Chapter

Reliability and Validity of Clinicopathological Features Associated with Frailty Syndrome in Elderly Population

Kelvin Leshabari

Abstract

Geriatrics is an applied science as its practice is an art of medicine. As a scientific discipline, there exists a potential race for measurements. Frailty stands as among poorly defined concepts in geriatric medicine. There are philosophical, circumstantial, and practical justifications behind this rather seemingly *clinical tragedy*. This chapter contributes toward reliability and validity aspects of currently applied frailty scales and indicators across different population base. It acknowledges the contribution of Fried's frailty scale. It also describes different frailty scales and indicators tested in America, Europe, and Asia. Lastly, the chapter contrasts the popular belief behind applications of Cronbach's α coefficient of test scores for reliability assessment in clinical research. Other research gaps are also highlighted including merging clinical research findings in geriatrics with psychosocial aspects under the emerging field of geropsychology. It also proposes a solution for usage in future studies that aim at assessing reliability of test scores in clinical and biomedical sciences.

Keywords: frailty, reliability, validity, multimorbidity, index, scale

1. Introduction

Geriatric medicine is a relatively younger sub-specialty of medicine. Unlike fields like general internal medicine or surgery, that are known to have existed since antiquity, geriatrics has gained significant popularity, in orthodox medical practice, around the second half of the twentieth century. Geriatrics is *an applied science* just as its practice is *an art* of medicine. For that matter, there exists a potential race for measurements. It follows logic therefore that for geriatric medical conditions and practice to be acceptable among scientists, it needs unified codes that are measurable. This chapter will provide basic aspects associated with measuring variables that are customarily prevalent in geriatric wards and corridors throughout the world. Specifically, it analyzes the reliability and validity of different scales of the commonest concept of *frailty* among senior citizens the world over.

The clinical characterization, of modern geriatric medicine, owes much to the pioneering work, of Professor Bernard Isaacs back in the 1960s. It was Isaacs who is credited in public literature, to have coined the term "*geriatric giants*," in common usage, to geriatricians the world over to date [1]. Simply stated, he referred "geriatric

giants" to conditions of *immobility*, *instability*, *incontinence*, and *impaired memory/intellect* that are relatively common, on statistical grounds, among senior citizens of any human society [1]. From 1960s onward, the conditions characterizing *geriatric giants* have changed several times, and currently used mainly among scholars and clinicians alike, in its modified form, as per Professor Mary Tinetti's keynote address of *Geriatric 5Ms,* at the Canadian Geriatric Society conference, back in April 2017 [2]. She addressed the Geriatric 5Ms to comprise the *Mobility*, *Mind*, *Medications*, *Multi-complexity*, and *Matters* most [2]. It must be understood that most of this characterization refers to measurable constructs that defines core aspects of geriatric medical research and practice, initially coined as *geriatric giants* by Isaacs in 1965.

There exists a lot of confusion in geriatric medicine to date, regarding the measurable construct of *geriatric giants* [3–7]. Part of this confusion has basis from failure to achieve standard definitions from its components. This is because, in almost all cases, disease conditions among senior citizens, unlike other groups in the population pyramid, tend to present *atypically* on clinical grounds. Whereas the exact cause of this trend among senior citizens is still ill understood, there exists evidence for a common pathway, originating from different organ systems, in etiopathogenesis of diseases in the elderly. Moreover, quite often there is also a disconnect between the original site of *malice* and the clinical presentations for symptoms and signs of pathologic conditions thereafter. The immediate effect of this rather anomalous conundrum is the rather *bizarre presentations* of most common clinical conditions, seen among senior citizens, as well as recorded in their morbidity and mortality statistics throughout the world.

Atypical presentation in the elderly can be exemplified, say by a Nonagenarian lady, presenting to the emergency department of a typical hospital, with symptoms and signs suggestive of acute confusional state like delirium, caused by *Escherichia coli* infection in her urinary tract. As it is commonly the case, once her bladder cystitis/urethritis is treated, using relatively simple treatment pathways, the acute confusional state disappears. The observation given, justifies not only the atypical nature of presentation, to most geriatric illnesses, but also the multiple organo-systemic involvement, in their pathogenesis. Besides, growing characterization of emotional, social, and cognitive aspects of aging is paramount in modern clinical practice. There are several theories and postulates that endeavor to link the interplay of the mind (and the central nervous system) in multisystemic etiopathogenesis of Geriatric 5Ms. The theories and postulates have evolved into a new sub-specialty named geropsychology. The details of which will be discussed further in the discussion section. However, to the betterment of science, there exists palpable evidence that probably the so-called *Geriatric 5Ms* has achieved a unified goal of standardizing the measurable construct of *geriatric giants*. It is on this basis that this chapter finds its pivot, on the attributable last aspect of *Matters most*, referred to as *Frailty*.

