The relative and absolute reliability of the Functional Independence and Difficulty Scale in community-dwelling frail elderly Japanese people using long-term care insurance services

Saito, Takashi
Izawa, P. Kazuhiro
Watanabe, Shuichiro

(Citation)
Aging Clinical and Experimental Research,29(3):549-556

(Issue Date)
2017-06

(Resource Type)
journal article

(Version)
Accepted Manuscript

Original Article

Aging Clinical and Experimental Research

The relative and absolute reliability of the Functional Independence and Difficulty Scale in community-dwelling frail elderly Japanese people using long-term care insurance services

Takashi Saito, RPT, MSc,[1*] Kazuhiro P. Izawa, RPT, PhD,[2] Shuichiro Watanabe, MD, PhD[3]

[1] Doctoral Course of Gerontology, Graduate School of Gerontology, J. F. Oberlin University, Japan

[2] Graduate School of Health Sciences, Kobe University, Kobe, Japan

[3] Graduate School of Gerontology, J. F. Oberlin University, Tokyo, Japan

***Corresponding author**: Takashi Saito

Doctoral Course of Gerontology, Graduate School of Gerontology, J. F. Oberlin University, 3758 Tokiwa-machi, Machida-shi, Tokyo 194-0294, Japan

Tel.: +81 03-5367-1321; FAX: +81 03-5367-1329; E-mail: takashi-riha@live.jp

ABSTRACT

Background The newly developed Functional Independence and Difficulty Scale is a tool for assessing the performance of basic activities of daily living in terms of both independence and difficulty. The reliability of this new scale has not been assessed.

Aims The aims of this study were to examine the relative reliability and absolute reliability of the newly developed scale in community-dwelling frail elderly people in Japan.

Methods Participants were 47 community-dwelling elderly subjects (22 for assessing test-retest reliability and 25 for assessing inter-rater reliability). As relative reliability indices, intra-class correlation coefficients were used. From an absolute reliability perspective, we conducted Bland-Altman analysis and calculated the limit of agreement or minimal detectable change to determine the acceptable range of error.

Results Intra-class correlation coefficients for test-retest and inter-rater reliability were 0.90 (P <0.001) and 0.97 (P <0.001), respectively. The limit of agreement for test-retest reliability was -5.2 to 1.8, representing an increase of over 6 points for improvement and a decrease of over 2 points for decline of basic activities of daily living ability. The minimal detectable change for inter-rater reliability was 3.7, indicating that a 3-point difference might be exist between difference raters. The results of this study demonstrated that the FIDS appeared to be a reliable instrument for use in Japanese community-dwelling frail elderly people.

Conclusions While further research using a large and more diverse sample of participants is needed, our findings support the use of FIDS in clinical practice or clinical research targeting frail elderly Japanese people.

Key Words: Activities of daily living, Relative reliability, Absolute reliability, Bland-Altman analysis, Kappa coefficient, Intra-class correlation coefficient

Introduction

Aging is associated with progressive loss of neuromuscular function that often leads to progressive disability and loss of independence in the elderly [1]. Several studies have shown that disability in basic activities of daily living (BADL) is correlated with multimorbidity [2], decreased muscle strength [3], impaired static and dynamic balance [4], slow walking speed [5] and falls [6]. Moreover, BADL disability has been identified as a predictor of adverse health outcomes, such as hospitalization [7], use of formal and informal home care services [8], institutionalization [9] and death [9] in older people. Therefore, the assessment of BADL disability has become increasingly important in both patient care and clinical research as the population of older people with multiple chronic diseases and infirmities grows [9].

Although BADL disability can be defined as dependence, i.e., "requiring help from another person" or difficulty with a BADL task [10], Gill and colleagues [9] emphasized the usefulness of a BADL scale for assessing both independence and difficulty. Using data from cross-sectional and longitudinal analyses, they found that elderly people who were BADL independent with difficulty had functional profiles, physical performance scores and rates of health care utilization and death that were intermediate to those of elderly people who were independent without difficulty and persons who were dependent [9].

