 criteria has not yet been established.

The reference test and the index tests were performed by the same investigator or completed by the participant during a single study visit. The study investigator also recorded clinical and demographic information needed for the Ishii formula. The results from the reference test and one of the index tests, the SARC-F, was directly available to the investigator. The SarQoL® Overall score and the Ishii score, were calculated some time after the end of the study visit.

Statistical analyses

Statistical analyses were carried out with the Statistical Package for the Social Sciences version 27.0.0.0 (SPSS Statistics; IBM, Armonk, NY). The distribution of variables in this analysis was examined by looking at the distance between median and mean, histogram, QQ-plot, and the Shapiro–Wilk test. Continuous variables are presented as mean ± standard deviation if normally distributed and as median (25th–75th percentile) if not normally distributed. The evaluation of the screening performance of the Overall QoL score of the SarQoL® questionnaire, the SARC-F tool and the Ishii screening test was based on their sensitivity (Se), specificity (Sp), positive likelihood ratio (LR +), negative likelihood ratio (LR-), positive predictive value (PPV), and negative predictive value (NPV) in relation to sarcopenia as diagnosed with the EWGSOP2 criteria. These values and the associated 95% confidence intervals were obtained through the GENLIN procedure, as outlined in document 422875 from IBM support [27]. Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC) provided the overall accuracy of the three screening instruments. An AUC value above 0.9 indicates high accuracy of the screening instrument, between 0.8 and 0.9 excellent accuracy and between 0.7 and 0.8 acceptable accuracy [19]. The Youden J statistic (sensitivity + specificity −1) was used to find the optimal cut-point for the Overall SarQoL score [28]. The analyses presented in this article

Table 1 Clinical characteristics

	Sarcopenic ($n=17$)	Not sarcopenic ($n=292$)	p-value*
Age (years)	80.07 (71.98–86.36)	73.55 (69.68–78.58)	0.011
Gender (women)	10 (58.8%)	170 (58.2%)	0.961
N° of drugs	9.00 (3.50–12.50)	6.00 (4.00–8.00)	0.035
N° of comorbidities	4.00 (3.00–7.00)	4.00 (2.00–5.00)	0.462
Gait speed (m/s)	0.70 ± 0.27	1.14 ± 0.28	< 0.001

*P-values from Mann–Whitney U-test, Pearson Chi-square or Student t-test, depending on variable characteristics

have been performed in all participants who were assessed for sarcopenia using the EWGSOP2 criteria, screened with the SARC-F questionnaire and who completed the SarQoL® questionnaire at the third follow-up of the SarcoPhAge study. A p-value of 0.05 was considered statistically significant.

Results

A total of 309 people were included in this analysis. All participants were assessed for sarcopenia with the EWGSOP2 criteria in the third yearly evaluation of the SarcoPhAge study, and 17 (5.5%) of them were diagnosed with sarcopenia. The sarcopenic participants were older than those not diagnosed as sarcopenic [80.07 (71.98 – 86.36) years versus 73.55 (69.68 –78.58) years, $p=0.011$]. They also took more medication and had a lower gait speed than those not diagnosed with sarcopenia. The complete clinical characteristics for the sample are detailed in Table 1.

The SARC-F questionnaire identified 48 participants (15.5% of the sample) with a score ≥ 4 points and thus suspected of having sarcopenia. A ROC curve of the SarQoL® Overall score and the SARC-F score is presented in Fig. 1. The AUC for the SarQoL® Overall score is 0.771 (95% CI: 0.652–0.889), and for the SARC-F 0.802 (95% CI: 0.696–0.909).

The Youden index was maximised at ≤ 52.4 points for the SarQoL® Overall score ($J_c=0.452$, Se = 0.647, Sp = 0.805). This threshold value, together with the prespecified threshold for the SARC-F, were used for the construction of Table 2, detailing the screening accuracy of the two instruments.

The SarQoL® Overall score, dichotomized at ≤ 52.4 points, had, in absolute numbers, slightly greater sensitivity than the SARC-F score (64.7% vs. 52.9%), because it correctly identified 11 out of the 17 sarcopenic participants, whereas the SARC-F correctly identified 9 out of 17. In terms of their specificity, the SARC-F had, in absolute numbers, slightly greater specificity than the SarQoL® Overall score (80.5% vs. 86.6%), with 253 non-sarcopenic subjects correctly identified compared to the 235 found by the SarQoL® questionnaire.

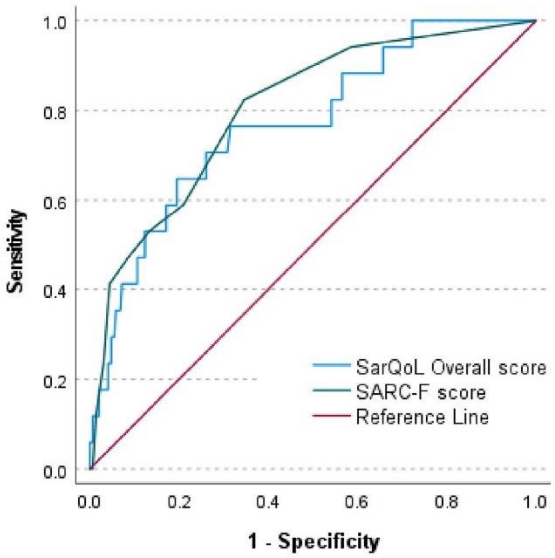


Fig. 1 ROC curves for the SarQoL® overall score and the SARC-F tool

Table 2 Screening accuracy of the SarQoL® overall score and the SARC-F instrument

	SarQoL	SARC-F
True positives	11	9
False positives	57	39
True negatives	235	253
False negatives	6	8
Sensitivity	0.647 (0.411–0.842)	0.529 (0.301–0.750)
Specificity	0.805 (0.757–0.847)	0.866 (0.824–0.902)
Positive predictive value	0.162 (0.088–0.261)	0.188 (0.095–0.313)
Negative predictive value	0.975 (0.950–0.990)	0.969 (0.944–0.986)
Positive likelihood ratio	3.315 (2.175–5.051)	3.964 (2.322–6.768)
Negative likelihood ratio	0.439 (0.230–0.837)	0.543 (0.327–0.901)
AUC	0.771 (0.652–0.889)	0.802 (0.696–0.909)

AUC: area under the ROC curve. Values between parentheses are the 95% confidence interval

The Ishii screening test outperformed both the SARC-F and the SarQoL® Overall score with an AUC of 0.884 (95% CI: 0.840–0.927). The Ishii screening test correctly identified all 17 sarcopenic individuals, and therefore had a sensitivity of 100%, and correctly identified 224 non-sarcopenic individuals for a specificity of 76.7%. It flagged a total of 85 people as being at high risk for sarcopenia, which is 27.5% of the total sample.

We also looked at the sensitivity and specificity of a range of threshold values for the SarQoL® Overall score, which are displayed in Table 3.

Discussion

This exploratory study showed that the SarQoL® questionnaire may be useful in screening potential candidates who are suspected of having sarcopenia for inclusion in clinical trials. The AUC of 0.771 (95% CI: 0.652–0.889) places it into the category of screening instruments with acceptable accuracy and confirms the primary study hypothesis. There might thus be a role for the SarQoL® questionnaire in a recruitment strategy of a clinical trial, certainly if it is already being considered to measure quality of life. We also found that the screening accuracy of the SarQoL® questionnaire in this sample was comparable to the SARC-F questionnaire but inferior to the Ishii screening test. The SarQoL® questionnaire was able to correctly identify more sarcopenic participants than the SARC-F (64.7% vs. 52.9%), but at the cost of a slightly lower specificity (80.5% vs 86.6%). The Ishii screening test, which relies on physical parameters, correctly identified all 17 sarcopenic participants, giving it a sensitivity of 100%, but had the lowest specificity of all three tests at 76.7%.

Table 3 Sensitivity and specificity for a range of threshold values for the SarQoL® overall score

Threshold value	Se	Sp	PPV	NPV
≤ 30 points	5.9%	100%	100%	94.8%
≤ 40 points	17.6%	95.9%	20.0%	95.2%
≤ 50 points	52.9%	85.6%	17.6%	96.9%
≤ 52.4 points (optimal threshold)	**64.7%**	**80.5%**	**16.2%**	**97.5%**
≤ 60 points	76.5%	65.8%	11.5%	98.0%
≤ 70 points	88.2%	40.1%	7.9%	98.3%
≤ 80 points	100%	21.2%	6.9%	100%
≤ 90 points	100%	7.9%	5.9%	100%
≤ 100 points	100%	NA	5.5%	NA

Se sensitivity, *Sp* specificity, *PPV* positive predictive value, *NPV* negative predictive value

Bold value indicates the optimal threshold identified by the Youden index and its associated Se, Sp, PPV and NPV

That the Ishii screening test outperforms the SARC-F and Overall SarQoL® score should not be a great surprise. In fact, the items in the Ishii test closely resemble those that make up the diagnosis of sarcopenia according to the EWGSOP2 criteria, namely grip strength and calf circumference (as an indicator of muscle mass) [26]. The Ishii screening test has also shown, in a Polish study, that it possesses good screening accuracy when used to find sarcopenic people diagnosed with the EWGSOP2 criteria [29]. However, any comparison between the Ishii screening test and the SARC-F and Overall SarQoL® score needs to take into account that the Ishii screening test necessitates a face-to-face contact between the researcher and the potential candidate to obtain grip strength and calf circumference measurements, whereas the SARC-F and the SarQoL® questionnaire can be administered via the postal service, through the internet or via telephone.

The screening efficacy of the SARC-F, one of the most widely used tools and recommend by several organizations, has been investigated for multiple diagnostic criteria and summarized in a meta-analysis published in 2021. The authors found that the screening accuracy of the SARC-F was characterized by relatively low sensitivity (27–39%) combined with relatively high specificity (86–91%) when used in conjunction with the EWGSOP, Asian Working Group on Sarcopenia, International Working Group on Sarcopenia, and the Foundation for the National Institutes of Health Sarcopenia Project criteria. Interestingly, when they calculated the pooled sensitivity and specificity of the SARC-F based on the EWGSOP2 criteria, they found inverse results: moderate sensitivity (77%) and lower specificity (63%), although these results were only based on 4 studies. It is also important to mention that 3 of the 4 included studies focused on hospitalized patients, and that the pooled prevalence of sarcopenia was higher than in the general population at 21.56% [25]. We are aware of two other studies that are not included in this meta-analysis, namely Piotrowicz et al who reported a sensitivity of 35.3% and a specificity of 85.7%, and Nguyen et al, with a sensitivity of 64.9% and a specificity of 68.2%, both of which recruited community-dwelling older people [29, 30]. It has been argued that the SARC-F is better suited to ruling out sarcopenia rather than case-finding, which seems to be the case for the last two articles mentioned, but not so for the 4 included in the meta-analysis of Lu et al [31, 32].

In our study, the SarQoL® questionnaire performed similarly to the SARC-F questionnaire, with slightly greater sensitivity but slightly lower specificity. The SarQoL® questionnaire was able to correctly identify more sarcopenic patients in the sample, but the PPV of 16.2% was lower than the PPV of 18.8% of the SARC-F instrument. This means that 68 people would have been singled out for further investigation by the SarQoL®

questionnaire, and 48 for the SARC-F, for two additional sarcopenic subjects to be found. Therefore, in our example, the SarQoL® questionnaire would have been preferable if the recruitment strategy called for finding the greatest number of sarcopenic participants in the shortest amount of time, accepting the extra cost in performing complete body composition and/or physical performance assessments on more people. The SarQoL® questionnaire also has the advantage that it is self-administered and, therefore, requires fewer hours of study personnel time than the SARC-F, which is interviewer-administered.

The specific purpose for which a screening instrument is used can influence which of its characteristics to prioritize. In an ideal situation, a screening instrument would be inexpensive, easy to administer, without side effects, reliable, valid, and both highly sensitive and specific. Oftentimes, however, a trade-off needs to be made between these characteristics. Both the SARC-F and the SarQoL® questionnaire are inexpensive, easy to administer and without side effects given that they are questionnaires. The SarQoL® questionnaire has also demonstrated to be reliable in multiple studies [11]. However, both the SARC-F and the SarQoL® questionnaire are not highly sensitive nor highly specific, and are not as sensitive as the Ishii screening test. Nonetheless, if its limitations are taken into account, the SarQoL® questionnaire could be useful within certain contexts.

There are some limitations to take into account when interpreting the results of this study. First off, this study was a secondary analysis of data collected previously, and not specifically designed to answer the research question. This has led to certain issues around the reduction of risk of bias, such as the fact that the research assistant was not blinded to the results of the body composition analysis, grip strength measurement and SARC-F score. A second issue is the fact that, because no pre-specified cut-off exists, we determined the optimal threshold for the Overall QoL score of the SarQoL® questionnaire with the Youden index. This reflects the best balance between sensitivity and specificity, but may not necessarily be generalizable. The various studies performed with the SarQoL® questionnaire have already shown that absolute quality of life scores can significantly differ between countries. Normative population data or pilot studies will be needed to inform the appropriate threshold value in different situations. Lastly, because of the design of this study, we did not perform sample size calculations but provided confidence intervals around the main outcome values to provide a measure of precision. For both the SarQoL® and the SARC-F questionnaire, relatively large confidence intervals are observed around their point estimates, owing to the small number of people diagnosed with sarcopenia according to the EWGSOP2 criteria in this sample.

This study shows the feasibility of using the SarQoL® questionnaire as a tool to select those people who may benefit from a complete sarcopenia evaluation. While this study presents an interesting new use for the SarQoL® questionnaire, caution should be used in applying the threshold value used in this study (≤ 52.4 points) to other populations.

Conclusion

In the population presented in this study, the SarQoL® Overall score, dichotomized at ≤ 52.4 points, performed roughly equal in terms of sensitivity and specificity to the SARC-F tool in identifying people considered sarcopenic with the EWGSOP2 criteria.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40520-021-01913-z.

Author's contributions AG, CB, OB and J-YR designed the study. ML and CB collected the data. AG performed the analysis and wrote the manuscript. All authors provided feedback on the manuscript and analyses and approved the final manuscript.

Funding AG is supported by a FRIA doctoral grant from the Fonds de la Recherche Scientifique (F.R.S-FNRS).

Availability of data and material Data are available as a supplementary file.

Code availability No specific code was written for this study.

Declarations

Conflict of interest CB, J-YR and OB are shareholders of SarQoL sprl. AG, BD-H and ML report no conflicts of interest related to this work.

Ethical approval The SarcoPhAge study was approved by the Ethics Committee of the University Teaching Hospital of Liège (n° 2012-277). Because this is a secondary analysis of previously collected data, no additional approval was sought for this specific analysis.

Statement of human and animal rights This study was approved by the ethical committee of the University Hospital of Liège and complied with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Written consent was obtained from all individual participants included in the study.

Consent to participate All participants provided written informed consent.

Consent for publication Not applicable.