Frailty is a poorly defined syndrome almost exclusively confined to the elderly population. There are dozens of descriptions given for frailty [8–15]. All of them were made for specified frameworks of interest by their original authors. On pedagogic sense, none could be used systematically, without a pinch of doubt, to any destitute clinician/researcher. However, out of dozens of frailty definitions available, the one proposed by Fried and colleagues in Cardiovascular Health Study Collaborative Research Group back in 2001 [13] is the most widely applied framework by clinicians/researchers in bio-gerontologists the world over. The underlying scientific framework for applying frailty syndrome to be discussed in this chapter has taken into account the famous fact derived by Sir George Box's seminal paper back in 1978 that *all models are wrong but some are useful* [16] in its philosophical sense. It is neither the intention of this chapter nor anywhere in the mind of its author to market the Cardiovascular Health Study Collaborative Research Group's

frailty postulate. Rather, and I do believe it to be safe, the plan is to sparingly appreciate strengths and weaknesses of some of these useful concepts, via estimation of their internal consistency and content validity in each of them. Thus, this chapter will guide the reader through reliability and validity aspects of clinicopathological features associated with frailty syndrome in senior citizens.

2. Reliability of frailty models in elderly population studies

Geriatrics, as a branch of clinical sciences, is a scientific discipline as physics or chemistry is to natural sciences. To that end, measurements are the core aspect for its sustainability. It is under this framework that most frailty scales are available the world over to date, and in future shall be assessed. Technically, assessing the quality aspects of the scale takes mainly two domains, namely, *reliability* and *validity*. Taken simply, reliability assessment refers to the process of determining the extent to which a measurement of a phenomenon provides stable and consistent results [17]. This definition, though adopted from Carmine and Zeller's publication of 1979 [17], who worked in the field of psychometrics, is as applicable, in its entirety to frailty scale assessment, as it was intended in assessing psychometric scales. Thus, whereas the intention here is to adopt quantifiable and logically consistent manner for reliability assessment of frailty scales, it is by no means the intention of the author to *sales pitch* the methods discussed nor should it be conclusive that the method described in this chapter is the only mechanism of achieving reliability assessment. To ensure clarity in this endeavor, the end of the chapter will also contain some vivid shortcomings and a potential solution to the reliability assessment method described in this chapter.

2.1 Internal consistency assessment of frailty scales used in elderly population

There are various ways of assessing reliability index of any given phenomenon/scale in nature. Some are well known in literature, and there are probably many others in production pipelines for usage in future. However, the most popular methods include test-retest reliability, split-half reliability, and internal consistency reliability tests. Out of these, this chapter will deal with internal consistency reliability. The decision to do so is derived from its conceptual meaning as opposed to the rest. Simply stated, internal consistency refers to the extent to which a measurement of a scale provides stable and consistent results across a specified condition [17]. One rule is important to be mentioned here, in that all accounts of assessing reliability of any given scale, the reliability score to be obtained is not reflective of a constancy but rather a mere statistic for a given test result. This translates to the fact that a given scale may end up with different scores, under different elderly population conditions, dependent on a number of factors, some known (e.g., test settings and gender) and others unknown even to the test itself. Thus, caution to the interpretation of the test scores is highly warranted.

2.1.1 Clinical Frailty Scale

Clinical Frailty Scale (CFS) is a clinical judgment-based tool (originally designed as an epidemiological tool) to screen for frailty and other adverse health events in opposition to fitness in older aged population. It is a direct replica of a frailty index that was part of the original design aspect of the first part of Canadian Study on Health and Aging (CSAH), with the aim of characterizing cognitive impairment and other important health issues, designed as a prospective 5-year follow-up of 10,263 people aged at least 65 years back in 1991 [18, 19]. At the time of going to

press, the Clinical Frailty Scale is composed of a 9-point scale, that was made public in 2007, an improvement from the original scale of a 5-point scale originally published in 2005 [20]. It was originally developed in the second half of the Canadian Study of Health and Aging (CSHA) as a quick means to assess frailty and other senile physical and mental challenges past clinical assessment [20]. The conceptual framework of the Clinical Frailty Scale relies on the "fitness and frailty" model, and the scale was designed by adopting the mechanism from Streiner and Norman [21]. It is for all practical purposes, not a questionnaire but a quantified summary write-up of an elderly overall health status in relation to mortality risks. Internal consistency scores for Clinical Frailty Scale among elderly population across different geographical areas are provided in **Table 1**.