These findings implied that older persons who were BADL independent with difficulty are more frail than those without difficulty and are a population at high risk for adverse health outcomes. To capture these elderly people and evaluate their BADL disability, scales assessing either BADL dependency or difficulty would not be adequate. Thus, scales that can assess both BADL independence and difficulty are needed.

Consequently, we have developed a new BADL assessment tool, the Functional Independence and Difficulty Scale (FIDS), for community-dwelling elderly people in Japan

68 [11]. In assessing both independence and difficulty in performing BADL, FIDS showed
69 acceptable item validity, internal consistency and external validity [11, 12]. These findings
70 suggested that FIDS might be a useful tool for assessing BADL disability in the field of
71 community health care or community welfare for elderly people.
72 However, little is known about the relative reliability and absolute reliability of FIDS.
73 Relative reliability is the degree to which individuals maintain their position in a sample with
74 repeated measurements [13], and absolute reliability is the degree to which repeated
75 measurements vary for individuals [13]. Researchers and clinicians are interested in the relative
76 reliability, as expressed by the intraclass correlation coefficient (ICC), and absolute reliability,
77 as expressed by Bland-Altman analysis [14] or the acceptable range of error, of a measurement
78 instrument. These measurement properties affect the interpretation of the value of a measure and
79 the results of an intervention program. To use FIDS as an index of the health state of elderly
80 people for descriptive purpose or for observing change over time or outcome, the reliability of
81 FIDS is a crucial measurement property and needs to be evaluated.
82 When using an assessment tool or evaluating its reliability, assessment modes should be
83 considered, i.e., interview modes or self-administered modes. Each of these modes has its merits
84 and demerits regarding cost-effectiveness, response rate [15] and validity of response [16, 17].
85 When assessing BADL disability of seriously ill or frail elderly people, the face-to-face
86 interview mode may be appropriate because these people often have visual or cognitive
87 impairment or are not good at writing due to physical impairment. BADL disability is one of the
88 matters of concern in seriously ill or frail elderly people, and evaluating the reliability of FIDS
89 in face-to-face modes is important for clinical practice or clinical research targeting them.
90 Therefore, in light of the lack of data relating to the reliability of the newly developed FIDS, the
91 aim of this study was to evaluate the relative and absolute reliability of the FIDS in face-to-face

92 modes for community-dwelling frail elderly people in Japan.

93

94 **Methods**

95 *Subjects*

96 The participants were community-dwelling frail elderly people using Japanese long-term care
97 insurance (LTCI) services [18]. The elderly people using LTCI services have physical or mental
98 impairment and need assistance in performing BADL and/or instrumental activities of daily
99 living. In Japan, a mandatory LTCI system was implemented in 2000. Municipalities are
100 responsible for certification of long-term care needs based on the evaluation results by the
101 Certification Committee for Long-term Care Need [18]. The nationally uniform level of
102 long-term care need [18] was based on the insured's mental and physical conditions and on
103 family doctors' letters of opinions.

104 We included participants from a home-visit nursing station in Kawasaki city, Kanagawa
105 prefecture, Japan. Kawasaki city is located 20 km south of Tokyo. The participants were
106 individuals who were registered as users of LTCI services including home-visit nursing care or
107 rehabilitation provided from the home-visit nursing station. The researchers visited each
108 participant's home and invited them to participate in this study.

109 Of the 63 subjects initially invited, 27 subjects participated in the assessment of test-retest
110 reliability and 36 subjects participated in the assessment of inter-rater reliability. The inclusion
111 criterion was a minimum age of 65 years, and the exclusion criterion was cognitive impairment
112 (error of Mental State Questionnaire (MSQ) ≥9 [19, 20]). Participants who did not match these
113 criteria or who did not want to participate in research procedures voluntarily were excluded
114 from the present study.

115

Raters

The raters were three physical therapists (PT), one (PT1) for assessing test-retest reliability, and the other two (PT2, PT3) for assessing inter-rater reliability. The three PTs each had at least 3 years of training and clinical experience in the rehabilitation of community-dwelling elderly people. The raters were instructed on how to assess FIDS score, and thereafter, the raters familiarized themselves with FIDS by using it in their clinical practice. Prior to the test sessions, the raters used FIDS and assessed BADL performance of at least five elderly people who were not participants in this study.