References

1. Cruz-Jentoft AJ, Bahat G, Bauer J et al (2018) Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 48:16–31. https://doi.org/10.1093/ageing/afy169
2. Van Ancum JM, Alcazar J, Meskers CGM et al (2020) Impact of using the updated EWGSOP2 definition in diagnosing sarcopenia: a clinical perspective. Arch Gerontol Geriatr 90. https://doi.org/10.1016/j.archger.2020.104125
3. Locquet M, Beaudart C, Petermans J et al (2019) EWGSOP2 Versus EWGSOP1: impact on the Prevalence of Sarcopenia and Its Major Health Consequences. J Am Med Dir Assoc 20:384–385. https://doi.org/10.1016/j.jamda.2018.11.027
4. Mohd Nawi SN, Khow KS, Shiong Lim W et al (2019) Screening tools for sarcopenia in community-dwellers: a scoping review screening tools for Sarcopenia-Siti N Mohd Nawi et al. Ann Acad Med Singapore 48:201
5. Locquet M, Beaudart C, Reginster JY et al (2018) Comparison of the performance of five screening methods for sarcopenia. Clin Epidemiol 10:71–82. https://doi.org/10.2147/CLEP.S148638
6. Beaudart C, Biver E, Reginster JY et al (2015) Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: The SarQoL. Age Ageing 44:960–966
7. Beaudart C, Biver E, Reginster J-Y et al (2017) Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle. https://doi.org/10.1002/jcsm.12149
8. Beaudart C, Edwards M, Moss C et al (2017) English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. Age Ageing 46:271–277. https://doi.org/10.1093/ageing/afw192
9. Fábrega R, Antonio C, Amat M et al (2019) Psychometric properties of the spanish version of the sarcopenia and quality of life, a quality of life questionnaire specific for sarcopenia. Calcif Tissue Int. https://doi.org/10.1007/s00223-019-00635-9
10. Geerinck A, Bruyère O, Locquet M et al (2018) Evaluation of the responsiveness of the sarqol® questionnaire, a patient-reported outcome measure specific to sarcopenia. Adv Ther 35:1842–1858. https://doi.org/10.1007/s12325-018-0820-z
11. Geerinck A, Alekna V, Beaudart C et al (2019) Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (Sarqol) questionnaire: an analysis of subjects from 9 validation studies. PLoS ONE 14:e0216065. https://doi.org/10.1371/journal.pone.0216065
12. Erdogan T, Eris S, Avci S et al (2021) Sarcopenia quality-of-life questionnaire (SarQoL)®: translation, cross-cultural adaptation and validation in Turkish. Aging Clin Exp Res. https://doi.org/10.1007/s40520-020-01780-0
13. Geerinck A, Scheppers A, Beaudart C et al (2018) Translation and validation of the Dutch SarQoL ®, a quality of life questionnaire specific to sarcopenia. J Musculoskelet Neuronal Interact 18:463–472
14. Tsekoura M, Billis E, Gliatis J et al (2018) Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. Disabil Rehabil. https://doi.org/10.1080/09638288.2018.1514076
15. Konstantynowicz J, Abramowicz P, Glinkowski W et al (2018) Polish validation of the SarQoL®, a quality of life questionnaire specific to sarcopenia. J Clin Med 7:323. https://doi.org/10.3390/jcm7100323
16. Gasparik AI, Mihai G, Beaudart C et al (2018) Correction to: Psychometric performance of the Romanian version of the SarQoL(R), a health-related quality of life questionnaire for sarcopenia. Arch Osteoporos 13:98
17. Alekna V, Kilaite J, Tamulaitiene M et al (2019) Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). Eur Geriatr Med 10:761–767
18. Safonova YA, Lesnyak OM, Baranova IA et al (2019) Russian translation and validation of SarQoL® - quality of life questionnaire for patients with sarcopenia. Nauchno-Prakticheskaya Revmatol 57:38–45. https://doi.org/10.14412/1995-4484-2019-38-45
19. Mandrekar JN (2010) Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 5:1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d
20. Beaudart C, Reginster JY, Petermans J et al (2015) Quality of life and physical components linked to sarcopenia: the SarcoPhAge study. Exp Gerontol 69:103–110
21. Bossuyt PM, Reitsma JB, Bruns DE et al (2015) STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 351:1–9. https://doi.org/10.1136/bmj.h5527
22. Beaudart C, Biver E, Reginster JY et al (2017) Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 8:238–244. https://doi.org/10.1002/jcsm.12149
23. Malmstrom TK, Miller DK, Simonsick EM et al (2016) SARC-F: A symptom score to predict persons with sarcopenia at risk for poor functional outcomes. J Cachexia Sarcopenia Muscle 7:28–36. https://doi.org/10.1002/jcsm.12048
24. Ida S, Kaneko R, Murata K (2018) SARC-F for Screening of Sarcopenia Among Older Adults: a Meta-analysis of Screening Test Accuracy. J Am Med Dir Assoc 19:685–689. https://doi.org/10.1016/j.jamda.2018.04.001
25. Lu JL, Ding LY, Xu Q et al (2021) Screening accuracy of SARC-F for sarcopenia in the elderly: a diagnostic meta-analysis. J Nutr Heal Aging 25:172–182. https://doi.org/10.1007/s12603-020-1471-8
26. Ishii S, Tanaka T, Shibasaki K et al (2014) Development of a simple screening test for sarcopenia in older adults. Geriatr Gerontol Int 14:93–101. https://doi.org/10.1111/ggi.12197
27. IBM Support (2018) Can SPSS statistics produce epidemiological statistics from 2x2 tables such as positive and negative predictive values, sensitivity, specificity and likelihood ratios? Retrieved from https://www.ibm.com/support/pages/can-spss-statistics-produce-epidemiological-statistics-2x2-tables-such-positive-and-negative-predictive-values-sensitivity-specificity-and-likelihood-ratios
28. Youden WJ (1950) Index for rating diagostic tests. Cancer 3:32–35. https://doi.org/10.1002/1097-0142(1950)3:1%3c32::AID-CNCR2820030106%3e3.0.CO;2-3
29. Piotrowicz K, Głuszewska A, Czesak J et al (2021) SARC-F as a case-finding tool for sarcopenia according to the EWGSOP2. National validation and comparison with other diagnostic standards. Aging Clin Exp Res. https://doi.org/10.1007/s40520-020-01782-y
30. Nguyen TN (2020) Reliability and validity of SARC-F questionnaire to assess sarcopenia among vietnamese geriatric patients. Clin Interv Aging 15:879–886
31. Piotrowicz K, Gryglewska B, Gąsowski J (2021) The usefulness of SARC - F. Aging Clin Exp Res. https://doi.org/10.1007/s40520-021-01839-6
32. Ünsal F, Murat M, Levent K (2021) SARC-F as a case-finding tool in sarcopenia: valid or unnecessary? Aging Clin Exp Res. https://doi.org/10.1007/s40520-021-01838-7

The short-form SarQoL questionnaire

Quality of Life Research
https://doi.org/10.1007/s11136-021-02823-3

Development and validation of a short version of the Sarcopenia Quality of Life questionnaire: the SF-SarQoL

A. Geerinck[1] · C. Beaudart[1] · J.-Y. Reginster[1,2] · M. Locquet[1] · C. Monseur[3] · S. Gillain[4] · O. Bruyère[1]

Accepted: 11 March 2021

Abstract

Purpose To facilitate the measurement of quality of life in sarcopenia, we set out to reduce the number of items in the previously validated Sarcopenia Quality of Life (SarQoL®) questionnaire, and to evaluate the clinimetric properties of this new short form.

Methods The item reduction process was carried out in two phases. First, information was gathered through item-impact scores from older people ($n = 1950$), a Delphi method with sarcopenia experts, and previously published clinimetric data. In the second phase, this information was presented to an expert panel that decided which of the items to include in the short form. The newly created SFSarQoL was then administered to older, community-dwelling participants who previously participated in the SarcoPhAge study. We examined discriminative power, internal consistency, construct validity, test–retest reliability, structural validity and examined item parameters with a graded response model (IRT).

Results The questionnaire was reduced from 55 to 14 items, a 75% reduction. A total of 214 older, community-dwelling people were recruited for the validation study. The clinimetric evaluation showed that the SF-SarQoL® can discriminate on sarcopenia status [EWGSOP2 criteria; 34.52 (18.59–43.45) vs. 42.86 (26.56–63.69); $p = 0.043$], is internally consistent ($\alpha = 0.915$, $\omega = 0.917$) and reliable [ICC = 0.912 (0.847–0.942)]. A unidimensional model was fitted (CFI = 0.978; TLI = 0.975; RMSEA = 0.108, 90% CI 0.094–0.123; SRMR = 0.055) with no misfitting items and good response category separation.

Conclusions A new, 14-item, short form version of the Sarcopenia Quality of Life questionnaire has been developed and shows good clinimetric properties.

Keywords Sarcopenia · Quality of life · Questionnaire development · Item response theory · Item reduction

Background

The process of ageing is associated with numerous physiological changes. One of these changes is the age-related decrease in muscle mass and function known as sarcopenia, which has received a great deal of interest in the past decade [1, 2].

Sarcopenia is described by the European Working Group on Sarcopenia in Older People (EWGSOP) as "*a progressive and generalized skeletal muscle disorder that is associated with increased likelihood of adverse outcomes including falls, fractures, physical disability and mortality*" [3]. The most recent consensus criteria of the EWGSOP2 state that low muscle strength is an indicator of probable sarcopenia, low strength in combination with low muscle mass is confirmed sarcopenia, and low muscle strength, low muscle mass and low physical performance is severe sarcopenia [3].

✉ A. Geerinck
anton.geerinck@uliege.be

1 Division of Public Health, Epidemiology and Health Economics, World Health Organization Collaborating Centre for Public Health Aspects of Musculoskeletal Health and Ageing, University of Liège, Liège, Belgium

2 Chair for Biomarkers of Chronic Diseases, Biochemistry Department, College of Science, King Saud University, Riyadh, Kingdom of Saudi Arabia

3 Department of Education Sciences, University of Liège, Liège, Belgium

4 Geriatrics Department, University Hospital of Liège, Liège, Belgium

Published online: 30 March 2021

Sarcopenia has been associated with increased mortality, functional decline, a higher rate of falls and a higher incidence of hospitalization [4, 5]. In the last few years, evidence has been accumulating on the adverse impact of sarcopenia on quality of life [6, 7].

In 2015, Beaudart and colleagues presented the Sarcopenia Quality of Life (SarQoL) questionnaire, an auto-administered patient-reported outcome measure specifically designed to measure quality of life in older, community-dwelling people [8]. It is still currently the only instrument measuring quality of life validated for sarcopenic samples and the only sarcopenia-specific QoL questionnaire available.

The clinimetric properties of the SarQoL questionnaire have been examined for 11 language-specific versions of the questionnaire and has demonstrated strong measurement properties [9–20]. The questionnaire has been extensively translated, and is available in 30 languages from the website www.sarqol.org.

The comprehensive nature of the SarQoL® questionnaire, which allows it to probe multiple facets of QoL in sarcopenia, means a trade-off has been made between its comprehensiveness and its response burden. Several factors may contribute to the perception of burden on the part of the respondent, such as the length of the questionnaire, the formatting, the instructions, the invasiveness of the questions and the cognitive load the questions put on the respondent [21]. While the developers estimated, based on the results of a pre-test in the target population, that it would take most patients about 10 min to complete the SarQoL®, in practice a considerable number of respondents need more time than this. Given that most clinical studies administer a number of tests and questionnaires, and thus need to take into consideration the response burden of each instrument so as not to jeopardize the accuracy of the obtained data and the percentage of missing responses, a shorter version of the SarQoL® questionnaire might prove valuable.

The first objective of this study was to extract a shorter version out of the 55 items of the SarQoL® questionnaire which safeguards the conceptual structure and the content validity of the original instrument. The second objective was to investigate the clinimetric properties of the newly developed short-form SarQoL.

Methods

Development phase

The SarQoL questionnaire

The short form described in this article was developed from the Sarcopenia Quality of Life (SarQoL) questionnaire. This auto-administered patient-reported outcome measure was developed with the specific aim of evaluating quality of life in sarcopenic, community-dwelling older people. The SarQoL measures QoL through 55 items categorized into seven domains of health-related dysfunction: physical and mental health, locomotion, body composition, functionality, activities of daily living, leisure activities, and fears [8]. The response options of the SarQoL questionnaire are a mix of Likert scales (3, 4, or 5 levels) and multiple-answer multiple-choice questions. The scoring algorithm calculates an overall QoL score which is scaled from 20 to 100 points (with complete data), and also provides seven domain scores, scaled from 0 (worst QoL possible) to 100 (best QoL possible) points. The scoring algorithm is not publicly available, but tools to calculate the scores are available by contacting info@sarqol.org. The clinimetric properties of the questionnaire have been evaluated in 11 different language-specific versions, and considerable information is available for known-groups validity, construct validity, internal consistency, floor and ceiling effects, test–retest reliability, standard error of measurement, smallest detectable change, and an evaluation of the responsiveness of the SarQoL has also been carried out [9–20]. Based on these results, the SarQoL is considered to be a valid, reliable and responsive instrument. The SarQoL questionnaire itself and additional information on the various publications are available from www.sarqol.org.

Item selection process

The objectives of the item reduction process were to create a significantly shorter version of the SarQoL questionnaire that would represent as much of the conceptual model of the Overall QoL score of the original questionnaire as possible, and thus also be highly correlated with the same score.