2.1.2 Edmonton Frail Scale (EFS)

Edmonton Frail Scale, an effort first conceptualized by Darryl Rolfson while at the University of Alberta, Canada back in 1999, was presented for the first time to peer review at the Canadian Geriatric Society in Edmonton, Canada, in 2000 [25]. Ever since its first time in press, the scale has been applied in research, educational and clinical settings for quantitative frailty assessment among senior citizens [15, 26–33]. Edmonton Frail Scale consists of nine domains and 11 items. The initial scale devised by Rolfson at Edmonton had 10 domains [25]. Each component may have a score of 0, 1, or 2 signifying normal health, mild/moderate impairment, or severe impairment, respectively. Domains include general health status; cognitive status; medication use; presence of social support; incontinence; nutrition and mood; functional dependency; and functional performance test [25]. The total scores are also classified into no frailty (0–3 points); pre-frailty (4–5 points); frailty (6–8 points), and severe frailty (9–17) [26]. The internal consistency scores of Edmonton Frail Scale for senior citizens across different geographical settings are as reported here in **Table 2**.

2.1.3 Groningen Frailty Indicator

Groningen Frailty Indicator (GFI) is a 15-item indicator for assessment of frailty developed by Professor Steverink and his colleagues at the University of Groningen, The Netherlands, first published in 2001 [34]. The internal consistency findings of GFI are as summarized in **Table 3**.

2.1.4 Tilburg Frailty Indicator

Tilburg Frailty Indicator (TFI) is a questionnaire for screening frail community dwelling older people that includes self-reported information, originally tested and validated from an elderly community of Roosendaal in The Netherlands, based on

Country/region	Cronbach's α-reliability score (95% C.I.)	Settings
1. Australia [22]	0.76 (0.7–0.81)	Perioperative (hospital-based)
2. Canada [20]	0.97 (not given)	Community-based
3. Singapore [23]	0.91 (0.86–0.95)	Hospital-based
4. Turkey [24]	0.811 (not given)*	Outpatient clinic (hospital-based)

**p-value cited as <0.001.*

Table 1.
Internal consistency scores for Clinical Frailty Scale across elderly population from different geographical areas.

Country/region	Cronbach's α-reliability score (95% C.I.)	Settings
1. Canada [25]	0.62	Hospital/clinic based
2. Ireland [31]	0.41	Hospital based
3. Italy [26]	0.98	Hospital based
4. Poland [28]	0.709	Hospital based
5. China [33]	0.95	Hospital based
6. Canada [20]	0.97	Hospital based

Table 2.
Internal consistency scores for Edmonton Frail Scale across senior citizens from different geographical areas.

Country/region	Construct validity index	Settings
1. The Netherlands [35]	0–1: disorder (median, range): 2, 1–4 2: (median, range): 4, 3–6 3: (median, range): 6, 4–8	Community-based
2. The Netherlands [36]	GFI ≥ 4	Community-based
3. Romania [37]	GFI score = 0.746	Physician-based

Table 3.
Construct validity scores for Groningen Frailty Indicator across senior citizens in different geographical areas.

Country/region	Construct validity index	Settings
1. China [39]	Physical domain: r = −0.39–0.57 (P < 0.001) Psychological domain: r = −0.47–0.49 (P < 0.001)	Community-based
2. The Netherlands [38]	Social domain: r = −0.35–0.71 (P < 0.001) Physical domain: r = −0.43–0.62 (P < 0.001) Psychological domain: r = −0.19–0.46 (P < 0.001) Social domain: r = 0.29–0.96 (P < 0.001) Physical domain: r = 0.31 (P < 0.001) Psychological domain: r = 0.24 (P < 0.001)	Community-based
3. Italy [40]	Social domain: r = 0.25 (P < 0.001)	Community-based

Table 4.
Construct validity scores of Tilburg Frailty Indicator across senior citizens in different geographical areas.

a working framework in development, developed by a team of Dutch scientist first published in 2009 [38]. Tilburg Frailty Indicator is unique among frailty indicators, in that it includes multiple domains of human functions but selectively excludes disability [38]. TFI consists of two parts, namely, multimorbidity and frailty domains. The first part (designated as part A) contains 10 questions on determinants of frailty in relation to disease states, while the second part is solely on frailty aspects [38]. The internal consistency score ratings of TFI across studies from different geographical areas are given in **Table 4**.

3. Validity aspects of frailty scales and indicators used in elderly population

Much as reliability may be loosely assumed to be synonymous to precision in measurements, it follows a natural pattern then to ensure validity by the assumption

of accuracy. It must be understood that geriatrics, just as other branches of clinical medicine, is essentially an applied science field. To this end, the reader is cautioned against making substantial error in reasoning, that of assuming measurement exactness of constructs made in its clinical measurements, just as natural scientists make, in say reaction time in subjects like physics or chemistry. It is on this basis, that all aspects of validity, discussed in this chapter, constitute a number of assumptions, some of them may be hard to prove, even when considered useful in the stated models. For instance, since most validation processes in constructing frailty scales and indicators consisted of a number of items, the assumptions made are such that those items, when taken collectively, refer to a construct of frailty, and that when applied to humans in their contextual nature, can distinguish those who are frail from those who are not. This section will deal with one important form of validity measurement, that of construct validity at most [41].