Testing procedure

All testing procedures were conducted in the participants' home. Data collection was carried out from August to October 2014.

The two test sessions (T1 and T2) were performed separately. Because our participants were frail and vulnerable elderly people, their physical condition and BADL ability might have changed if a long interval have been set between T1 and T2. We limited the interval between T1 and T2 to 7 days to avoid changes in BADL ability and to minimize recall bias of the rater and participants.

In the T1 session, the rater assessed the participants' basic characteristics and FIDS. Basic characteristics included age, sex, height, weight, chronic disease (cerebrovascular disease, musculoskeletal disease, internal disease, neuromuscular disease and others), long-term care need level certification [18] and degree of independence of daily living.

The criteria for long-term care need level certification in Japan are as follows: Requiring support 1, Requiring support 2, Requiring long-term care 1, Requiring long-term care 2, Requiring long-term care 3, Requiring long-term care 4 and Requiring long-term care 5 [18].

Benefits according to the long-term care level are set to minimum for Requiring support 1 and maximum for Requiring long-term care 5. The people classified into the two support levels are able to independently perform BADL and are considered to need some support to prevent an increase in eligibility level due to physical or mental impairments, whereas people classified into the long-term care levels need assistance to perform basic activities of daily living [21]. Typically, the elderly people in care levels 1–2 can walk independently, whereas those in care levels 3–5 have difficulty in walking alone [21]. Some studies [22, 23] distinguished between care levels 1–2 and 3–5, with the former called "moderately disabled" and the later called "severely disabled."

Criteria for assessing the degree of independence of daily living were as follows: independent: going outside independently; house-bound: needing help to go outside but, in general, living independently in their house; and bed-bound: needing help for any BADL and mainly living in bed.

FIDS was assessed through face-to-face interview. The FIDS [11] comprises 14 items of BADL: getting up from bed, standing up from a chair, standing up from the floor, dressing, putting on pants, eating, cleaning after toileting, washing, brushing teeth, opening a PET bottle, cutting toenails, walking inside, walking outside and going up and down 4 to 6 steps. The FIDS has two questions for each item, one question about independence (A: Do you need someone's help to do activity X?) and one about difficulty (B: Do you have any difficulty in doing activity X?). The response to each question was designed to be simply "yes" (need help/unable or have difficulty in doing) or "no" (need no help or have no difficulty). Scores were assigned as follows: 1 if the participant reported being dependent or unable to perform the activity (response to question A is "yes" and response to question B is not considered), 2 if independence with difficulty was reported (response to question A is "no" and response to question B is "yes") and

3 if independence without difficulty was reported (responses to questions A and B are both "no"). Thus, function scores for the FIDS range from 14-42, with higher scores representing better function.

In the T2 session, the rater assessed FIDS again. In assessing test-retest reliability, the rater (PT1) was instructed to assess FIDS without referring to the FIDS score obtained in the T1 session. In assessing inter-rater reliability, the two raters (PT2, PT3) were instructed not to communicate with each other about the FIDS score obtained during the T1 and T2 testing procedures. Moreover, PT2 and PT3 alternated between which rater performed the test first to minimize any bias from the order of raters.

Statistical analysis

All statistical tests were conducted using IBM SPSS Statistics (Version 22, IBM Japan Ltd.) and Microsoft® Excel (Version 2007, Microsoft Japan). A two-tailed *P* value of <0.05 was considered significant.

To assess the agreement for each item of the FIDS between T1 and T2, the un-weighted Kappa coefficient was calculated. To determine the relative reliability, ICC (1, 1) and ICC (2, 1) were calculated. To assess the absolute reliability, Bland-Altman analysis [14], limit of agreement (LOA) [14] and 95% confidence interval (CI) of the minimal detectable change (MDC_{95}) [24] were used.