The item selection process was carried out in two phases, presented in Fig. 1. The first phase served to collect and collate as much information on the properties of the items and domains in the SarQoL questionnaire. This phase started off with the calculation of item-impact scores to determine which items in the SarQoL questionnaire are the most relevant and impactful for sarcopenic people. For this purpose, we combined data collected in Brazil, the Czech Republic, the UK, Belgium (two separate cohorts), Poland, Spain and Switzerland. All data were collected in non-interventional studies (transversal and cohort) from community-dwelling older people (60 years and older) who were evaluated for sarcopenia according to the EWGSOP criteria [22]. In total, data from 1950 participants were included in this dataset, of which 267 were diagnosed as sarcopenic. By calculating the prevalence of an item occurring (those that experienced an item divided by those that did not) and dividing this by the mean impact, a ranking was established from most relevant

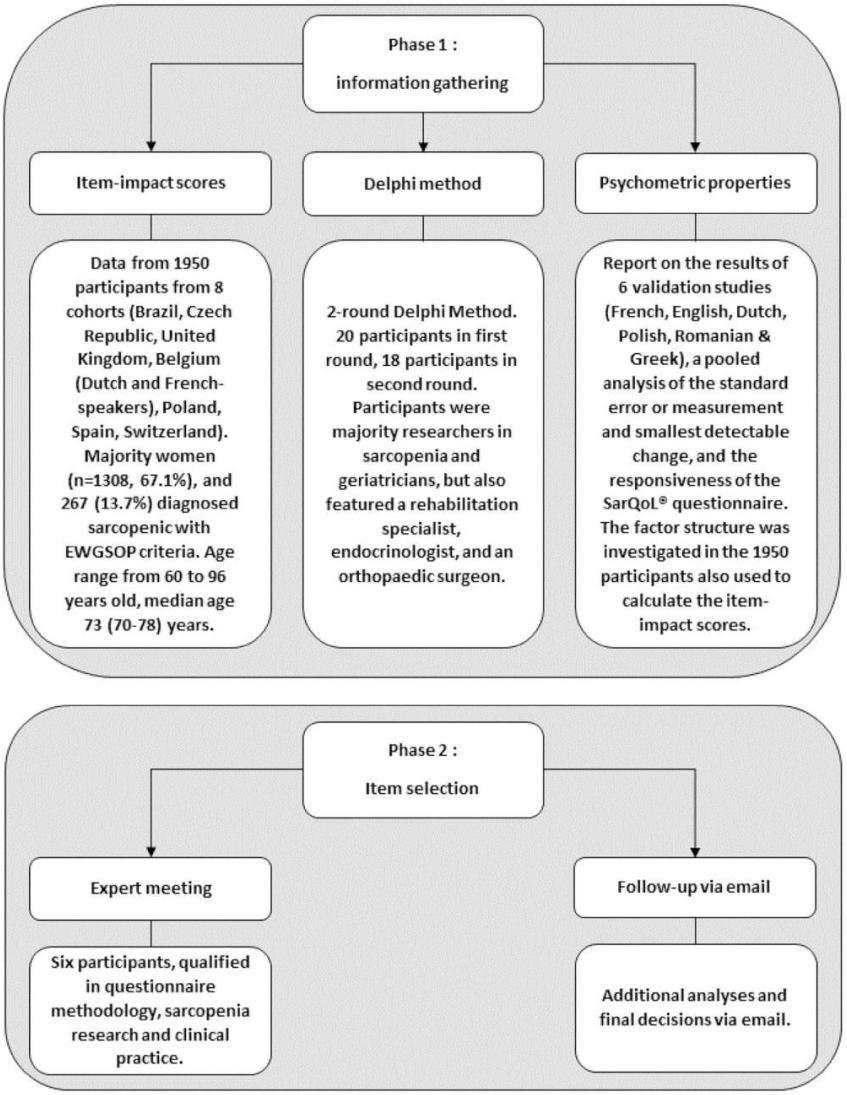


Fig. 1 Development of the SF-SarQoL

and impactful to least [23]. The first phase of the item selection process continued with a 2-round modified Delphi method, so that the patient's perspective quantified by the item-impact scores could be complemented with the opinion of health care professionals and researchers. We targeted researchers and clinicians involved in sarcopenia research who had previous experience with the SarQoL questionnaire, through use, translation, validation or development, and invited them to participate. The participants were provided with an Excel file wherein they were able to categorize each of the 55 items as either "must absolutely be kept in a short form" or "could be discarded". Items were organized and presented per domain. In the second round, the participants were once again asked to categorize the items in the SarQoL questionnaire (keep or discard), but were now also provided the item-impact scores as well as the percentage of participants who agreed on whether to keep or discard an item in the first round. Consensus at the end of the second round was defined as 70% agreement. During both rounds, participants were able to add comments on their choices. The information from the Delphi method, the item-impact scores, and the already published information concerning the clinimetric properties of the SarQoL questionnaire was summarized into a report at the end of the first stage.

In the second phase of the item reduction process the report compiled at the end of phase one was presented to an expert group consisting of researchers specialized in sarcopenia and QoL, a clinical practitioner and a questionnaire

methodologist (AG, CB, OB, ML, CM, SG). These discussed the available information and decided on the inclusion or exclusion of a number of items. As recommended in the guidelines formulated by Goetz et al., the expert group was asked to consider content validity (i.e. the results from the item-impact study and the Delphi method) as having the most weight in the decision-making process, followed by clinimetric properties and finally any additional analyses (factor analysis, correlations, or subgroup analyses) that were performed. To ensure an important reduction of the length of the questionnaire, an arbitrary goal of at least a 65% reduction was chosen at the start of the selection process, while maintaining the relative weight of the seven domains in the SarQoL questionnaire.

Validation phase

Population and study design

For the validation of the SF-SarQoL, we contacted the 314 participants who had previously participated in the fourth and/or fifth year of follow-up of the SarcoPhAge (Sarcopenia and Physical impairment with advancing Age) study [24]. In short, this study recruited older, community-dwelling volunteers from the Liège province of Belgium, and invited them once a year for a battery of physical and other measurements. Given that sarcopenia was the main focus of the SarcoPhAge study, body composition, muscle strength and physical performance were evaluated at each visit with dual-energy X-ray absorptiometry, a hydraulic hand-dynamometer and the Short Physical Performance Battery. Details on the SarcoPhAge study design and results have been reported previously [24, 25]

We provided the participants, through the postal service, with study packets composed of the short form SarQoL questionnaire, the EQ-5D and EQ-VAS questionnaire which are preference-based measures of health status, and the original SarQoL questionnaire. The study packets were accompanied by an explanatory letter and a pre-stamped envelope with which to return the study documents [26]. The people who consented to participate and sent back the completed questionnaires received a second packet by mail about 10 days after the date on which they completed the first packet. The second study packet consisted of the SF-SarQoL and a query on whether their health had changed in the interval between the two administrations of the SF-SarQoL. Demographic and clinical data were obtained from the existing datasets collected during the fourth or fifth year follow-up visits of the SarcoPhAge study. Sarcopenia was diagnosed with the revised consensus criteria from the EWGSOP2 (handgrip strength below 27 kg for men or 16 kg for women, together with low muscle mass defined as appendicular skeletal muscle mass divided by height-squared (ASM/Ht^2) < 7.0 kg/m^2 for men or < 5.5 kg/m^2 for women) [3]. The research protocol (no 2012/277) and its amendment (dated 19/12/2019) were approved by the Ethics Committee of the University Teaching Hospital of Liège.

Clinimetric properties from classical test theory

The clinimetric properties of the SF-SarQoL have been examined with the following indicators from classical test theory:

(1) Item characteristics have been evaluated with percentage of missing responses. Floor and ceiling effects for the overall QoL score of the SF-SarQoL were considered to be present if more than 15% of respondents obtained the lowest (0 points) or highest (100 points) score [27].
(2) Discriminative power (also known as known-groups validity), which measures an instrument's ability to distinguish among distinct groups, has been examined in three separate comparisons: sarcopenic versus non-sarcopenic, probably sarcopenic (low grip strength in the EWGSOP2 algorithm) versus probably non-sarcopenic (normal grip strength), and at high risk of sarcopenia (SARC-F score ≥ 4) versus at low risk of sarcopenia [3, 28]. We expected to find significantly lower QoL scores on the SF-SarQoL for sarcopenic participants, those with low grip strength and those at high risk of sarcopenia. Significant differences in QoL were established with the Student *t* test or the Mann–Whitney *U* test, depending on normality of distribution of the scores. Point biserial correlation coefficients (*r*) were calculated to provide a measure of the strength of association between group status and QoL.
(3) Internal consistency was measured with both the Cronbach's alpha value and the McDonald omega value. We decided on this approach because the alpha value allows comparison to previous validation studies, while the omega value avoids some of the problems associated with the alpha value and is considered to be a more accurate reflection of internal consistency [29]. For both indicators, values between 0.7 and 0.95 indicate that the items in the questionnaire are closely interrelated and measure the same concept [27].
(4) Test–retest reliability has been quantified with the intraclass correlation coefficient (ICC—two-way mixed model and absolute agreement type) for the total score of the SF-SarQoL, and with weighted kappa coefficients (using quadratic weights) for the individual items. An ICC value greater than 0.7 indicates acceptable reliability [27]. For the weighted kappa coefficients, a value ≥ 0.8 is almost-perfect agreement, ≥ 0.6 and < 0.8 is substantial agreement, ≥ 0.4 and < 0.6 is

moderate agreement, ≥ 0.2 and < 0.4 is fair agreement and < 0.2 is slight agreement [30]. Only those participants who participated in both administrations of the SF-SarQoL, whose health did not change in the interval period, and who completed the second questionnaire a maximum of 3 weeks after the first, were eligible for inclusion in this analysis. A Bland–Altman analysis was also carried out to detect whether there was systematic bias in the test–retest data [31].

(5) The construct validity of the SF-SarQoL has been investigated through three approaches. First, we evaluated criterion validity, where the instrument scores are compared to those of a gold standard. This was measured with the ICC (two-way mixed model and consistency type) between the overall QoL scores of the short form and the original SarQoL questionnaire [27]. Secondly, we tested hypotheses on the expected correlation between the SF-SarQoL and the EQ-5D and EQ-VAS questionnaires, assuming that we will find strong correlations between them [27]. Lastly, we evaluated the structural validity of the SF-SarQoL. We hypothesized that the SF-SarQoL is unidimensional, with all items loading on the latent construct of quality of life, and have carried out a confirmatory factor analysis using the diagonally weighted least squares estimator (WSLMV) for ordinal data using the R package "Lavaan" (version 0.6–6). Model fit was evaluated with the Chi-square test ($p \geq 0.05$ indicates good fit), the comparative fit index (CFI; good fit if ≥ 0.95), the Tucker-Lewis index (TLI; good fit if ≥ 0.95), the root mean square error of approximation (RMSEA; good fit if ≤ 0.08) and the standardized root mean square residual (SRMR; good fit if ≤ 0.08) [32, 33].

Clinimetric properties from modern measurement theory

Before constructing and testing an IRT model, it is important to verify that the items meet the assumptions of unidimensionality, local independence and monotonicity [34].

(1) Most IRT applications require a factor structure with a single latent trait, hence the need to establish whether the instrument in question is unidimensional. This was established using the results of the CFA described in the previous paragraph, supplemented with an exploratory factor analysis. Before launching the EFA, we inspected the suitability of the data using Bartlett's test of sphericity and the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy. The EFA was executed on the polychoric correlation matrix with the WLSMV estimator from the R package "Psych" (version 1.9.12.31). The number of factors present was evaluated with parallel analysis (PA) and Velicer's minimum average partial (MAP) test [35].

(2) The second assumption, local independence, means that there should be no correlation between two items after the effect of the underlying trait is filtered out. In other words, the item responses should be entirely a function of the underlying trait, and not (partly) dependent on a second factor [34]. To determine this, we looked at the residual correlation matrix from the previously described single-factor CFA, and considered a value of 0.2 above the average residual correlation as the cut-off for local independence [36].

(3) Lastly, the concept of monotonicity was examined. This concept states that the probability of endorsing a higher item response category should increase with increasing levels of the underlying construct [34]. Monotonicity was evaluated with Mokken scaling carried out with the R package "Mokken" (version 3.0.2), using the scalability coefficient H for each item and the questionnaire in its entirety. The assumption of monotonicity was confirmed if the item scalability coefficients were ≥ 0.3 and the scalability coefficient H_i for the entire questionnaire was ≥ 0.5 [36].

After confirming unidimensionality, local independence and monotonicity, a logistic Graded Response Model (GRM) was fit to the data using the R package "mirt" (version 1.32.1). This model calculates both item thresholds (b) as well as item slopes (a). For the purpose of this analysis, the response options "I do not undertake these types of physical activities" in item 2.1 and 2.2, "not applicable" in item 3.1 and 3.2, "I am unable to walk" in item 4, and "I have never participated in leisure activities" in item 8 were treated as missing responses. The encoding of the responses on item 8 was also re-ordered, going from decreased participation to increased participation. Item fit was examined with the $S - X^2$ indicator, where $p \leq 0.001$ indicates poor fit, and by examining the category characteristic curves. For all items, 3 thresholds were estimated, except for item 8, where only two thresholds were estimated.

Statistical analysis

All analyses were executed with SPSS version 27.0.0, R version 4.0.0. and JASP version 0.13.1.

In addition to the statistical manipulations described in the preceding paragraphs, we also verified normality of distribution for quantitative variables with the Shapiro–Wilk test, by comparing mean and median, and by evaluating the histogram and Q–Q plot. Continuous variables following a Gaussian distribution are reported as mean ± standard deviation, while skewed variables are reported as median (25th percentile–75th percentile). Nominal variables are reported

as absolute (n) and relative (%) frequencies. All comparisons were considered significant at the 5% level ($p \leq 0.5$).

Results

Development

Twenty experts participated in the first round of the modified Delphi method, and eighteen of them participated in both rounds. The panel reached consensus on the inclusion of 13 items and the exclusion of 23 items, with 19 items not reaching the 70% agreement threshold for either option. Together with the item-impact scores, calculated separately for the sarcopenic ($n = 267$) and non-sarcopenic ($n = 1584$) participants, and the clinimetric information already available from previous validation studies, these allowed the expert panel to reach a final decision on the inclusion of 14 items from six domains (physical and mental health, locomotion, body composition, functionality, activities of daily living, and leisure activities), which together constitute the short-form SarQoL questionnaire. The expert panel made the decision to deviate from the original conceptual model by not including an item from domain seven (fears) because the format of the question (items are conditional upon the previous question) and the response options (only a positive answer is identified, a negative response or missing data cannot be separated) rendered item-level analysis problematic. The summarized results from the Delphi method, the item-impact ranking and the final decisions of the expert panel are shown in Table 1. The SF-SarQoL is available in online supplementary 1 and from www.sarqol.org.

Clinimetric evaluation

Participants

A total of 214 older people participated in the validation study for the SF-SarQoL. The median age of the participants was 76 (73–81) years and 63.1% were women. We found 70 (32.7%) participants with probable sarcopenia (low grip strength in the EWGSOP2 algorithm), of whom 21 (9.8%) had confirmed sarcopenia. With the help of the SARC-F questionnaire, we found 30 (14.0%) participants at high risk of sarcopenia. The complete clinical and QoL characteristics are reported in Table 2.

Relationship between short and long form scoring algorithm

To ease interpretation of the QoL scores of the short form questionnaire, it was decided to use a scale going from zero to 100, a deviation from the 20–100 scale of the long form questionnaire. Within the scale, lower scores represent persons whose quality of life is significantly impacted by sarcopenia, and higher scores indicate people with better QoL and a smaller impact of sarcopenia. Figure 2 shows the scatter plot of the short and long form Overall QoL score. From this figure, it can be observed that the short form scores Overall QoL scores are roughly parallel but below the dotted equivalence line, which represents perfect correspondence between the 2 scores.

Clinimetric properties classical test theory

The per-item percentage of missing responses ranged between 0 and 5.6%. Five (2.3%) participants scored zero points on the Overall QoL score of the SF-SarQoL, and 1 (0.5%) person scored 100 points, indicating that there are no floor or ceiling effects in this sample. We found excellent discriminative power when comparing probably sarcopenic versus probably not [32.74 (20.15–43.15) vs. 48.81 (28.57–70.24); $p < 0.001$; $r = -0.342$], sarcopenic versus not sarcopenic [34.52 (18.59–43.45) vs. 42.86 (26.56–63.69); $p = 0.043$; $r = -0.144$] and at high risk of sarcopenia versus low risk [17.86 (6.64–24.05) vs. 46.43 (30.95–65.48); $p < 0.001$; $r = -0.444$]. Internal consistency among the items was excellent with a Cronbach's alpha of 0.915 (95% CI = 0.896–0.930) and a McDonalds' omega value of 0.917 (95% CI = 0.897–0.933). Test–retest reliability was calculated among 133 participants. Within this sub-sample, we found excellent test–retest reliability with an ICC of 0.912 (95% CI = 0.847–0.942) for the overall QoL score of the SF-SarQoL. On an item level, we found moderate to almost-perfect agreement between the first and second administration with weighted kappa coefficients, detailed in Table 3.