Construct validity is a way of measuring a disposition/character/trait/belief such that its accuracy can be estimated with quantifiable degrees of confidence. In simplistic fashion, it is a way of measuring a test for what it claims to quantify. Construct validity differs from other forms of validity in applied sciences, namely, criterion and content validity, since in construct validation, there is an aspect of quantifying the quality of a measuring instrument toward what it claims to measure. Thus, for all practical purposes, this chapter will endeavor to quantify aspects hypothesized to assess frailty, as applied to the community of senior citizens living in different geographical communities. This notion, inter alia, follows the appreciable level of acceptance in reliability indices prior to its undertaking, lest of that, it may be deemed invalid in practice. In this sub-section, an analysis of different frailty scales/indicators in construct validity will be determined here underneath.

4. The triumph and controversies surrounding reliability and validity of frailty scales/indicators in elderly population

It is important to underscore the importance of association between what is characterized as *frailty* and increased susceptibility to ill health among humans of advanced age. From the earlier sections of this chapter, an account of the term *frailty* has been made. However, it is important to analyze those numbers, in a bid to express not only what they suggest but also not to overexpress their usefulness in science. One important caution needs special attention here in that myself as a clinical researcher, positively influenced by biased affection to make judgment using numbers and experimental findings, maybe at risk of committing a *self-fulfilling prophecy*, a typical form of rather known *pygamalion effect,* quite common among scientists dealing with quantitative research methods and applications. Therefore, this section will be dealt with not only interpretation of reliability and validity of different frailty scales/indicators shown before but also the challenges of assuming the score results to individuals as *sin qua non* in frailty assessment to prospective readers be it practicing geriatricians, bio-gerontologists and/or policy makers, and other decision makers in aging field.

First, on reliability aspects, it is important for geriatricians, other clinicians handling senior citizens, clinician-scientists, policy makers as well as other readers alike to be aware of the fact that frailty scales/indicators scores derived from cited studies above do not in actual sense measure reliability at best. There is no doubt that no other statistic in published literature has been a subject of wide confusion than *coefficient* α for reliability test scores. Specifically, *Cronbach's* α coefficient that at best displays homogeneity of test scores has been incorrectly associated with a quality indicator of internal stability score, a direct reflection of a reliability

estimate. This rather subtle cognitive error has been in existence in science for at least 60 years. It was formerly described by Cronbach via a seminal paper published in 1951 [42]. To simply describe the extent of the spread of the flaw, as well as the confusion therein, until the time this line you read was first typed by the author, at the peak of a COVID19 pandemic, the Cronbach's paper published in Psychometrika back in 1951, had been cited more than 45,000 times in published literature worldwide. Details about the flaws (as well as the resulting confusion) of *Cronbach's* α as an index measure of reliability of test scores are beyond the scope of this book. However, just to give a glimpse to readers, I have decided to provide a narrative account of the fallacy behind usage of Cronbach's α coefficient as a measure of reliability of test results.

Cronbach's α coefficient as a test statistic in principle is consistently and incorrectly taken as a measure of internal stability of test scores, and therefore an estimate of internal consistency. It has been shown that Cronbach's α coefficient cannot provide investigators that sort of information [43]. Cronbach's α coefficient is at best the *greatest lower bound* to reliability estimate, and therefore almost always an underestimate of a reliability coefficient α for internal consistency of test scores [43]. At this juncture, it is important to remind readers on what exactly is internal consistency of test scores results. Simply written, internal consistency of test score results refer to interrelatedness of a set of items, be it test scores results or any other of non-singular matrix scores [44]. It therefore follows that much of the confusion surrounding Cronbach's α coefficient dates back to Cronbach's paper of 1951 [42] in that Cronbach used internal consistency and homogeneity synonymously [44]. It is clear nowadays, therefore, that Cronbach's α coefficient may attain values that are outside the scope of possible reliability scores from a single test result. I would like just to mention a solution to this challenge, just sparingly to include standard measurement errors in the form of the following equation:

$$\sigma_y = \sigma_x \left(1 - \rho_x + x_{,+}\right)^{1/2} \tag{1}$$

where $\rho_{x+}x_{,+}$—test score reliability in a population, σ_x—standard deviation of a population of interest, and σy—standard measurement error of a sample of interest.

It is important to remind readers that application of standard measurement error as a measure of internal consistency of test scores assumes each individual score results originated from a test with the same accuracy [43]. Details of this method of assessing internal consistency, and therefore inherent reliability of any given test scores, are given in other published findings of the past [43–47].