We investigated the types of error, which can be classified roughly into two types: accidental error and systematic error. Accidental errors include biological variation and measurement error and represent deviations in a non-specific direction. Accidental errors cannot be controlled and can exist in all measurements. In contrast, systematic errors represent deviation in a specific direction and are divided into fixed bias and proportional bias. Fixed bias represents deviation

188 with a specific width in a specific direction regardless of the true value, whereas proportional
189 bias represents deviation that changes in a specific direction in proportion to the true value.
190 Systematic errors, i.e., fixed bias and proportional bias, between the T1 and T2 values were
191 analyzed with Bland-Altman analysis. When 0 was not included in the 95% CI of the mean
192 difference between the T1 and T2 values (difference of T1 minus T2 score), this was considered
193 to indicate the presence of a fixed bias. To investigate the presence of proportional bias, a test of
194 no correlation was conducted. Proportional bias was considered to exist if the regression was
195 significant. When neither fixed bias nor proportional bias was detected, only accidental error
196 was considered to exist.
197 LOA and MDC_{95} were calculated to determine the acceptable range of error. If a systematic
198 error was present, we calculated the LOA. The LOA was adapted to the most optimistic range
199 between the lower LOA and upper LOA [25]. When an accidental error was considered to exist,
200 MDC_{95} was calculated.
201 The formulae used to calculate 95% CI of the mean difference between the T1 and T2 values
202 [25], LOA [14] and MDC_{95} [24] are as follows:
203 ・95% CI of the mean difference between the T1 and T2 values = $\bar{d} \pm t \times \sqrt{SD_d^2/n}$
204 ・Lower LOA = $\left(\bar{d} - 1.96 \times SD_d\right) \pm t \times \sqrt{3SD_d^2/n}$
205 ・Upper LOA = $\left(\bar{d} + 1.96 \times SD_d\right) \pm t \times \sqrt{3SD_d^2/n}$
206 ・Optimistic LOA = $\left(\bar{d} - 1.96 \times SD_d\right) + t \times \sqrt{3SD_d^2/n} \sim \left(\bar{d} + 1.96 \times SD_d\right) -$
207 $t \times \sqrt{3SD_d^2/n}$
208 ・$MDC_{95} = 1.96 \times SD_d$
209 where n = the number of measured subjects, d = the difference between T1 and T2 values
210 (T1-T2), $\bar{d}$ = the mean of the difference between the T1 and T2 values, SD_d = the standard
211 deviation of the difference between the T1 and T2 values, and t = the point of the t distribution

(degree of freedom is n-1 and P <0.05).

RESULTS

Subjects

Of the 63 participants, we excluded 14 subjects from the analysis because of the presence of cognitive impairment as indicated by the MSQ, one subject because of refusal to participate in this research and one subject because of missing values. Therefore, the final sample for analysis comprised 47 respondents (22 for assessment of test-retest reliability and 25 for assessment of inter-rater reliability) (Figure 1).

Characteristics and other variables in the subjects

Table 1 presents the characteristics of the subjects and results of the comparison tests between the two study groups. Except for the age of the participants, there were no other significant differences between the two groups. Subjects of the test-retest reliability group were significantly older than subjects of inter-rater reliability group (P <0.05).

In assessing test-retest reliability (n = 22; 14 women, 8 men), the average age was 84.1 years. Among these 22 subjects, 6 satisfied the criteria of independence degree as “Independent,” 12 as “House-bound” and 4 as “Bed-bound.” Seven were registered for LTCI care levels “requiring support 1-2,” 10 as “requiring long-term care 1-2” and 5 as “requiring long-term care 3-5.” Average total scores (SD) of the FIDS were 29.7 (7.7) for T1 and 31.4 (7.4) for T2.

In assessing inter-rater reliability (n = 25; 14 women, 11 men), the average age was 79.1 years. Among these 25 subjects, 11 satisfied the criteria of independence degree as “Independent,” 13 as “House-bound” and 1 as “Bed-bound.” Eleven were registered for LTCI care levels “requiring support 1-2,” 11 as “requiring long-term care 1-2” and 3 as “requiring

long-term care 3-5." Average total scores (SD) of the FIDS were 32.0 (7.0) for T1 and 31.9 (7.3) for T2.