A Bland–Altman analysis revealed the presence of a systematic bias of 4.11 (95% CI 2.51; 5.72) points, with higher average scores for the retest scores (50.47 ± 24.82) compared to the test scores (46.36 ± 23.30).

The criterion construct validity, measuring the strength of relationship between the SarQoL overall QoL score and its short form equivalent, was excellent with an ICC of 0.835 (95% CI = 0.789–0.871). It should be noted that the scoring algorithm for the short form and the original SarQoL questionnaire are not on the same metric, and are thus not interchangeable. We also found strong correlations between the SF-SarQoL overall score and the EQ-5D index score ($r = 0.671$; $p < 0.001$) and the EQ-VAS ($r = 0.697$; $p < 0.001$). A confirmatory factor analysis of a one-dimensional model resulted in the following fit indices ($\chi^2 = 269.330$, df = 77, $p < 0.001$; CFI = 0.978; TLI = 0.975; RMSEA = 0.108, 90% CI = 0.094–0.123; SRMR = 0.055). As the five items of question 1 share a common stem, we hypothesized that they would be highly correlated, with would lead to a deterioration of fit indices. To overcome this issue, an alternative

Table 1 Development SF-SarQoL

Domain/item	Delphi method[a]		Item-impact ranking[b,c]		Final decision
	Consensus inclusion	Consensus exclusion	Sarcopenic group	Non-sarcopenic group	
Physical and mental health					
1.1 Loss of arm strength	x		3	6	IN
1.2 Loss of leg strength	x		1	4	IN
1.4 Loss of energy		x	4	5	
2 Muscle pain		x	2	3	
6 Feeling old		x	6	2	
7 Feeling of muscle weakness		x			
8 Feeling of being physically weak			5	1	IN
16 Feeling of being frail			7	7	
Locomotion					
9.1 Limitation in walking time	x		4	4	
9.2 Limitation in number of outings		x	6	6	
9.3 Limitation in walking distance	x		2	2	
9.4 Limitation in walking speed	x		1	1	IN
9.5 Limitation in steps length		x	7	7	
10.1 Feeling of fatigue when walking	x		3	3	IN
10.2 Need of recovery time when walking			7	8	
10.3 Difficulties to cross a road fast enough			9	9	
10.4 Difficulties to walk on uneven ground		x	4	5	
Body composition					
1.3 Loss of muscle mass			2	2	IN
13 Physical change		x	1	1	
14 Weight change (loss or gain)		x			
15 Upset with change		x			
Functionality					
1.5 Loss of physical capacity	x		2	2	IN
1.6 Loss of flexibility		x	3	1	
11 Balance problems			5	4	IN
12 Falls occurrence	x		13	8	
17.1 Climbing one flight of stairs	x		11	13	
17.2 Climbing several flights of stairs		x	6	6	
17.3 Climbing stairs without a banister			8	11	
17.4 Crouching or kneeling			4	5	
17.5 Stooping			10	10	
17.6 To stand up from the floor without any support			1	3	IN
17.7 Get up from a chair	x		9	7	
17.8 To stand from a sitting position			12	12	
18 Limitation of movement	x		7	9	IN
20 Sexuality		x	14	14	
Activities of daily living					
17.11 Take public transportation		x	14	14	
17.12 To get in/out a car		x	12	12	
3.1 Difficulty during light physical effort			5	9	
3.2 Fatigue during light physical effort	x		2	7	
3.3 Pain during light physical effort		x	4	8	
4.1 Difficulty during moderate physical effort			6	6	IN
4.2 Fatigue during moderate physical effort	x		3	3	IN

Table 1 (continued)

Domain/item		Delphi method[a]		Item-impact ranking[b,c]		Final decision
		Consensus inclusion	Consensus exclusion	Sarcopenic group	Non-sarcopenic group	
4.3 Pain during moderate physical effort			x	7	5	
5.1 Difficulty during intense physical effort			x	10	2	
5.2 Fatigue during intense physical effort			x	9	1	
5.3 Pain during intense physical effort			x	11	4	
17.9 Carrying heavy objects			x	1	10	IN
17.10 Opening a bottle or a jar				8	11	
17.13 Shopping			x	15	15	
17.14 Household tasks				13	13	
Leisure activities						
21 Change in physical activities			x	1	1	
22 Change in leisure activities				2	2	IN
Fears						
19	Fear of getting hurt Fear of not succeeding Fear of being tired Fear of falling					

[a]Empty cells indicate that the 70% agreement threshold was not reached

[b]Because certain questions in the SarQoL questionnaire are conditional on other questions (i.e. "If yes on previous question, then …"), item-impact scores could not be calculated for items 7, 14, 15 and 19

[c]Items are ranked from most impactful (1) to least impactful

model was tested, with the five items of question 1 loading on a first latent variable, and the remaining questions on a second latent variable (factor 1: items 1.1 to 1.5; factor 2: items 2.1 to 8) and a correlated residual variance between items 1.5 and 4. This model obtained adequate fit indices: ($\chi^2 = 161.847$, df = 75, $p < 0.001$; CFI = 0.990; TLI = 0.988; RMSEA = 0.074, 90% CI = 0.058–0.089; SRMR = 0.042). The 2 latent variables in this model are highly correlated at $r = 0.894$. Standardized factor loadings for both models are reported in Table 3.

Clinimetric properties modern measurement theory

Confirmatory factor analysis did not conclusively indicate that the SF-SarQoL is unidimensional. Therefore, we investigated further with an exploratory factor analysis, which was considered appropriate when the Bartlett's test returned a p value < 0.001 and the KMO test a value of 0.87. Parallel analysis identified a single factor in the data, as did the Velicer's MAP test, which achieved a minimum of 0.05 with 1 factor. There were no locally dependent items found, with no residual correlations greater than the cut-off of 0.184 or − 0.216 (average residual correlation = − 0.016). The monotonicity assumption was confirmed when scalability coefficients H_i between 0.517 ("balance problems") and 0.716 ("reduction physical capacity") were found, alongside a Mokken scalability coefficient H for the entire short form of 0.635.

After fitting the logistic Graded Response Model to the data, we found no misfitting items, as evidenced by the fact that no p values for the $S - X^2$ indicator were smaller than 0.001. The item with the lowest discriminative ability was found to be "leisure activities" ($a = 1.518$) and the most discriminative item was "reduction of physical capacity" ($a = 3.791$). The item thresholds were spread out from − 1.889 ("Carrying heavy objects") to 1.756 ("Tired moderate effort"). Detailed results on the model fit and item parameters are reported in Table 4. The category characteristics curves, a visual representation of the item parameters, are shown in Fig. 3.

Discussion

This article describes the development of a 14-item short form version of the SarQoL® questionnaire, and the subsequent examination of its clinimetric properties.

The item reduction process follows the guidelines formulated by Goetz et al. by, among other things, prioritizing content validity over statistical properties [37]. The 2-phase process employed led to the inclusion of 14 items from six domains, preserving, as much as possible,

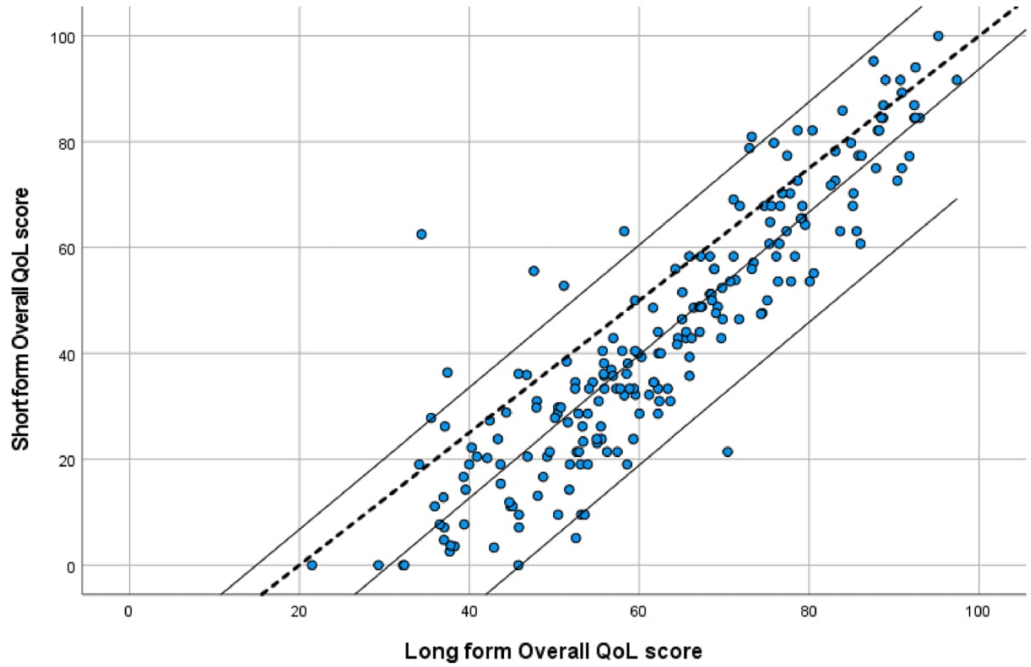
Table 2: Characteristics of the sample

	n (%)	Median (P25–P75)
Gender		
Male	80 (36.9%)	
Female	137 (63.1%)	
Age (years)		76 (73–81)
Probable sarcopenia (with EWGSOP2)		
Yes	70 (32.7%)	
No	143 (66.8%)	
Sarcopenia (with EWGSOP2)		
Yes	21 (9.8%)	
No	193 (90.2%)	
At risk of sarcopenia (with SARC-F)		
Yes	30 (14.0%)	
No	184 (86.0%)	
EQ-5D index score		0.800 (0.747–0.827)
EQ-VAS		70 (60–80)
SarQoL		
Physical and mental health		60.54 (48.87–73.30)
Locomotion		55.56 (41.67–75.70)
Body composition		62.50 (48.96–70.83)
Functionality		66.69 (55.36–82.28)
Activities of daily living		60.00 (48.21–76.67)
Leisure activities		33.25 (33.25–66.50)
Fears		87.50 (75.00–100.00)
Overall QoL score		61.97 (51.57–75.64)
SF-SarQoL overall QoL score first administration		40.24 (23.81–62.64)
SF-SarQoL overall QoL score second administration		47.62 (31.55–70.24)

the conceptual structure of the original SarQoL® questionnaire in the short form. One domain (D7: fears) did not contribute to the short form because, in the original questionnaire, it is dependent on the response of a different item that is not a part of domain seven. This type of conditional question ("If yes to the previous question, then …) combined with the fact that the response options for the items in question 19 make it impossible to distinguish between missing data and negative responses, made it inopportune in the eyes of the expert committee to include this domain. On top of the problems caused by its phrasing and response option, the participants in the Delphi method did not reach consensus on its inclusion, so these items and domain was not included in the short form. The questionnaire was thus reduced from 55 to 14 items, a 75% reduction.

In contrast with the original questionnaire, the newly created SF-SarQoL does not provide domain scores, but only an Overall QoL score. This is a conscious choice because, in our estimation, the original SarQoL® questionnaire is better suited when researchers wish to look at QoL on a domain-level. The SF-SarQoL is better suited to studies that use QoL as a secondary outcome, or in association with a general QoL instrument, and, in this vein, it privileges a single QoL score.

The validation part of this study found good to excellent results for discriminative power, construct validity, internal consistency, test–retest reliability and an absence of floor and ceiling effects. However, despite an ICC of 0.912 (95% CI = 0.847–0.942) for the test–retest reliability, we did find a systematic bias of 4.11 (95% CI = 2.51; 5.72) points. An earlier analysis of the original SarQoL® questionnaire in a sample of 274 sarcopenic participants demonstrated no such bias [0.18 (− 0.26; 0.63) points], so this result was unexpected [11]. It is unclear how this bias originated and whether it is a feature of the questionnaire or a one-off event, specific to this sample. It is possible that the higher QoL scores recorded during the second administration of the SF-SarQoL may be due to the packet length (19 pages for the first packet versus 6 pages for the second packet), or due to the information on sarcopenia received with the first packet, and which was absent in the second packet. Future validation studies should prioritize investigating test–retest reliability and, hopefully, clarify this issue. Confirmatory factor analysis did not conclusively confirm the unidimensional nature of the SF-SarQoL, with a 2-factor model showing better



Legend: The dotted line provides a reference for equivalence between short and long form scores (from 0:20 to 100:100). A linear fit line with 95% prediction interval is also provided (R^2=0.816).

Fig. 2 Relation between short form scores and the long form Overall QoL scores

fit than the unidimensional model. The graded response model did not indicate any misfitting items. The item trace lines show good separation between the different response categories.

Overall, the SF-SarQoL displays adequate to good clinimetric properties, allowing its use in research, clinical trials and clinical practice. Potential users should consider the objectives of their research when choosing between the 55-item or the 14-item SarQoL® questionnaire. If QoL is a primary outcome, the original SarQoL® questionnaire provides a superior level of detail and precision, as well as scores for the seven QoL domains on top of the overall QoL score. However, if QoL is not the main objective, and response burden is a serious consideration, the SF-SarQoL could be the more appropriate tool.

An important remark to make is that the scores on the original SarQoL® questionnaire and the newly developed SF-SarQoL are not interchangeable and should not be compared head-to-head. During the discussions on the scoring algorithm to be created for the short form SarQoL questionnaire, we examined the complexities of the original scoring algorithm, and a choice was made to place the SF-SarQoL on a 0–100 scale where the score range for the original SarQoL® questionnaire is about 20–100 points.

This study has several strengths: we followed the guidelines by Goetz et al., prioritized content validity, administered the SF-SarQoL in an independent sample and performed as complete a validation as possible with elements from both classical test theory and modern measurement theory.