In this chapter, I have hesitated myself from committing a rather common statistical crime. It is well known that meta-analysis of findings from individual studies, customarily using forest plots, is an efficacious way of deriving effect size as well as identifying small and insignificant statistical results. However, I must admit there have been strong attempts to pool reliability and validity estimates from different studies here. The decision at the end, of not to include forest plots, from meta-analysis in this chapter, is based on the same philosophy, behind the chapter, namely, *reliability and validity* for test scores of sample estimates. On a frank note, there is profound heterogeneity reported from the study sample used for assessment of reliability and validity of frailty scales and indicators in publication database. This made all attempts toward "*forest plotting*" a futile exercise on philosophical grounds. It is quite obvious that given wealthy of statistical tests available to date, there were remedial measures to account for heterogeneity of those referred studies. However, given the fact that data were different from how they were conceptualized, and

not only in the way they were analyzed, made all those statistical tests available for estimating heterogeneity, a *non-starter* in this endeavor.

At large, these studies differ significantly on the basis of their designs. For instance, whereas findings in **Table 1** reflected assessment of Cronbach's alpha coefficient, for what was referred to as internal consistency, out of studies targeting Clinical Frailty Scale, the study by Rockwood and colleagues in Canada was conceived as a prospective observational study [20]. Moreover, Chong and colleagues' study conducted in Singapore was designed in a retrospective fashion [23]. It therefore comes out automatic that *total population at risk* was a distinguishing feature between these two studies. Clearly, with prospective data, one can quantify *population at risk*, whereas in retrospective data, such a count is not possible. The difference in probability counts of risk between those two studies does not end in risk estimates. It blows out in any calculation involving probabilistic appointments, including the early stages of obtaining Cronbach's α coefficient. Pooling out estimates from these two study designs (prospective vs. retrospective) is no difference from mixing oranges and mangoes together. Whereas the idea may seem useful in gastronomy, it is *a statistical crime*, equivalent to a third-degree murder in jurisprudence [48]. Details about the flaws in pooling estimates of retrospective and prospective designs together are described in length in mathematical statistical literature [49–52].

Apart from the design differences between studies whose estimates were pooled as means to assess reliability and validity in this chapter, heterogeneity is also suspected to be present from publication bias. Quite commonly in biomedical research and databases, studies are only published if they attain positive outcomes as per research questions designed by investigators. Whereas the message here is not to support the idea, as I personally believe in learning from findings with negative results from their hypothesized questions, I found it an important message to remind readers. It is quite possible that there were other studies left behind simply because they either failed to appear in press for what so ever reasons or they were left behind merely out of ignorance by the author during retrieval of information used to pool these data. At this point, it should be clear that there are quantitative mechanisms of assessing heterogeneity in statistical data [39, 40, 52–61]. However, those techniques are far behind abilities to correct what went wrong during design stage. It was therefore futile to justify application of those techniques to data that was conceived in either retrospective fashion or out of publication bias.

On a positive note, however, the findings from these studies do probably highlight an important construct that is related to diminished ability of various body systems, currently coined as *frailty*. This is because in most of these studies, all of their domains (physical, psychological, or social) do reflect some deficiencies that are commonly associated with those of advanced age, who we may safely assume, to reflect a true concept of frailty. Until now, it must be born to the minds of readers that the lack of gold standard decision rule for assessing frailty forces scholars to make comparisons to available tools. It is therefore a call to action for future researchers in aging research to consider design and development of more innovative concepts and tools in assessment of frailty [62–64].

Lastly, and as a matter of urgent priority, geriatricians, aging research scientists as well as other practitioners and decision makers in health need to consider different population base in their future research on frailty. At present, there appears to be palpable evidence that demographic transition has started, and likely to mature soon, in parts of sub-Saharan Africa [65]. For instance, it is quite evident that Tanzania, just like other sub-Saharan African countries, has its population undergoing *demographic transition* [65], perhaps at a faster rate than what was

seen in Europe in the nineteenth century and early parts of twentieth century. It is clear that part of what may be termed as *residual effects*, in ascertaining factors associated with frailty and other aging-related concepts, to be better explained by environmental milieu found in sub-Saharan Africa, rather than the developed North. It is therefore a matter of intellectual maturity that future studies on quality indicators in frailty assessment will also be tested and validated in population found in the South.

Likewise, on a pioneering scale, global efforts in the interplay of *mind* and *organosystemic degeneration*s in later life need a critical eye among aging researchers. At present, there is a lot of confusion not only among geropsychologists but even among clinician-scientists caring for the senior citizens across nations. For instance, there is a clear gap in research evidence on inability to characterize the cognitive domain in the illustration of the concept of frailty, in addition to clear discrepancies in how best to handle cognitive abnormalities in the oldest old. It also follows the logic that the current interventions and strategies in psychosocial interventions the world over to be *porous* at best and segregate the senior citizens at worst. To this end, I propose that psychosocial challenges arising directly or indirectly from the aging process to be handled using data-based findings. Moreover, there is a desperate call for inductive research to deductive thinking in the science of geropsychology the world over. Short of that, most scales/indicators of frailty will have a *lack-of-fit* on the basis of their missing domain of the psyche.