Agreement

The agreement for each item of the FIDS between T1 and T2 is shown in Table 2. In assessing test-retest reliability, the un-weighted Kappa coefficient ranged from 0.46 to 0.92 (mean [standard deviation; SD] was 0.66 [0.15]). In assessing inter-rater reliability, the un-weighted Kappa coefficient ranged from 0.41 to 0.77 (mean [SD] was 0.64 [0.10]).

Relative reliability

For assessing test-retest reliability, the ICC (1, 1) was 0.90 (P <0.001), and the 95% CI was 0.78 to 0.96. For assessing inter-rater reliability, the ICC (2, 1) was 0.97 (P<0.001), and the 95% CI was 0.96 to 0.99.

Absolute reliability

The results of Bland-Altman analysis are shown in Table 3. In assessing test-retest reliability, fixed bias was detected, and the LOA was -5.2 to 1.8. In assessing inter-rater reliability, no fixed bias or proportional bias was detected, and the MDC_{95} was 3.7.

Discussion

This is the first demonstration of the relative and absolute reliability of a new BADL assessment tool, the FIDS. The results of agreement and relative reliability were satisfactory. Moreover, results of absolute reliability suggested useful information for detecting real change of BADL ability. Our findings support the use of FIDS as an indicator of BADL performance in clinical

260 practice or clinical research targeting frail elderly Japanese people using LTCI services.

261 The agreement and relative reliability of the FIDS were calculated by un-weighted Kappa
262 coefficient and ICC, respectively. A Kappa coefficient above 0.81 is considered as “almost
263 perfect,” that between 0.61 and 0.80 as “substantial” and that between 0.41 and 0.60 as
264 “moderate [26].” An ICC above 0.9 was considered as good reliability, between 0.7 and 0.9 as
265 average reliability and less than 0.7 as poor reliability [27]. Moreover, the recommended ICC
266 for an assessment tool is ICC>0.7 for a large group (as in research) or ICC> 0.9 for individuals
267 [28]. Our Kappa coefficient and ICC results met these criteria, as “moderate” to “almost perfect”
268 and as “excellent reliability,” respectively. These findings supported that FIDS is an assessment
269 tool that ensures the agreement and the relative reliability for use in individuals.

270 Although relative reliability is a measure of the degree to which individuals maintain their
271 test results with repeated measurements, relative reliability cannot, in practical use, realize
272 certain analytical goals. For the clinical practitioner or researcher, information on absolute
273 reliability, which can be used to interpret the magnitude of “real change” in individual patients
274 or subjects, is needed. It would be important to consider the values of the acceptable range of
275 error for use in individuals.

276 Our Bland-Altman analysis showed an acceptable range of error for test-retest reliability of
277 -5.2 to 1.8 and inter-rater reliability of 3.7. This suggested that when assessing BADL ability in
278 FIDS for pre/post intervention effect, an increase of over 6 points indicates improvement, and a
279 decrease of over 2 points indicates decline. Similarly, when different raters assess FIDS, there
280 may be a 3-point difference between them.

281 Although our findings of absolute reliability could imply that FIDS can be used to detect real
282 changes in an older population, some points should be noted when interpreting these results.
283 First, although no systematic error was detected in inter-rater reliability, systematic error (fixed

bias) was detected in test-retest reliability. This discrepancy may be because that the alternation of raters avoided bias from order effect in inter-rater reliability, but learning effect might influence the FIDS score between the two sessions in test-retest reliability. Second, our results of acceptable range of error were relatively wide. This may because our subjects had various chronic diseases and a wide range in their levels of cognitive impairment and independence degree of daily living. This heterogeneity in background variables might influence the acceptable range of error of the FIDS.