However, this study also has some limitations: we did not perform differential item functioning analysis because of concerns about the sample size. We fully intend to rectify this once we are able to assemble sufficient data, preferably from multiple countries. We were unable to integrate the domain "fears" into the short-form, so a certain amount of content was lost during the item reduction process. Our sample size of 214 participants is sufficient for the performed statistical manipulations, but does not permit subgroup analyses. The members of the Delphi panel were selected

Table 3 Test–retest reliability and construct validity

	Concordance of items between test and retest ($n = 133$)		Standardized factor loadings		
	Weighted kappa (95% CI)	Interpretation[a]	Model 1	Model 2[b]	
				Factor 1	Factor 2
1.1 Reduction strength arms	0.794 (0.658–0.840)	Substantial	0.695	0.725	
1.2 Reduction strength legs	0.735 (0.637–0.834)	Substantial	0.897	0.930	
1.3 Reduction muscle mass	0.682 (0.590–0.773)	Substantial	0.806	0.827	
1.4 Reduction physical capacity	0.613 (0.495–0.732)	Substantial	0.917	0.951	
1.5 Reduction length of walks	0.750 (0.673–0.828)	Substantial	0.867	0.873	
2.1 Difficulty moderate effort	0.691 (0.541–0.842)	Substantial	0.901		0.915
2.2 Tired moderate effort	0.646 (0.485–0.808)	Substantial	0.856		0.864
3.1 Get up from floor	0.683 (0.512–0.854)	Substantial	0.786		0.802
3.2 Carrying heavy objects	0.546 (0.335–0.756)	Moderate	0.821		0.833
4 Tired when walking	0.798 (0.732–0.865)	Substantial	0.874		0.866
5 Feel weak	0.791 (0.709–0.873)	Substantial	0.877		0.900
6 Balance problems	0.867 (0.812–0.921)	Almost perfect	0.673		0.689
7 Limit movements	0.728 (0.637–0.819)	Substantial	0.850		0.868
8 Leisure activities	0.406 (0.185–0.627)	Moderate	0.594		0.605

[a]Kappas interpreted according to Landis and Koch, where ≥ 0.8 is almost-perfect agreement, ≥ 0.6 and < 0.8 is substantial agreement, ≥ 0.4 and < 0.6 is moderate agreement, ≥ 0.2 and < 0.4 is fair agreement, and < 0.2 is slight agreement

[b]Model 2 is a 2-factor model with correlated residual variance between items 1.5 and 4

Table 4 Graded response model

Item	Monotonicity	Model fit	Item parameters			
	H_i	p value $S - X^{2a}$	a	b_1	b_2	b_3
1.1 Reduction strength arms	0.526	0.061	1.691	− 1.519	0.277	1.579
1.2 Reduction strength legs	0.681	0.407	3.515	− 0.618	0.314	1.388
1.3 Reduction muscle mass	0.590	0.460	2.278	− 1.140	0.292	1.499
1.4 Reduction physical capacity	0.716	0.365	3.791	− 1.012	0.494	1.594
1.5 Reduction length of walks	0.653	0.204	2.940	− 0.543	0.415	1.461
2.1 Difficulty moderate effort	0.695	0.176	3.592	− 0.478	0.361	1.262
2.2 Tired moderate effort	0.651	0.072	2.790	− 0.416	0.618	1.756
3.1 Get up from floor	0.591	0.001	2.219	− 1.002	0.193	1.378
3.2 Carrying heavy objects	0.653	0.497	2.544	− 1.889	− 0.314	1.012
4 Tired when walking	0.645	0.068	3.176	− 0.673	0.247	1.425
5 Feel weak	0.687	0.476	3.386	− 0.967	0.234	1.210
6 Balance problems	0.517	0.632	1.581	− 1.362	0.022	1.298
7 Limit movements	0.697	0.269	2.954	− 1.335	− 0.110	0.709
8 Leisure activities	0.557	0.435	1.518	0.017	3.229	NA

[a]$S - X^2$ statistic calculated on 160 complete observations

for their previous knowledge of the SarQoL® questionnaire, and were not necessarily representative of the wider community of sarcopenia researchers and geriatricians. Due to the transversal nature of the performed validation study, we were unable to examine the responsiveness of the new SF-SarQoL. Evaluating this property of the SF-SarQoL should be a priority for future research.

In conclusion, this article presented the development process and the validation of a 14-item short form version of the SarQoL® questionnaire. In an independent sample, the SF-SarQoL demonstrated adequate measurement properties to allow its use. While its responsiveness should still be investigated, we fully recommend its use in situations where the original 55-item SarQoL® questionnaire is deemed to be too much of a burden on the respondents.

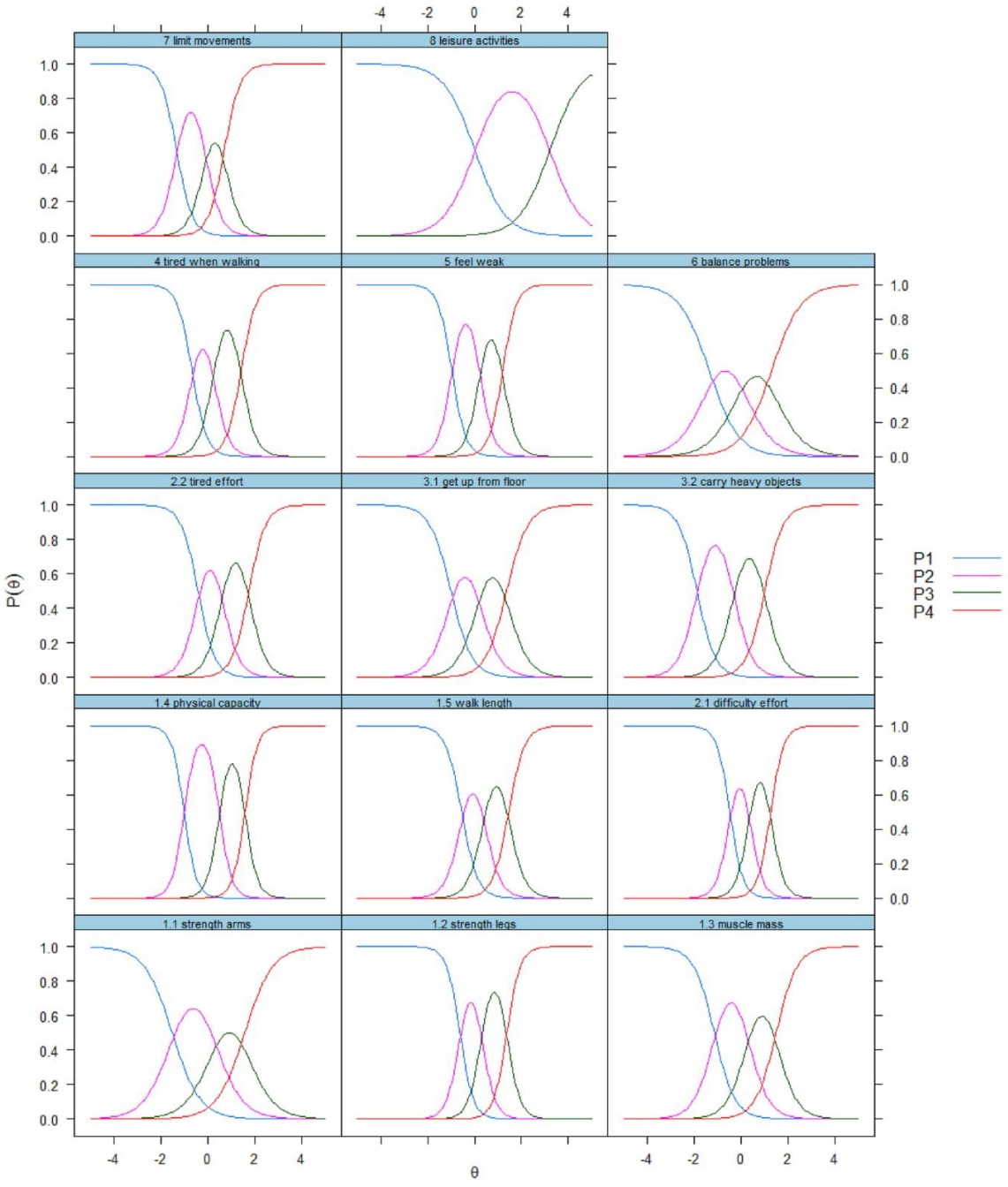


Fig. 3 Category characteristic curves of the 14 items analyzed in the SF-SarQoL

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11136-021-02823-3.

Author contributions This study was conceptualized by AG, CB, OB and J-YR. All authors (AG, CB, J-YR, ML, CM, SG, OB) participated in the development phase. AG, CB and ML collected the data. AG was responsible for data analysis and wrote the first draft of the manuscript. All authors revised the draft article and approved the final product.

Funding AG is supported by a FRIA doctoral grant from the F.R.S-FNRS. A portion of the expenses associated with the validation study were covered by the Léon Fredericq foundation, a non-profit organization associated with the university teaching hospital of Liège.

Data availability The dataset has been deposited on the Open Science Network (OSF) and can be consulted via the following link: http://www.doi.org/10.17605/OSF.IO/3PSZM

Code availability R scripts have been deposited on the Open Science Network (OSF) and can be consulted via the following link: http://www.doi.org/10.17605/OSF.IO/3PSZM

Declarations

Conflict of interest CB, J-YR and OB are shareholders of SarQoL sprl, a spin-off of the University of Liège. All other authors declare no conflicts of interests.

Ethical approval This study received the approval of the Ethics Committee of the University Teaching Hospital of Liège in an amendment approved on 19/12/2019 to research protocol n° 2012/277.

Informed consent Informed consent was obtained from all individual participants included in the study.

Informed consent Data collected in the study were anonymized and do not allow identification of individual study participants.

References

1. Peña Ordóñez, G. G., Bustamante Montes, L. P., Ramírez Duran, N., Sánchez Castellano, C., & Cruz-Jentoft, A. J. (2017). Populations and outcome measures used in ongoing research in sarcopenia. *Aging Clinical and Experimental Research, 29*, 695–700.
2. Reginster, J. Y., Beaudart, C., Al-Daghri, N., Avouac, B., Bauer, J., Bere, N., et al. (2020). Update on the ESCEO recommendation for the conduct of clinical trials for drugs aiming at the treatment of sarcopenia in older adults. *Aging Clinical and Experimental Research*. https://doi.org/10.1007/s40520-020-01663-4.
3. Cruz-Jentoft, A. J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., et al. (2018). Sarcopenia: Revised European consensus on definition and diagnosis. *Age and Ageing, 48*, 16–31.
4. Beaudart, C., Zaaria, M., Pasleau, F., Reginster, J.-Y., Bruyère, O., & Stenroth, L. (2017). Health outcomes of sarcopenia: A systematic review and meta-analysis. *PLoS ONE, 12*, e0169548.
5. Zhao, Y., Zhang, Y., Hao, Q., Ge, M., & Dong, B. (2019). Sarcopenia and hospital-related outcomes in the old people: A systematic review and meta-analysis. *Aging Clinical and Experimental Research, 31*, 5–14.
6. Woo, T., Yu, S., & Visvanathan, R. (2016). Systematic literature review on the relationship between biomarkers of sarcopenia and quality of life in older people. *The Journal of Frailty & Aging, 5*, 88–99.
7. Tsekoura, M., Kastrinis, A., Katsoulaki, M., Billis, E., & Gliatis, J. (2017). Sarcopenia and its impact on quality of life. *Advances in Experimental Medicine and Biology, 987*, 213–218.
8. Beaudart, C., Biver, E., Reginster, J. Y., Rizzoli, R., Rolland, Y., Bautmans, I., et al. (2015). Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: The SarQoL. *Age and Ageing, 44*, 960–966.
9. Beaudart, C., Biver, E., Reginster, J.-Y., Rizzoli, R., Rolland, Y., Bautmans, I., et al. (2017). Validation of the SarQoL, a specific health-related quality of life questionnaire for Sarcopenia. *Journal of Cachexia, Sarcopenia and Muscle, 8*, 238–244.
10. Beaudart, C., Edwards, M., Moss, C., Reginster, J. Y., Moon, R., Parsons, C., et al. (2017). English translation and validation of the SarQoL®, a quality of life questionnaire specific for sarcopenia. *Age and Ageing, 46*, 271–277.
11. Geerinck, A., Alekna, V., Beaudart, C., Bautmans, I., Cooper, C., De Souza, O. F., et al. (2019). Standard error of measurement and smallest detectable change of the Sarcopenia Quality of Life (SarQoL) questionnaire: An analysis of subjects from 9 validation studies. *PLoS ONE, 14*, e0216065.
12. Erdogan, T., Eris, S., Avci, S., Oren, M. M., Kucukdagli, P., Kilic, C., et al. (2021). Sarcopenia quality-of-life questionnaire (SarQoL)®: Translation, cross-cultural adaptation and validation in Turkish. *Aging Clinical and Experimental Research*. https://doi.org/10.1007/s40520-020-01780-0.
13. Geerinck, A., Scheppers, A., Beaudart, C., Bruyère, O., Vandenbussche, W., Bautmans, R., et al. (2018). Translation and validation of the Dutch SarQoL ®, a quality of life questionnaire specific to sarcopenia. *Journal of Musculoskeletal and Neuronal Interactions, 18*, 463–472.
14. Tsekoura, M., Billis, E., Gliatis, J., Tsepis, E., Matzaroglou, C., Sakkas, G. K., et al. (2018). Cross cultural adaptation of the Greek sarcopenia quality of life (SarQoL) questionnaire. *Disability and Rehabilitation, 42*, 1–7.
15. Konstantynowicz, J., Abramowicz, P., Glinkowski, W., Taranta, E., Marcinowicz, L., Dymitrowicz, M., et al. (2018). Polish validation of the SarQoL®, a Quality of Life Questionnaire specific to sarcopenia. *Journal of Clinical Medicine, 7*, 323.
16. Gasparik, A. I., Mihai, G., Beaudart, C., Bruyere, O., Pop, R.-M., Reginster, J.-Y., et al. (2018). Correction to: Psychometric performance of the Romanian version of the SarQoL(R), a health-related quality of life questionnaire for sarcopenia. *Archives of Osteoporosis, 13*, 98.
17. Alekna, V., Kilaite, J., Tamulaitiene, M., Geerinck, A., Mastaviciute, A., Bruyère, O., et al. (2019). Validation of the Lithuanian version of sarcopenia-specific quality of life questionnaire (SarQoL®). *European Geriatric Medicine, 10*(5), 761–767.
18. Safonova, Y. A., Lesnyak, O. M., Baranova, I. A., Suleimanova, A. K., & Zotkin, E. G. (2019). Russian translation and validation

of SarQoL®—Quality of life questionnaire for patients with sarcopenia. *Nauchno-Prakticheskaya Revmatol, 57*, 38–45.
19. Fábrega-Cuadros, R., Martínez-Amat, A., Cruz-Díaz, D., Aibar-Almazán, A., & Hita-Contreras, F. (2020). Psychometric properties of the Spanish version of the sarcopenia and quality of life, a Quality of Life Questionnaire Specific for Sarcopenia. *Calcified Tissue International, 106*, 274–282.
20. Geerinck, A., Bruyère, O., Locquet, M., Reginster, J. Y., & Beaudart, C. (2018). Evaluation of the responsiveness of the SarQoL® questionnaire, a patient-reported outcome measure specific to sarcopenia. *Advances in Therapy, 35*, 1842–1858.
21. U.S. Department of Health and Human Services F and DA, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health (2009). Guidance for industry use in medical product development to support labeling claims guidance for industry. Retrieved from https://www.fda.gov/media/77832/download.
22. Cruz-Jentoft, A., Baeyens, J. P., Bauer, J., Boirie, Y., Cederholm, T., Landi, F., et al. (2010). Sarcopenia: European consensus on definition and diagnosis. *Age and Ageing, 39*, 412–423.
23. Broder, H. L., McGrath, C., & Cisneros, G. J. (2007). Questionnaire development: Face validity and item impact testing of the child oral health impact profile. *Community Dentistry and Oral Epidemiology, 35*, 8–19.
24. Beaudart, C., Reginster, J. Y., Petermans, J., Gillain, S., Quabron, A., Locquet, M., et al. (2015). Quality of life and physical components linked to sarcopenia: The SarcoPhAge study. *Experimental Gerontology, 69*, 103–110.
25. Locquet, M., Beaudart, C., Hajaoui, M., Petermans, J., Reginster, J. Y., & Bruyère, O. (2018). Three-year adverse health consequences of sarcopenia in community-dwelling older adults according to 5 diagnosis definitions. *Journal of the American Medical Directors Association, 20*, 43-46.e2.
26. The EuroQol Group. (1990). EuroQol: A new facility for the measurement of health-related quality of life. *Health Policy (New York), 16*, 199–208.
27. de Vet, H. C. W., Terwee, C. B., & Mokkink, L. B. (2011). *Measurement in medicine: A practical guide*. . University Press.
28. Malmstrom, T. K., & Morley, J. E. (2013). SARC-F: A simple questionnaire to rapidly diagnose sarcopenia. *Journal of the American Medical Directors Association, 14*, 531–532.
29. Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*, 399–412.
30. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159.
31. Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 2802*, 135–160.
32. Brown, T. (2015). *Confirmatory factor analysis for applied research*. (2nd ed.). The Guildford Press.
33. Kline, R. (2016). *Principles and practice of structural equation modeling*. (4th ed.). The Guildford Press.
34. DeMars, C. (2010). *Item response theory*. . Oxford University Press.
35. Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human Computer Interaction, 32*, 51–62.
36. Lameijer, C. M., Van Bruggen, S. G. J., Haan, E. J. A., Van Deurzen, D. F. P., Van Der Elst, K., Stouten, V., et al. (2020). Graded response model fit, measurement invariance and (comparative) precision of the Dutch-Flemish PROMIS® upper extremity V2.0 item bank in patients with upper extremity disorders. *BMC Musculoskeletal Disorders, 21*, 1–17.
37. Goetz, C., Coste, J., Lemetayer, F., Rat, A. C., Montel, S., Recchia, S., et al. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *Journal of Clinical Epidemiology, 66*, 710–718.