Acknowledgements

I wish to convey my sincere vote of thanks to my fellow trustees of **Ultimate Family Health Care**, namely, *Drs Godfrey Swai* and *Mathew Mwanjali*, whose inspiration and courage during the chapter preparation, notwithstanding their tolerance and experts' opinion, to review the chapter manuscript, is highly appreciated. Special thanks goes to these two personalities for their inception of the concept of "*family health*" in Tanzania, including their time-to-time "*free mentorship*," both in clinical practice and clinical research works, I have been making with them, out of which my current practice in Tanzania follows and grows. It is from this group interaction, the Dar es Salaam Longitudinal Ageing Study has been conceptualized.

I also wholeheartedly thank all patients and other clients in geriatric/endocrinology clinics at Moyo Safi Hospital, Alshifa Medical and Dialysis Centre as well as AB hospitals in Dar es Salaam, Tanzania, for their openness and very welcoming ground to work and appreciate my geriatric practice. Special vote of thanks goes to Dr Chuor Garang de Alier for his tireless efforts in tracing some of the full text documents used as references in this chapter. His helping hand, out of his tight schedule, at times via midnight calls, as a student at St. Hugh's College in Oxford, aiming at ensuring all references cited, have been read in context, is highly appreciated.

Conflict of interest

No conflict of interest declared in the preparation of this manuscript.

Author details

Kelvin Leshabari[1,2]

1 Ultimate Family Health Care Trust, Dar es Salaam, Tanzania

2 Clinical Research, H3 Research Unit of I-Katch Technology Ltd., Dar es Salaam, Tanzania

*Address all correspondence to: celsius_lx@yahoo.co.uk

References

[1] A giant of geriatric medicine – Professor Bernard Isaacs (1924-1995) British Geriatrics Society. Available from: https://www.bgs.org.uk/geriatricmedicinearchive/bgsarchive/biographies/a-giant-of-geriatric-medicine-professor-bernard-isaacs-1924-1995 [Accessed: March 10, 2020]

[2] Morley J. The new geriatric giants. Clinics in Geriatric Medicine. 2017;**33**(3):xi-xii

[3] Emmet K. Nonspecific and atypical presentation of disease in the older patient. Geriatrics. 1998;**53**(2):50-60

[4] Jarrett P, Rockwood K, Carver D, Stolee P, Cosway S. Illness presentation in elderly patients. Archives of Internal Medicine. 1995;**155**(10):1060-1064

[5] Gavazzi G, Krause K. Ageing and infection. The Lancet Infectious Diseases. 2002;**2**(11):659-666

[6] Schattner A, Klein Y, Herskovitz P. An atypical presentation of a common disease. QJM. 2012;**105**(9):891-893

[7] Hughes L, McMurdo M, Guthrie B. Guidelines for people not for diseases: The challenges of applying UK clinical guidelines to older people with multiple co-morbidities. Age and Ageing. 2013;**42**(1):62-69

[8] Bortz W II. A conceptual framework of frailty: A review. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences. 2002;**57**:M283-M288

[9] Fried L, Tangen C, Walston J, Newman A, Hirsch C, Gottdiener J, et al. Frailty in older adults: Evidence for a phenotype. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences. 2001;**56**(3):M146-M156

[10] Rockwood K, Andrew M, Mitniski A. A comparison of two approaches to measuring frailty in elderly people. Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2007;**62**(7):738-743

[11] Rockwood K, Mitniski A. Frailty in relation to the accumulated deficits. Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2007;**62**(7):722-727

[12] Soong J. Functional assessment in older people. BMJ. 2011;**343**:d4681

[13] Fried L, Xue Q, Cappola A, Ferrucci L, Chaves P, Varadhan R, et al. Nonlinear multisystem physiological dysregulation associated with frailty in older women: Implications for etiology and treatment. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences. 2009;**64A**(10):1049-1057

[14] Rumero-Ortuno R, Walsh C, Lawlor B, Kenny R. A frailty instrument for primary care: Findings from the survey of health, ageing and retirement in Europe (SHARE). BMC Geriatrics. 2010;**10**(1):57. DOI: 10.1186/1471-2381-10-57

[15] Rolfson D, Majumdar S, Tsuyuki R, Tahir A, Rockwood K. Validity and reliability of the Edmonton Frail Scale. Age and Ageing. 2006;**35**(5):526-529

[16] Box G. Robustness in the strategy of scientific model building. In: Launer R, Wilkinson G, editors. Robustness in Statistics. Academic Press; 1979. pp. 201-236

[17] Carmines E, Zeller R. Reliability and Validity Assessment. SAGE Publications; 1979. DOI: 10.4135/9781412985642