This study has several limitations. First, the study sample was small and comprised only frail elderly using LTCI services. Therefore, the generalization of our results was restricted by known and unknown selection bias. Further research using a large and more diverse sample of participants is needed. This additional research may reveal a more specific type of reliability of FIDS, e.g., reliability of FIDS by LTCI care levels, reliability of FIDS in healthy elderly using no LTCI services and disease-specific reliability of FIDS. Second, we assessed the reliability of FIDS in the face-to-face mode, and thus it may not be generalizable to self-administered modes. Further research assessing the reliability of FIDS in self-administered modes is needed. Moreover, if the reliability of FIDS in different modes, i.e., face-to-face mode and self-administered mode, were verified, a FIDS score based on the different modes could be considered equivalent and comparable within the same subject.

Conclusion

In conclusion, the results of this study demonstrated that the FIDS appeared to be a reliable instrument for Japanese community-dwelling frail elderly people using LTCI services. Although further research using a large and more diverse sample of participants is needed, our findings support the use of FIDS in clinical practice or clinical research targeting frail elderly Japanese

people.

Acknowledgments The authors wish to thank all of the subjects who participated in the study. We also thank the staff of the Department of Rehabilitation of the Visiting Nursing and Rehabilitation Network for their help with data collection.

Authorship T.S. and S.W. designed the study, collected and analyzed the data, interpreted the results and prepared the manuscript. K. I. and S.W. supervised data analysis and manuscript preparation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The study protocol was approved by the ethics Committee of J. F. Oberlin University (approval no. 14036). The participants received oral and written explanations of the research procedures by the researchers. Participation in the study was voluntary. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Doherty TJ (2003) Invited review: Aging and sarcopenia. J Appl Physiol 95:1717–1727.
2. St John PD, Tyas SL, Menec V, Tate R (2014). Multimorbidity, disability, and mortality in community-dwelling older adults. Can Fam Physician 60:e272–e280.
3. Hairi NN, Cumming RG, Naganathan V, Handelsman DJ, Le Couteur DG, Creasey H, Waite LM, Seibel MJ, Sambrook PN (2010) Loss of muscle strength, mass (sarcopenia), and quality (specific force) and its relationship with functional limitation and physical disability: the Concord Health and Ageing in Men Project. J Am Geriatr Soc 58:2055–2062. doi: 10.1111/j.1532-5415.2010.03145.x.
4. Candela F, Zucchetti G, Magistro D, Ortega E, Rabaglietti E (2014) Real and perceived physical functioning in Italian elderly population: associations with BADL and IADL. Adv Aging Res 3:349–359. doi: 10.4236/aar.2014.35045.
5. Seino S1, Yabushita N, Kim MJ, Nemoto M, Jung S, Osuka Y, Okubo Y, Matsuo T, Tanaka K (2011) Comparison of a combination of upper extremity performance measures and usual gait speed alone for discriminating upper extremity functional limitation and disability in older women. Arch Gerontol Geriatr 55:486–491. doi: 10.1016/j.archger.2011.10.011.
6. Langlois JA, Norton R, Campbell AJ, Leveille S (1999) Characteristics and behaviours associated with difficulty in performing activities of daily living among older New Zealand women. Disabil Rehabil 21:365–371.
7. Hirose T, Hasegawa J, Izawa S, Enoki H, Kuzuya M (2011) Investigation of whether depressive status contributes to mortality and hospitalization in community-dwelling dependent older people. Nihon Ronen Igakkai Zasshi 48:163–169 (in Japanese with English abstract).

8. Coughlin TA, McBride TD, Perozek M, Liu K (1992) Home care for the disabled elderly: predictors and expected costs. Health Serv Res 27:453–479.
9. Gill TM, Robison JT, Tinetti ME (1998) Difficulty and dependence: two components of the disability continuum among community-living older persons. Ann Intern Med 128:96–101.
10. Gill TM (2010) Assessment of function and disability in longitudinal studies. J Am Geriatr Soc 58:S308–S312. doi: 10.1111/j.1532-5415.2010.02914.x.
11. Saito T, Izawa KP, Omori Y, Watanabe S (2015) The Functional Independence and Difficulty Scale: Instrument Development and Validity Evaluation. Geriatr Gerontol Int. doi: 10.1111/ggi.12605.
12. Saito T, Izawa KP, Matsui N, Arai K, Ando M, Morimoto K, Fujita N, Takahashi Y, Kawazoe M, Watanabe S (2016) Comparison of the measurement properties of the Functional Independence and Difficulty Scale with the Barthel Index in community-dwelling elderly people in Japan. Aging Clin Exp Res. Doi: 10.1007/s40520-016-0558-x
13. Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 26:217–238.
14. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 8:307–310.
15. Christensen AI, Ekholm O, Glümer C, Juel K (2014) Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys. Eur J Public Health 24:327-332. doi: 10.1093/eurpub/ckt067.
16. Bergmann MM, Jacobs EJ, Hoffmann K, Boeing H (2004) Agreement of self-reported medical history: comparison of an in-person interview with a self-administered questionnaire. Eur J Epidemiol 19:411-416.