The relative importance of aspects of quality of life

European Geriatric Medicine

Patients' preferences for quality of life aspects in sarcopenia: a best-worst scaling study

--Manuscript Draft--

Manuscript Number:	EGEM-D-21-00232R1
Full Title:	Patients' preferences for quality of life aspects in sarcopenia: a best-worst scaling study
Article Type:	Research Paper
Funding Information:	Fonds Léon Fredericq — Mr. Anton Geerinck Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture — Mr. Anton Geerinck
Abstract:	Purpose As information on patients' preferences regarding quality of life aspects in sarcopenia is lacking, this study aims to assess the relative importance of the 14 items of a QoL questionnaire designed for sarcopenia (the SF-SarQoL) using a best-worst scaling (BWS) survey. Methods Participants, aged 65 years or older and community-dwelling, who previously participated in the SarcoPhAge study, received a BWS survey via the mail. An object case BWS was selected in which participants completed 12 choice tasks, picking the most and least important aspect from 4 out of 14 SF-SarQoL items for each task. Relative importance scores (RIS) were estimated using Hierarchical Bayes modelling. A cluster analysis was also conducted to investigate whether several profiles with regards to QoL preferences were present. Results A total of 163 participants were included, aged 75 (IQR: 73-81) years old, and mostly women (n=107; 65.6%). Two items were found to be significantly more important than others: "feeling a reduction of physical capacity" (RIS=11.26), and "having balance problems" (RIS=11.09). The least important items were "experiencing difficulty carrying heavy objects" (RIS=2.89), and "feeling a reduction in muscle mass" (RIS=3.82). We found a relatively weak evidence for the presence of two clusters. One cluster prioritized items related to falls where the second prioritized items related to feeling physically capable. Conclusion Not all QoL aspects in sarcopenia were equally important. The relative weight of each QoL aspect may be used to interpret QoL results obtained with the SF-SarQoL or inform target outcomes in interventional studies.
Corresponding Author:	Anton Geerinck Université de Liège: Universite de Liege Liège, Belgie BELGIUM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Université de Liège: Universite de Liege
Corresponding Author's Secondary Institution:	
First Author:	Anton Geerinck
First Author Secondary Information:	

Order of Authors:	Anton Geerinck
	Médéa Locquet
	Mickaël Hiligsmann
	Jean-Yves Reginster
	Olivier Bruyère
	Charlotte Beaudart
Order of Authors Secondary Information:	
Author Comments:	Dear Editor, We are pleased to resubmit our manuscript on patients' preferences for quality of life aspects in sarcopenia as established through a best-worst scaling survey. We hope the editor and the reviewers consider the modifications made in response to the remarks from the peer reviewers adequate. Thank you in advance for your consideration.
Response to Reviewers:	Dear reviewers, Thank you again for your remarks and suggestions concerning our manuscript on patient preferences with regards to aspects of quality of life. We have responded to each of the points raised during the peer review in a seperate document, submitted as a supplementary file. We hope you agree with our responses and the modifications we have made to the manuscript. Thank you in advance for your consideration.
Suggested Reviewers:	

Cover page

Patients' preferences for quality of life aspects in sarcopenia: a best-worst scaling study

Authors:

Anton Geerinck [1], Médéa Locquet [1], Mickaël Hiligsmann [2], Jean-Yves Reginster [1,3], Olivier Bruyère [1], Charlotte Beaudart [1,2]

1 Division of Public Health, Epidemiology and Health Economics, World Health Organization Collaborating Centre for Public Health aspects of musculoskeletal health and ageing, University of Liège, Belgium

2 Department of Health Services Research, CAPHRI Care and Public Health Research Institute, Maastricht University, The Netherlands

3 Chair for Biomarkers of Chronic Diseases, Biochemistry Department, College of Science, King Saud University, Riyadh, KSA, Saudi Arabia.

Title:

Patients' preferences for quality of life aspects in sarcopenia: a best worst scaling study

Corresponding author:

Anton Geerinck – anton.geerinck@uliege.be – ORCID 0000-0002-1669-0224

Patients' preferences for quality of life aspects in sarcopenia: a best-worst scaling study

Authors:

Anton Geerinck [1], Médéa Locquet [1], Mickaël Hiligsmann [2], Jean-Yves Reginster [1,3], Olivier Bruyère [1], Charlotte Beaudart [1,2]

1 *Division of Public Health, Epidemiology and Health Economics, World Health Organization Collaborating Centre for Public Health aspects of musculoskeletal health and ageing, University of Liège, Belgium*

2 *Department of Health Services Research, CAPHRI Care and Public Health Research Institute, Maastricht University, The Netherlands*

3 *Chair for Biomarkers of Chronic Diseases, Biochemistry Department, College of Science, King Saud University, Riyadh, KSA, Saudi Arabia.*

Title:

Patients' preferences for quality of life aspects in sarcopenia: a best worst scaling study

Corresponding author:

Anton Geerinck – anton.geerinck@uliege.be – ORCID 0000-0002-1669-0224

Key summary points

Aim:

To assess the relative importance of the 14 items of the SF-SarQoL, a short-form quality of life questionnaire for sarcopenia.

Findings:

Overall, community-dwelling older people considered feeling a reduction of physical capacity, balance problems and reduction of leg strength as the most important quality of life aspects.

Message:

Older people considered some QoL aspects to be more important than others. Adequate management of sarcopenia should take into account the patient's own priorities to maximize benefit to the patient

Abstract

Purpose: As information on patients' preferences regarding quality of life aspects in sarcopenia is lacking, this study aims to assess the relative importance of the 14 items of a QoL questionnaire designed for sarcopenia (the SF-SarQoL) using a best-worst scaling (BWS) survey.

Methods: Participants, aged 65 years or older and community-dwelling, who previously participated in the SarcoPhAge study, received a BWS survey via the mail. An object case BWS was selected in which participants completed 12 choice tasks, picking the most and least important aspect from 4 out of 14 SF-SarQoL items for each task. Relative importance scores (RIS) were estimated using Hierarchical Bayes modelling. A cluster analysis was also conducted to investigate whether several profiles with regards to QoL preferences were present.

Results: A total of 163 participants were included, aged 75 (IQR: 73-81) years old, and mostly women (n=107; 65.6%). Two items were found to be significantly more important than others: "feeling a reduction of physical capacity" (RIS=11.26), and "having balance problems" (RIS=11.09). The least important items were "experiencing difficulty carrying heavy objects" (RIS=2.89), and "feeling a reduction in muscle mass" (RIS=3.82). We found relatively weak evidence for the presence of two clusters. One cluster prioritized items related to falls where the second prioritized items related to feeling physically capable.

Conclusion: Not all QoL aspects were equally important. The relative weight of each QoL aspect may be used to interpret QoL results obtained with the SF-SarQoL or to inform target outcomes in interventional studies.

Keywords: sarcopenia, quality of life, SarQoL, best-worst scaling, cluster analysis

Declarations

Funding: Anton Geerinck is supported by a FRIA doctoral grant from the Fonds de la Recherche Scientifique (F.R.S-FNRS). This study was made possible through the financial support of the Léon Fredericq foundation, a non-profit organization associated with the University Teaching Hospital of Liège.

Conflicts of interest/Competing interests: CB, J-YR & OB are shareholders of SarQoL sprl, a spin-off of the University of Liège. All other authors declare no conflicts of interests.

Ethics approval: This study received the approval of the Ethics Committee of the University Teaching Hospital of Liège in an amendment approved on 19/12/2019 to research protocol n° 2012/277.

Consent to participate: Informed consent was obtained from all individual participants included in the study.

Consent for publication: Not applicable

Availability of data and material: Data and materials are available upon request from the corresponding author.

Code availability: Not applicable

Authors' contributions: AG, CB, OB, MH and J-YR designed the study. AG, ML and CB collected the data. AG and MH performed statistical analysis. AG wrote the first draft of the manuscript. All authors provided feedback on the manuscript and approved the final draft.

1. Introduction

Sarcopenia, the skeletal muscle disorder characterized by a loss of muscle strength and function, can have a significant impact on those affected. It has been shown to be associated with a number of adverse outcomes such as mortality, functional decline, disability, falls and hospitalization [1]. This impact on a personal level cascades into impact on the health systems that provide care to people with sarcopenia, and economic studies have found significantly higher healthcare costs for sarcopenic people both in a hospital setting as well as in the community [2].

Previous research in sarcopenia has mainly focused on so-called hard outcomes (such as mortality or hospitalizations), but interest in the lived experience of sarcopenic patients has been steadily growing. More and more studies are reporting results for quality of life (QoL), mostly concluding that sarcopenic people have lower quality of life compared to non-sarcopenic people [3]. Other examples of patient-reported outcomes are pain, physical function, satisfaction with care, etc. A recent working group organized by the European Society for Clinical and Economic Aspects of Osteoporosis, Osteoarthritis and Musculoskeletal Diseases (ESCEO) emphasized that inclusion of a patient-reported outcome measure (PROM) in clinical trials of pharmaceutical interventions for sarcopenia is highly desirable [4]. The FDA has also encouraged the appropriate use of PROMs in regulatory studies, and has observed a 500% increase in the number of pre-market submissions that include PROMs between 2009 and 2015 [5].

A number of generic QoL questionnaires (i.e., designed for use across different populations) are regularly used in sarcopenia research, most notably the SF-36 and the EQ-5D. A QoL questionnaire specifically designed for sarcopenia, called the Sarcopenia Quality of Life (SarQoL®) questionnaire, has also been available since 2015 [6]. The SarQoL® is recommended for use with older, community-dwelling individuals experiencing a loss in muscle strength and function. It is based on a multidimensional concept of QoL, encompassing 55 items from 7 domains of health-related dysfunction: physical and mental health, locomotion, body composition, functionality, activities of daily living, leisure activities, and fears [6,7]. Recently, a shorter version of the SarQoL® questionnaire was developed, which reduced the length of the questionnaire from 55 to 14 items [8]. The SF-SarQoL questionnaire is available from the website www.sarqol.org in multiple languages.

Most QoL instruments translate the individual responses gathered with the tool in question into one or several scores, representing domains of QoL or the global level of QoL of the respondent. This approach is necessary for quantitative research on groups of people but reduces the complex concept of QoL to a number on a scale. While very useful, it should not be controversial to say that a single score does not tell us the

whole story about a person's QoL. Researchers can often delve deeper into the gathered QoL results, by looking at domain scores or even the item responses themselves, which is already an improvement over an overall score. However, this does not take into account that not all aspects of QoL are created equal: some items are likely to be considered more important by patients than others.

This type of information, the importance of one aspect/item/outcome in relation to others, can be obtained through choice modelling, of which the most frequently used designs are the discrete choice experiment (DCE) and best-worst scaling (BWS). DCE and BWS are already regularly employed to gauge patients' preferences regarding treatments [9,10]. Recently, a DCE was also used to look at which clinical outcomes were considered important by sarcopenic older persons, the first study of its kind in sarcopenia [11,12]. Interestingly, the participants of this study identified QoL as one of the 5 most important outcomes for sarcopenia interventions [12]. In comparison to a DCE, the BWS method is considered to be less cognitively demanding, gathers additional information on the least preferred option and is capable of capturing preferences for a longer list of items/attributes [10,13].

The primary objective of the present study was to establish a ranking from most to least important for the 14 aspects of QoL included in the SF-SarQoL® questionnaire using the best-worst scaling technique. The secondary objective of this study was to explore whether different profiles were present within the sample with regards to their ranking of the 14 aspects of QoL with the help of a cluster analysis.

2. Methods

2.1 Population

This study recruited older, community-dwelling people who had previously participated in the Sarcopenia and Physical Impairment with Advancing Age (SarcoPhAge) study. This is a 5-year cohort study, carried out in the Liège region of Belgium, which focused on a range of musculoskeletal indicators. All participants were aged 65 years and older at inclusion, with a body mass index below 50 kg/m² and without amputated limbs. Details on this study and several articles on different results have previously been published [14]. For the best-worst scaling study presented in this article, 314 individuals who had participated in the interviews for the 4th (July 2017 to September 2018) and/or 5th (June 2018 to November 2019) year of follow-up of the SarcoPhAge study, and for whom demographic and clinical data from these interviews were available, were contacted with an invitation to participate in February/March 2020. The research protocol (n° 2012/277) and its amendment dated 19/12/2019 were approved by the Ethics Committee of the University Teaching Hospital of Liège.

2.2 Study design

Patient preferences were elicited through an object (case 1) BWS survey. This type of choice experiment was first developed by Jordan J. Louvière in 1987, and its use in health care research was proposed in 2005 [13,15]. The objective of this type of choice experiment is to place objects (in this case different aspects of QoL) on an underlying, subjective, latent scale by having volunteers complete choice tasks in which they are asked to indicate the "best" (in this case: most important for QoL) and "worst" (in this case: least important for QoL) object from 3 or more options [15]. By analyzing choice frequency, for both "best" and "worst" choices, utility values can be calculated for each object, and a ranking from best to worst can be established [13].

The 14 items of the SF-SarQoL questionnaire constituted the list used to create the choice tasks in the BWS survey [8]. Twelve choice tasks of 4 items were presented to each participant to strike a balance between obtaining as much information as possible, without creating too much response burden. An example of a choice task from the BWS survey can be found in figure 1. Sawtooth Software was used to generate 2 versions of the BWS survey. The design algorithm of the Sawtooth software is considered to be similar to

the Balanced Incomplete Block Design and takes into account frequency balance, orthogonality, connectivity and positional balance [16]. Participants were randomly assigned to receive either version A or B using IBM's SPSS software.