[18] CSHA Working Group. Canadian Study of Health and Aging: Study

methods and prevalence of dementia. CMAJ. 1994;**150**(6):899-913

[19] Rockwood K, Wolfson C, McDowell I. The Canadian Study of Health and Aging: Organizational lessons from a national, multicenter, epidemiologic study. International Psychogeriatrics. 2001;**13**(suppl 1):233-237

[20] Rockwood K, Song X, MacKnight C, Bergman H, Hogan D, McDowell I, et al. A global clinical measure of frailty in elderly people. CMAJ. 2005;**173**(5):489-495

[21] Streiner D, Norman G. Health Measurement Scales: A Practical Guide to their Development and Use. 3rd ed. Oxford: Oxford University Press; 2003. pp. 4-13

[22] Darvall J, Loth J, Bose T, Braat S, de Silva A, Story D, et al. Accuracy of the Clinical Frailty Scale for perioperative frailty screening: A prospective observational study. Canadian Journal of Anesthesia. 2020;**67**:694-705. DOI: 10.1007/s.12630-020-01610-x

[23] Chong E, Chia J, Law F, Chew J, Chan M, Lim W. Validating a standardized approach in administration of the Clinical Frailty Scale in hospitalized older adults. Annals of the Academy of Medicine, Singapore. 2019;**48**:115-124

[24] Ozsurekci C, Balci C, Kizilarslanoglu C, Caliskan H, Dogrul R, Aycicek G, et al. An important problem in an aging country: Identifying the frailty via a 9 point Clinical Frailty Scale. Acta Clinica Belgica. 2019;**75**:200-204. DOI: 10.1080/17843286.2019.1597457

[25] Rolfson D, Majumdar S, Tahir A, Tsuyuki R. Development and validation of a new instrument for frailty. Clinical and Investigative Medicine. 2000;**23**:336

[26] Perna S, Francis M, Bologna C, Moncaglieri F, Riva A, Morazzoni P, et al. Performance of Edmonton Frail Scale on frailty assessment: Its association with multi-dimensional geriatric conditions assessed with specific screening tools. BMC Geriatrics. 2017;**17**:2. DOI: 10.1186/s 12877-016-0382-3

[27] Fabricio-Wehbe S, Schiaveto F, Vendrusculo T, Haas V, Dantas R, Rodriguez R. Cross-cultural adaptation and validity of 'Edmonton Frail Scale – EFS' in a Brazilian elderly sample. Revista Latino-Americana de Enfermagem. 2009;**17**(6):1043-1049

[28] Polanska B, Uchmanowicz B, Kujawska-Danecka H, Nowicka-Sauer K, Chudiak A, Dudek K, et al. Assessment of frailty syndrome using Edmonton Frailty Scale in Polish elderly sample. The Aging Male. 2018;**22**:177-186. DOI: 10.1080/13685538.2018.1450376

[29] Aygor H, Fadiloglu C, Sahin S, Aykar F, Akcicek F. Validation of Edmonton Frail Scale into elderly Turkish population. Archives of Gerontology and Geriatrics. 2018;**76**:133-137. DOI: 10.1016/j. archger.2018.02.003

[30] Hilmer S, Perera V, Mitchell S, Murnion B, Dent J, Bajorek B, et al. The assessment of frailty in older people in acute care. Australasian Journal on Ageing. 2009;**28**(4):182-188

[31] O'Brien M, Mallett V, Coghlan M, Yen D, Doran E, Williams D, et al. Edmonton Frail Scale as a predictor of adverse events in older patients undergoing systemic cancer therapy in Ireland. Age and Ageing. 2016;**45**:i13-i56

[32] Ramirez R, Cadena S, Ochoa M. Edmonton Frail Scale in Colombian older people: Comparison with the Fried criteria. Archives of Gerontology and Geriatrics. 2017;**75**:91-95

[33] Zhang X, Yang Y, Zhang C, Luo R, Liu Y. Development of a frailty scale

for elderly people in China. Chinese Nursing Research. 2017;**4**(2):64-70. DOI: 10.1016/j.cnre.2017.06.003

[34] Steverink N, Slaets J, Schuurmans H, Lis M. Measuring frailty: Developing and testing the GFI (Groningen Frailty Indicator). The Gerontologist. 2001;**41**:236

[35] Peters L, Boter H, Buskens E, Slaets J. Measurement properties of the Groningen Frailty Indicator in home dwelling and institutionalized elderly people. JAMDA. 2012;**13**:546-551

[36] Bielderman A, Schans C, Lieshout M, Greef M, Boersma F, Krijnen W, et al. Multidimensional structure of the Groningen Frailty Indicator in community-dwelling older people. BMC Geriatrics. 2013;**13**:86. DOI: 10.1186/1471-2381-13-86

[37] Olaroiu M, Ghinescu M, Naumov V, Brinza I, Heuvel W. The psychometric qualities of the Groningen Frailty Indicator in Romanian community-dwelling older citizens. Family Practice. 2014;**31**(4):490-495