379 17. Kraus L, Augustin R (2001) Measuring alcohol consumption and alcohol-related problems:
380 comparison of responses from self-administered questionnaires and telephone interviews.
381 Addiction 96:459-471.
382 18. Ministry of Health, Labour and Welfare. Long-Term Care, Health and Welfare Services for
383 the Elderly [Monograph on the internet]. Ministry of Health, Labour and Welfare. [cited
384 2016 Jan 30]. Available from: http://www.mhlw.go.jp/english/wp/wp-hw6/dl/10e.pdf.
385 19. Kahn RL, Goldfarb AI, Pollack M, Peck A (1960) Brief objective measures for the
386 determination of mental status in the aged. Am J Psychiatry 117:326–328.
387 20. Suzukawa M, Shimada H, Tamura M, Suzuki T, Inoue N (2011) The relationship between
388 the subjective risk rating of specific tasks and falls in frail elderly people. J Phys Ther Sci
389 23:425–429. doi: http://doi.org/10.1589/jpts.23.425.
390 21. Kamiya K, Sasou K, Fujita M, Yamada S (2013) Predictors for increasing eligibility level
391 among home help service users in the Japanese long-term care insurance system. Biomed
392 Res int. doi:10.1155/2013/374130.
393 22. Hayashi Y, Suzukawa M, Hato S, Ishimoto M, Kanaya Y, Shimada H (2013) Physical
394 performance associated with a decline in ADL in frail elderly people using long-term care
395 insurance with a day-care service. J Jpn Phys Ther Assoc 40:407–413 (in Japanese with
396 English abstract).
397 23. Suzukawa M, Shimada H, Makizako H, Watanabe S, Suzuki T (2009) Incidence of falls
398 and fractures in disabled elderly people utilizing long-term care insurance. Nihon Ronen
399 Igakkai Zasshi 46:334–340 (in Japanese with English abstract).
400 24. Faber MJ, Bosscher RJ, van Wieringen PC (2006) Clinimetric properties of the
401 performance-oriented mobility assessment. Phys Ther 86:944–954.
402 25. Shimoi T, Tani H (2008) The intra-rater and inter-rater reliability of tandem gait test with

403 the Bland-Altman analysis. Rigakuryoho Kagaku 23:625–631 (in Japanese with English
404 abstract). doi: 10.1589/rika.23.625.
405 26. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data.
406 Biometrics 33:159–174.
407 27. Nikneshan S, Mohseni S, Nouri M, Hadian H, Kharazifard MJ (2015) The effect of emboss
408 enhancement on reliability of landmark identification in digital lateral cephalometric
409 images. Iran J Radiol 12: doi: 10.5812/iranjradiol.19302.
410 28. Lopes AD, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, Dos Reis FB (2008) Validity
411 and reliability of the Western Ontario Rotator Cuff Index (WORC) for use in Brazil. Clin J
412 Sport Med 18:266–272. doi: 10.1097/JSM.0b013e31817282f4.