Figure 1: Example of a choice task in the BWS survey (translated from the original French).

In the table below, the participant has indicated that experiencing difficulty during activities of moderate effort is the most important aspects with regards to their quality of life, and reducing their leisure activities the least important.		
Least important		**Most** important
	Feeling a reduction of the strength in your arms	
	Experiencing difficulty during activities of moderate effort	**X**
	Having problems with your balance	
X	Reducing your leisure activities	

Participants received a paper copy of the BWS questionnaire through the postal service. They completed the questionnaire at home and returned it through the mail using an included pre-paid envelope.

Participants also received and completed the SF-SarQoL questionnaire itself at the same time as the BWS survey. This shorter version of the SarQoL® measures overall QoL through 14 items and has been validated for use in sarcopenia [8]. It provides a single score between 0 and 100 points, with greater scores indicating better QoL.

Clinical and demographic information was obtained from the interviews conducted at the 5th year of the SarcoPhAge study. If no data was collected at the 5th year interview (because of drop-out or missing data), the information collected at the 4th year of follow-up was used. Muscle mass was evaluated with dual x-ray absorptiometry, and muscle strength with a hydraulic hand dynamometer. We used the EWGSOP2 criteria to determine the sarcopenia status of each participant. Those with low grip strength, defined as <27 kg for men and <16 kg for women, were considered to be probably sarcopenic. If the persons with low grip strength also presented with low muscle mass, defined as an appendicular skeletal muscle mass divided by height-squared (ASM/Ht²) <7.0 kg/m² for men and <5.5 kg/m² for women, they were diagnosed as sarcopenic [17]. Participants also completed the SARC-F questionnaire, a screening tool which identifies those with a high probably of being sarcopenic through 5 questions on strength, assistance with walking, rising from a chair,

climbing stairs and falls. Participants who scored 4 or more points (on a scale between 0 and 10) were considered to be likely sarcopenic [18].

2.3 Statistical analysis

The distribution of the continuous demographic and clinical variables was evaluated by looking at the Shapiro-Wilk test, histograms, Q-Q plots, and the distance between mean and median. Variables that were normally distributed are presented as mean ± standard deviation, those that are not presented as median (25^{th} percentile – 75^{th} percentile). Binary variables are presented as absolute and relative frequencies [n(%)].

Relative importance scores (RIS) were estimated using Hierarchical Bayes estimation modelled using multinomial logit. The raw RIS were rescaled so that the sum of all RIS was 100 [16]. RIS are presented as mean (95% confidence interval of the mean). A fit statistic was calculated for each respondent, quantifying the probability that a participant has answered in a random manner. Surveys with a fit statistic below 0.25, indicating a significant probability of random responses by the participant, were excluded [19].

Subgroup analyses were conducted between men and women, as well as between those with normal and low grip strength (defined as <27 kg for men and <16 kg for women). These two variables were chosen because of their importance in interpreting any QoL outcomes if these subgroups showed to place different importance on aspects of QoL. Additional subgroup analyses were performed (and presented in appendix 1) comparing RIS between version A and version B of the BWS survey, between SARC-F score ≥4 points and <4 points, between sarcopenic and non-sarcopenic participants (EWGSOP2 diagnostic criteria), between those aged ≤75 years and > 75 years old and between those with lower QoL (≤47 points for the SF-SarQoL) and those with higher QoL (>47 points). P-values were calculated with Student T-test and Mann-Whitney U-test.

We carried out a cluster analysis on the obtained RIS using the Two-Step cluster strategy with the log-likelihood distance measure using logarithmically transformed versions of the 14 RIS. The number of clusters is selected by the software using the Bayesian Information Criterion. The overall goodness-of-fit of the cluster solution was evaluated with the silhouette measure of cohesion and separation, which ranges from -1 to 1. In a good cluster solution, the intra-cluster distances are small (high cohesion between elements in the same cluster) and the inter-cluster distances are large (good separation between elements from different clusters) [20]. A silhouette coefficient <0.25 indicates the absence of a substantial cluster structure; a value from 0.26 to 0.50 is considered a weak structure that could be artificial; from 0.51 to 0.70 translates to a reasonable structure; and from 0.71 to 1 the cluster solution is considered to be strong [21].

RIS were estimated and rescaled using Sawtooth Software. All statistical manipulations were carried out using SPSS v27.0.0.0. P-values ≤0.05 were considered to be significant.

3. Results

3.1 Population

Out of the 314 study candidates contacted, 163 (52%) sent back the completed BWS survey and were included in the dataset. Detail on the flow of participants throughout this study is provided in figure 2. Of these 163 people, 74 (45.4%) completed version A of the BWS survey, and 89 (54.6%) completed version B. The missing response rate per choice task ranged from 0 (%) to 6 (3.7%) observations, which makes for an average completion rate of 98.3% for the "best" choices and 98.1% for the "worst" choices. The mean fit statistic was 0.537 ± 0.110 and no participant was excluded because of a fit statistic below 0.25. Participants had a median age of 75.0 (73.0-81.0) years, and most were women (n=107, 65.6%). Additional characteristics are provided in table 1.

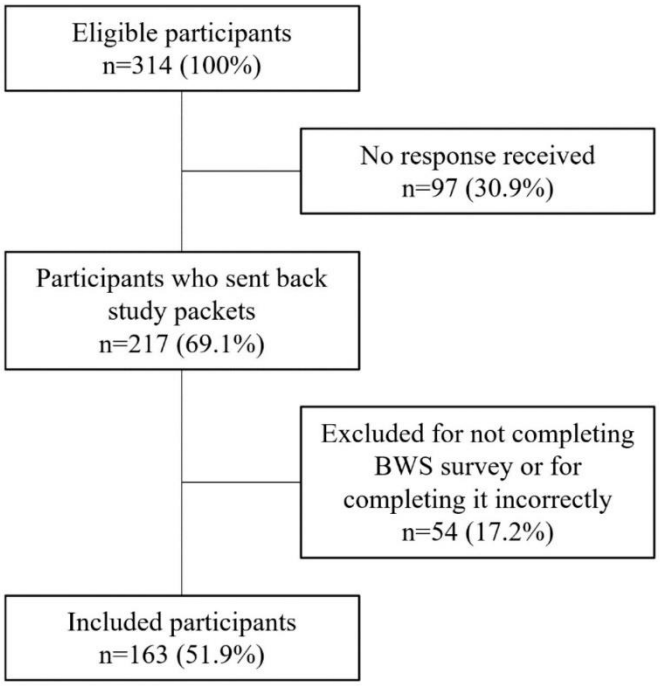


Figure 2: Flow-chart of participants eligible and included in this study.

Table 1: clinical and demographic characteristics (n=163)	
	Median (IQR) or n(%)
Age (years)	75.0 (73.0 – 81.0)
SF-SarQoL QoL (0-100 points)	46.9 (27.0 – 66.1)
Gender	
Men	56 (34.4%)
Women	107 (65.6%)
Grip strength	
Low *	49 (30.1%)
Normal	113 (69.3%)
Sarcopenia	
Yes	11 (6.7%)
No	152 (93.3%)
* <27 kg for men and <16 kg for women	

3.2 Relative importance of the 14 QOL aspects

Relative importance scores calculated for the 14 aspects of QOL included in the SF-SarQoL questionnaire are presented in table 2 and figure 3. The participants considered that "feeling a reduction in their physical capacity" [11.26 (10.37-12.14)], "having balance problems" [11.09 (9.91-12.27)], and "feeling a reduction of the strength in your legs" [9.03 (8.30-9.77)] were the 3 most important aspects of QoL in sarcopenia. On the other end of the spectrum, they considered "feeling a reduction of the strength in your arms" [4.35 (3.75-4.96)], "feeling a reduction in your muscle mass" [3.82 (3.15-4.49)], and "having difficulty carrying heavy objects" [2.89 (2.19-3.59)] as the least important aspects of QoL. Relatively large 95% confidence intervals were found, and consequently an important number of items have overlapping intervals. Roughly speaking, items can be grouped together in 3 groups: the 2 items on feeling a reduction in physical capacity and experiencing balance problems are significantly more important than all other items. Next up are 8 items whose confidence intervals overlap: leg strength, difficulty during moderate effort, feeling weak, difficulty getting up from the floor, limiting movements, fatigue during moderate effort, fatigue while walking, and walking distance. Lastly, a third group of items are clearly less important than the items mentioned so far: leisure activities, arm strength, muscle mass and carrying heavy objects. It is important to add that this is a relative assessment, rating whether one item is more important than another, not an absolute assessment, rating whether an item is important or not.

Table 2 also presents the results separated by gender and grip strength. We did not find important differences in the way men and women or people with low and normal grip strength valued the 14 QoL aspects. Only the item "limiting movement" was significantly different in terms of relative importance score between men and women [8.29 (7.10 – 9.48) vs 6.60 (5.89 – 7.31); p=0.011]. The comparison on grip strength also yielded a single significantly different RIS, in this case for the item "difficulty getting up from the floor, which was

considered more important by participants with low grip strength [9.89 (7.48 – 12.30) vs 6.94 (5.60 – 8.27); p=0.024].

The results of the additional subgroup analyses on BWS version (A versus B), SARC-F (≥4 points versus <4 points), sarcopenia status (sarcopenia versus no sarcopenia), age (≤75 years versus >75 years), and QoL (SF-SarQoL score ≤47 points versus >47 points) are available in appendix 1. In short, while we did find minor differences between the RIS values when comparing between groups on several characteristics, none of these differences upend the global results of the analysis on the complete sample.

Table 2: Relative Importance Scores – n=163

Label		Gender			Grip strength		
	All	Men	Women	p-value	Normal	Low	p-value
Reduction physical capacity	11.26 (10.37 - 12.14)	11.35 (9.75 - 12.94)	11.21 (10.13 - 12.30)	0.889	11.55 (10.44 - 12.65)	10.55 (9.01 - 12.09)	0.314
Balance problems	11.09 (9.91 - 12.27)	10.96 (8.77 - 13.15)	11.15 (9.73 - 12.58)	0.881	11.05 (9.62 - 12.48)	11.34 (9.11 - 13.56)	0.826
Reduction strength legs	9.03 (8.30 - 9.77)	9.42 (8.12 - 10.71)	8.83 (7.93 - 9.74)	0.458	8.58 (7.71 - 9.45)	10.10 (8.69 - 11.50)	0.064
Difficulty during moderate efforts	8.60 (7.88 - 9.32)	8.85 (7.68 - 10.02)	8.48 (7.55 - 9.41)	0.634	8.89 (8.00 - 9.78)	7.96 (6.66 - 9.26)	0.247
Feeling physically weak	8.06 (7.19 - 8.92)	7.08 (5.52 - 8.63)	8.57 (7.52 - 9.62)	0.108	7.84 (6.74 - 8.93)	8.50 (7.03 - 9.97)	0.494
Difficulty getting up from the floor	7.78 (6.61 - 8.96)	7.56 (5.72 - 9.4)	7.90 (6.36 - 9.44)	0.786	6.94 (5.60 - 8.27)	9.89 (7.48 - 12.30)	0.024
Limiting movement	7.18 (6.56 - 7.8)	8.29 (7.10 - 9.48)	6.60 (5.89 - 7.31)	0.011	7.54 (6.76 - 8.32)	6.43 (5.40 - 7.45)	0.108
Fatigue during moderate effort	7.09 (6.41 - 7.77)	6.67 (5.55 - 7.8)	7.31 (6.44 - 8.18)	0.382	7.40 (6.54 - 8.26)	6.41 (5.28 - 7.53)	0.190
Fatigue while walking	7.00 (6.09 - 7.92)	6.98 (5.35 - 8.61)	7.02 (5.89 - 8.14)	0.971	6.85 (5.70 - 7.99)	7.34 (5.76 - 8.93)	0.627
Reduction walking distance	6.82 (5.82 - 7.83)	7.01 (5.17 - 8.86)	6.73 (5.51 - 7.95)	0.792	6.72 (5.49 - 7.96)	7.02 (5.17 - 8.86)	0.794
Reduction leisure activities	5.02 (3.90 - 6.14)	4.16 (2.38 - 5.95)	5.47 (4.03 - 6.92)	0.275	3.86 (6.62 - 4.55)	4.19 (2.29 - 6.08)	0.391
Reduction strength arms	4.35 (3.75 - 4.96)	4.63 (3.67 - 5.59)	4.20 (3.42 - 4.99)	0.512	4.53 (3.81 - 5.26)	3.96 (2.80 - 5.13)	0.398
Reduction muscle mass	3.82 (3.15 - 4.49)	3.93 (2.65 - 5.2)	3.76 (2.97 - 4.55)	0.820	3.67 (2.86 - 4.49)	4.12 (2.88 - 5.37)	0.546
Difficulty carrying heavy objects	2.89 (2.19 - 3.59)	3.12 (1.95 - 4.29)	2.76 (1.87 - 3.65)	0.630	3.20 (2.33 - 4.07)	2.20 (0.98 - 3.43)	0.201

Average importance (100/14) is 7.14. The double line indicates the cut-point between above and below average importance for RI scores. Items are presented from most important to least important according to RIS. CI= confidence interval

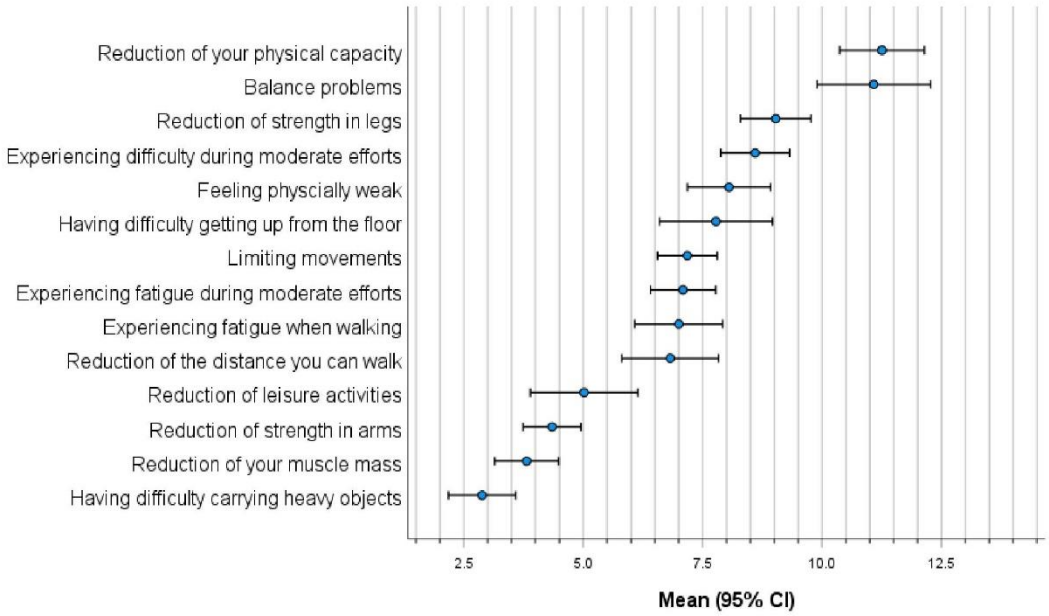


Figure 3: relative importance scores

3.3 Cluster analysis

The cluster analysis detected 2 distinct clusters within the sample. The value for the silhouette measure of cohesion and separation was 0.3, indicating that the cluster solution found is relatively weak and should be interpreted with caution. The largest cluster had 88 members, while the second cluster was slightly smaller at 75 members. Relative importance scores and rank for the 14 aspects of quality of life are presented for each cluster in table 3.