[38] Gobbens R, van Assens M, Luijkx K, Wijnen-Sponselee M, Schols M. The Tilburg Frailty Indicator: Psychometric properties. JAMDA. 2009;**11**:344-355. DOI: 10.1016/j.jamda.2009.11.003

[39] Dong L, Liu N, Tian X, Qiao X, Gobbens R, Kane R, et al. Reliability and validity of the Tilburg Frailty Indicator (TFI) among Chinese community-dwelling older people. Archives of Gerontology and Geriatrics. 2017;**73**:21-28

[40] Mulasso M, Roppolo M, Gobbens R, Rabaglietti E. The Italian version of Tilburg Frailty Indicator: Analysis of psychometric properties. Research on Aging. 2015;**38**(8):842-863

[41] Bouillon K, Kivimaki M, Hamer M, Sabia S, Fransson E, Singh-Manoux A, et al. Measures of frailty in a population-based studies: An overview. BMC Geriatrics. 2013;**13**:64. DOI: 10.1186/1471-2381-13-64

[42] Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;**16**:297-334

[43] Sijtsma K. On the use, misuse and very limited use of Cronbach's alpha. Psychometrika. 2009;**74**(1):107-120

[44] Schmitt N. Uses and abuses of coefficient alpha. Psychological Assessment. 1996;**8**:350-353

[45] Lord F. An empirical study of the normality and independence of errors of measurement in test scores. Psychometrika. 1960;**25**:91-104

[46] Lord F, Novick M. Statistical Theory of Mental Test Scores. Reading: Addison-Wesley; 1968

[47] Cortina J. What is coefficient alpha? An examination of theory and practice. Journal of Applied Psychology. 1993;**78**:98-104

[48] Moreau D, Kirk I, Waldie K. Seven pervasive statistical flaws in cognitive training interventions. Frontiers in Human Neuroscience. 2016;**10**:153

[49] Sedgwick P. Relative risks versus odds ratio. BMJ. 2014;**348**:g1407

[50] Ranganathan P, Aggarwal R, Pramesh C. Common pitfalls in statistical analysis: Odds versus risk. Perspectives in Clinical Research. 2015;**6**(4):222-224

[51] Sackett D, Deeks J, Altman D. Down with odds ratio! Evidence-Based Medicine. 1996;**1**(6):164-166

[52] Streiner D. Risky business: Making sense of estimates of risk. Canadian Journal of Psychiatry. 1998;**43**(4):411-415

[53] Anselmi P, Colledani D, Robusto E. A comparison of classical and modern measures of internal consistency. Frontiers in Psychology. 2019;**10**:2714. DOI: 10.3389/fpsyg.2019.02714

[54] Peters L, Boter H, Burgerhof J, Slaets J, Buskens E. Construct validity of the Groningen Frailty Indicator established in a large sample of home-dwelling elderly persons: Evidence of stability across age and gender. Experimental Gerontology. 2015;**69**:129-141

[55] Strauss M, Smith G. Construct validity: Advances in theory and methodology. Annual Review of Clinical Psychology. 2009;**5**:1-25

[56] Sechrest L. Validity of measures is no simple matter. Health Services Research. 2005;**40**(5Pt 2):1584-1604

[57] Kelly P, O'Malley K, Kallen M, Ford M. Integrating validity theory with use of measurement instruments in clinical settings. Health Services Research. 2005;**40**(5 Pt 2):1605-1619

[58] Westen D, Rosenthal R. Quantifying construct validity: Two simple measures. Journal of Personality and Social Psychology. 2003;**84**(3):608-618

[59] Elasy T, Gaddy G. Measuring subjective outcomes: Rethinking reliability and validity. Journal of General Internal Medicine. 1998;**13**(11):757-761

[60] Snow A, Cook K, Lin P, Morgan R, Magaziner J. Proxies and other external raters: methodological considerations. Health Services Research. 2005;**40**(5Pt2):1676-1693

[61] Ioannidis J, Patsopoulos N, Rothstein H. Reasons or excuses for avoiding meta-analysis in forest plots. BMJ. 2008;**336**(7658):1413-1415

[62] Edwardes M. Meta-analysis and the reversed theorem of the means. Research Synthesis Methods. 2014;**5**(4):313-321

[63] Pettiti D. Meta-analysis, Decision Analysis and Cost-effectiveness Analysis. 2nd ed. Oxford University Press; 1999

[64] Esterhuizen T, Thabane L. Meta-analysis: Some key limitations and potential solutions. Nephrology, Dialysis, Transplantation. 2016;**31**:882-885

[65] Leshabari K, Biswas A, Gebuis E, Leshabari S, Ohnishi M. Challenges in morbidity and mortality statistics of the elderly population in Tanzania: A call to action. Quality in Ageing and Older Adults. 2017;**18**(3):171-174