Table 3: RIS and ranking per cluster				
	Cluster 1 (n=88)		Cluster 2 (n=75)	
	RIS	Ranking	RIS	Ranking
Reduction physical capacity	8.84 (7.66 - 10.02)	4	14.1 (13.05 - 15.14)	1
Balance problems	10.04 (8.32 - 11.76)	3	12.31 (10.71 - 13.91)	2
Reduction strength legs	11.21 (10.2 - 12.22)	2	6.48 (5.73 - 7.23)	10
Difficulty during moderate efforts	7.43 (6.57 - 8.29)	5	9.98 (8.83 - 11.12)	4
Feeling physically weak	4.84 (3.98 - 5.71)	13	11.82 (10.72 - 12.93)	3
Difficulty getting up from the floor	11.54 (9.93 - 13.16)	1	3.37 (2.26 - 4.48)	11
Limiting movement	6.35 (5.63 - 7.08)	8	8.15 (7.12 - 9.18)	6
Fatigue during moderate effort	5.87 (5.10 - 6.63)	10	8.53 (7.41 - 9.65)	5
Fatigue while walking	6.64 (5.34 - 7.95)	7	7.43 (6.12 - 8.73)	7
Reduction walking distance	6.95 (5.54 - 8.35)	6	6.68 (5.19 - 8.17)	8
Reduction leisure activities	3.64 (2.22 - 5.05)	14	6.65 (4.89 - 8.40)	9
Reduction strength arms	6.26 (5.39 - 7.14)	9	2.11 (1.63 - 2.59)	12
Reduction muscle mass	5.39 (4.43 - 6.35)	11	1.97 (1.22 - 2.73)	13
Difficulty carrying heavy objects	4.98 (3.86 - 6.11)	12	0.43 (0.22 - 0.63)	14

Overall, cluster 1 found items related to falls (i.e., getting up from the floor, leg strength and balance) to be the most important and cluster 2 prioritized feeling physically capable. Both clusters shared the item "balance problems" in their top 3 of most important items, and "carrying heavy objects" as 1 of the 3 least important items.

4. Discussion

This study suggests that older people do not consider all items of musculoskeletal QoL represented in the BWS survey to be equally important. The ranking established in this study showed the QoL aspects "reduction of your physical capacity" and "experiencing balance problems" to be the most important. Within the sample described in this article, two different profiles were found with regards to the importance placed on certain aspects of musculoskeletal QoL. While the silhouette measure indicated that the structure found was weak, and that it could be artificial, it is not hard to imagine that there are likely different groups with different sets of priorities with regards to QoL. While we would caution against over-interpreting these results based on this sample alone, the choices made within the 2 clusters seem to make sense in that they coalesce around two themes: falls and physical capacity. The first one, falls, had already been identified in a previous study using focus groups, but the second one, physical capacity, had not yet been put forward [12].

To our knowledge, this is the first study to demonstrate the relative importance of different aspects of QoL in a quantative manner in sarcopenia. Unfortunately, because of the highly specific nature of the SF-SarQoL, and its focus on musculoskeletal aspects of QoL that are relevant to sarcopenic patients, we are unable to directly compare our findings with other studies, because of the heterogeneity of the items studied under the umbrella of the concept of QoL. There are however a limited number of studies which have investigated how older people think about QoL and what aspects they consider to be more or less important, employing broader concepts of QoL than used in our own BWS survey.

A thematic synthesis by Van Leeuwen and colleagues compiled a number of qualitative studies on the subject and is the most thorough overview of what QoL means to older people. The authors included 48 studies, incorporating the perspectives of more than 3400 older community-dwelling people from Western countries. From this vast amount of information, they distilled nine QoL domains: health perception, autonomy, role and activity, relationships, attitude and adaptation, emotional comfort, spirituality, home and neighborhood, and financial security. They also stress the interconnections between domains and the ripple effect of changes in a particular domain on the other domains. This exhaustive synthesis however was not set up to indicate which aspects or domains of QoL are the most important, or to establish a hierarchy among the nine domains, favoring instead the broadest possible concept of QoL [22].

In terms of quantitative research, there are three studies that have surveyed the relative importance of different aspects of QoL in the specific population of older people. Molzahn and colleagues published the results of a secondary analysis of the WHOQOL-OLD pilot study in 2011. In this article, they present data collected from 7401 people aged 60 years or older from 22 countries on the importance of 31 facets of QoL. The participants in this study considered ADL, general health, sensory abilities, mobility, autonomy, and

energy to be the most important QoL facets, in the order presented. With regards to the least important facets, they singled out sex-life, opportunity to learn new skills, social participation, and a positive body image and appearance [23]. While the items in the Molzahn study and our own survey are too dissimilar to compare head-to-head, it is interesting to note that the concepts considered important to the older people in the Molzahn study, such as ADL, general health, mobility, and autonomy, are well represented in the SF-SarQoL, while the concepts considered less important are not represented. A second study, carried out by Ratcliffe and colleagues and published in 2017, recruited 500 younger people (18-64 years) and 500 older people (65+ years) who performed two preference elicitation experiments (ranking and successive BWS) aimed at establishing a hierarchy of 12 quality of life dimensions. The older sample found the dimensions independence, physical mobility, control, and mental health particularly important in the ranking experiment, with similar results for the BWS task. While the items in this study are again too dissimilar to our own BWS survey, we note the importance that the participants of our study placed on their physical capacity, balance, and strength in the legs, and hypothesize that these items may be considered as prerequisites for independence and physical mobility, considered important in the Ratcliffe study. This study also demonstrated that the preferences of younger and older people with regards to QoL are different [24]. Lastly, Uy and colleagues conducted a BWS experiment in Singapore of which they published the result in 2018. They sought to establish a ranking of 27 health-related QoL domains and recruited 603 participants aged between 21 and 88 years old to do this. The BWS results placed the domain "self-care" at the top of the hierarchy, followed by "healing and resistance to illness" and "social relationships". At the other end of the scale, the participants considered "having a satisfactory sex life" as the least important aspect of HRQoL, followed by "having a normal physical appearance", and "interacting with others" [25]. However, because of the earlier finding that QoL preferences are different between older and younger participants, these results should be interpreted with caution.

As with any study, there are some methodological and practical limitations that need to be addressed. A first limitation is that, because of the recruitment and administration methods of this study, there is the potential for non-response bias. A total of 314 potential participants were contacted, and we received responses from 217 of them, a 69% response rate. Out of those 217, a further 54 participants either did not complete the BWS survey at all, or failed to complete the survey correctly (e.g., multiple "best" choices for a single choice task). This means that we were able to include 52% of the people we contacted, and 75% of the people who participated, in the final analyses. When we compared the 163 participants included in this analysis with the 54 that responded but were excluded, we did not find a significant difference for age (p=0.300), gender (p=0.183) and probable sarcopenia/low grip strength (p=0.155). We did however find that a larger proportion of the sarcopenic participants in the sample were unable or unwilling to complete the choice tasks, compared to the non-sarcopenic participants (52% completion rate versus 78%; p=0.011).

The 54 excluded participants also had significantly lower QoL [33.33 (18.27-44.55) vs. 45.99 (27.65-65.38) points on a scale from 0 to 100 measured with the SF-SarQoL; p=0.001] compared to the 163 included participants. This phenomenon may be related to the relative burden of the choice task, which may have been perceived as greater by sarcopenic participants and by those that already had substantially reduced quality of life. A second limitation is the sample size itself. Although there are currently no guidelines for minimum sample size for BWS surveys available in the literature, a review from 2016 found a median sample size of 175 participants (range: 15 to 803) for 26 object case BWS studies, in line with our own sample of 163 participants [10]. However, the relatively large confidence intervals found for the relative importance scores, which prohibit us from clearly separating some items, would likely have been narrower with a greater sample. This is especially noticeable for the items ranked at the middle of the importance hierarchy, where there are 8 items with overlapping confidence intervals. A third limitation of this study is that it was conducted in a single setting, namely older, community-dwelling volunteers from the Liège province in Belgium. Without further data it is uncertain whether our results can be generalized to the wider population of older people in Belgium or whether the results of this study are transferable to other countries.

This study could however open up some perspectives for the future. The ranking established could assist in a more detailed analysis of QoL data obtained with the SF-SarQoL, either by an item-based analysis taking into account the relative ranking of the item in question, or by creating a preference-weighted overall QoL score for the SF-SarQoL. It could also inform specific targets for improvement in interventional studies or inspire the design of interventions so as to increase the effect on physical capacity, balance, and leg strength.

In conclusion, this study provides the first data on the relative importance of different aspects of QoL in the context of sarcopenia from the subjective perspective of the patient. We established a ranking of 14 aspects of QoL on importance and showed that there were two clusters present in the sample with different priorities with regards to QoL.

5. Bibliography

[1] Beaudart C, Zaaria M, Pasleau F, Reginster J-Y, Bruyère O. Health outcomes of sarcopenia: a systematic seview and meta-analysis. PLoS One 2017;12:e0169548.

[2] Bruyère O, Beaudart C, Ethgen O, Reginster JY, Locquet M. The health economics burden of sarcopenia: a systematic review. Maturitas 2019;119:61–9. https://doi.org/10.1016/j.maturitas.2018.11.003.

[3] Woo T, Yu S, Visvanathan R. Systematic literature review on the relationship between biomarkers of sarcopenia and quality of life in older people. J Frailty Aging 2016;5:88–99. https://doi.org/10.14283/jfa.2016.93.

[4] Reginster JY, Beaudart C, Al-Daghri N, Avouac B, Bauer J, Bere N, et al. Update on the ESCEO recommendation for the conduct of clinical trials for drugs aiming at the treatment of sarcopenia in older adults. Aging Clin Exp Res 2020. https://doi.org/10.1007/s40520-020-01663-4.

[5] Food and Drug Administration. Value and Use of Patient-Reported Outcomes (PROs) in Assessing Effects of Medical Devices. 2017.

[6] Beaudart C, Biver E, Reginster J-YJ-Y, Rizzoli R, Rolland Y, Bautmans I, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. Age Ageing 2015;44:960–6. https://doi.org/10.1093/ageing/afv133.

[7] Beaudart C, Biver E, Reginster JY, Rizzoli R, Rolland Y, Bautmans I, et al. Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. J Cachexia Sarcopenia Muscle 2017;8:238–44. https://doi.org/10.1002/jcsm.12149.

[8] Geerinck A, Beaudart C, Reginster J-Y, Locquet M, Monseur C, Gillain S, et al. Development and validation of a short version of the Sarcopenia Quality of Life questionnaire : the SF - SarQoL. Qual Life Res 2021. https://doi.org/10.1007/s11136-021-02823-3.

[9] Hiligsmann M, Pinto D, Dennison E, Al-Daghri N, Beaudart C, Branco J, et al. Patients' preferences for osteoarthritis treatment: the value of stated-preference studies. Aging Clin Exp Res 2019;31:1–3. https://doi.org/10.1007/s40520-018-1098-3.

[10] Cheung KL, Wijnen B, Hollin I, Janssen E, Bridges J, Evers S, et al. Using Best – Worst Scaling to Investigate Preferences in Health Care. Pharmacoeconomics 2016;34:1195–209. https://doi.org/10.1007/s40273-016-0429-5.

[11] Hiligsmann M, Beaudart C, Bruyère O, Biver E, Bauer J, Cruz-Jentoft AJ, et al. Outcome Priorities for Older Persons With Sarcopenia. J Am Med Dir Assoc 2020;21:267-271.e2. https://doi.org/10.1016/j.jamda.2019.08.026.

[12] Beaudart C, Bruyère O, Cruz-Jentoft AJ, Vaquero-Pinto MN, Locquet M, Bauer J, et al. Patient's Engagement in the Identification of Critical Outcomes in Sarcopenia. J Am Med Dir Assoc 2020;21:284–6. https://doi.org/10.1016/j.jamda.2019.09.004.

[13] Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: What it can do for health care research and how to do it. J Health Econ 2007;26:171–89. https://doi.org/10.1016/j.jhealeco.2006.04.002.

[14] Beaudart C, Reginster JY, Petermans J, Gillain S, Quabron A, Locquet M, et al. Quality of life and physical components linked to sarcopenia: The SarcoPhAge study. Exp Gerontol 2015;69:103–10.

[15] Louvière JJ, Flynn TN, Marley AAJ. Best-worst scaling: theory, methods and applications. Cambridge University Press; 2015.

[16] Sawtooth Software. Lighthouse Studio Manual n.d. https://www.sawtoothsoftware.com/help/lighthouse-studio/manual/index.html?hid_web_maxdiff_numberofsets.html (accessed November 9, 2020).

[17] Cruz-Jentoft AJ, Bahat G, Bauer J, Boirie Y, Bruyère O, Cederholm T, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019;48:16–31. https://doi.org/10.1093/ageing/afy169.

[18] Malmstrom TK, Miller DK, Simonsick EM, Ferrucci L, Morley JE. SARC-F: A symptom score to predict persons with sarcopenia at risk for poor functional outcomes. J Cachexia Sarcopenia Muscle 2016;7:28–36. https://doi.org/10.1002/jcsm.12048.

[19] Sawtooth Software. Identifying bad respondents: fit statistic and identifying random responders n.d. https://sawtoothsoftware.com/help/lighthouse-studio/manual/hid_web_maxdiff_badrespondents.html.

[20] Norusis MJ. IBM SPSS Statistics 19 Statistical Procedures Companion: Cluster Analysis, Chapter 17 2011:361–91. https://doi.org/10.4135/9781412983648.

[21] Kaufman L, Rousseeuw P. Finding groups in data: an introduction to cluster analysis. Wiley-Interscience; 1990.

[22] Van Leeuwen KM, Van Loon MS, Van Nes FA, Bosmans JE, De Veti HCW, Ket JCF, et al. What

does quality of life mean to older adults? A thematic synthesis. vol. 14. 2019. https://doi.org/10.1371/journal.pone.0213263.

[23] Molzahn AE, Kalfoss M, Makaroff KS, Skevington SM. Comparing the importance of different aspects of quality of life to older adults across diverse cultures. Age Ageing 2011;40:192–9. https://doi.org/10.1093/ageing/afq156.

[24] Ratcliffe J, Lancsar E, Flint T, Kaambwa B, Walker R, Lewin G, et al. Does one size fit all? Assessing the preferences of older and younger people for attributes of quality of life. Qual Life Res 2017;26:299–309. https://doi.org/10.1007/s11136-016-1391-6.

[25] Uy EJ, Bautista DC, Xin X, Cheung YB, Thio ST, Thumboo J. Using best-worst scaling choice experiments to elicit the most important domains of health for health-related quality of life in Singapore. PLoS One 2018;13:1–18. https://doi.org/10.1371/journal.pone.0189687.