Table 1 Characteristics of the subjects and results of comparison test between the two study groups

	Test-retest reliability (n = 22)	Inter-rater reliability (n = 25)	Test statistic	*P*-value
Age (years)	84.1 ± 7.6	79.1 ± 7.6	2.28[a]	<0.05[d]
Sex, n (%)				
Male	8 (36.4)	11 (44.0)	0.28[b]	0.60[e]
Female	14 (63.6)	14 (56.0)		
BMI (kg/m^2)	21.2 ± 4.0	22.5 ± 3.8	-0.77[c]	0.44[f]
Independence degree of daily living, *n* (%)				
Independent	6 (27.3)	11 (44.0)		
House-bound	12 (54.5)	13 (52.0)	-1.57[c]	0.12[f]
Bed-bound	4 (18.2)	1 (4.0)		
Level of LTC, *n* (%)				
Requiring support 1-2	7 (31.8)	11 (44.0)		
Requiring long-term care 1-2	10 (45.5)	11 (44.0)	-1.28[c]	0.20[f]
Requiring long-term care 3-5	5 (22.7)	3 (12.0)		
Chronic disease, *n* (%)				
Cerebrovascular diseases	6 (27.3)	4 (16.0)		
Musculoskeletal disease	11 (50.0)	12 (48.0)		
Internal disease	2 (9.1)	4 (16.0)	0.92[b]	0.21[e]
Neuromuscular disease	2 (9.1)	4 (16.0)		
Others	1 (4.5)	1 (4.0)		
Error of MSQ, *n* (%)				
0-2	15 (68.2)	22 (88.0)		
3-5	6 (27.3)	2 (8.0)	-1.57[c]	0.12[f]
6-8	1 (4.5)	1 (4.0)		

Total FIDS score (points)				
Session 1	29.7 (7.7)	32.0 (7.0)	-1.08[a]	0.29[d]
Session 2	31.4 (7.4)	31.9 (7.3)	-0.26[a]	0.80[d]

Values are means ± standard deviation and percentage (%). *BMI* body mass index, *FIDS* Functional Independence and Difficulty Scale, *LTC* long-term care, *MSQ* Mental State Questionnaire.

[a] *t* value

[b] Chi-square value

[c] z value

[d] Two-sample *t*-test.

[e] Chi-square test.

[f] Unpaired Mann-Whitney test.

413

Table 2 Unweighted Kappa coefficient for each item of the Functional Independence and Difficulty Scale

	Test-retest reliability (n = 22)		Inter-rater reliability (n = 25)	
	Kappa coefficient	*P*-value	Kappa coefficient	*P*-value
Getting up from bed	0.53	<0.001	0.60	0.001
Standing up from a chair	0.81	<0.001	0.67	<0.001
Standing up from the floor	0.66	<0.001	0.41	0.007
Dressing	0.80	<0.001	0.71	<0.001
Putting on pants	0.49	0.002	0.75	<0.001
Eating	0.48	<0.001	0.60	0.001
Cleaning after toileting	0.67	<0.001	0.51	0.002
Washing	0.92	<0.001	0.72	<0.001
Brushing teeth	0.56	<0.001	0.60	0.001
Opening a PET bottle	0.65	<0.001	0.77	<0.001
Cutting toenails	0.75	<0.001	0.74	<0.001
Walking inside	0.46	0.001	0.59	<0.001
Walking outside	0.62	<0.001	0.66	<0.001
Going up and down 4 to 6 steps	0.90	<0.001	0.63	<0.001
Mean ± standard deviation	0.66 ± 0.15		0.64 ± 0.10	

414

Table 3 Absolute reliability of the Functional Independence and Difficulty Scale

	Bland-Altman analysis					
	Fixed bias		Proportional bias		Acceptable range of error	
	95% CI	Bias	r (P-value)	Bias	LOA	MDC_{95}
Test-retest reliability (n = 22)	-3.0 to -0.4	Exists	0.12 (0.60)	Does not exist	-5.2 to 1.8	-
Inter-rater reliability (n = 25)	-0.7 to 0.9	Does not exist	-0.18 (0.39)	Does not exist	-	3.7

95% CI 95% confidence interval of the mean difference between T1 and T2 values, *LOA* limit of agreement, MDC_{95} 95% confidence interval of the minimal detectable change

Figure Legend

Figure 1 Flow diagram of the study selection process and testing procedure.

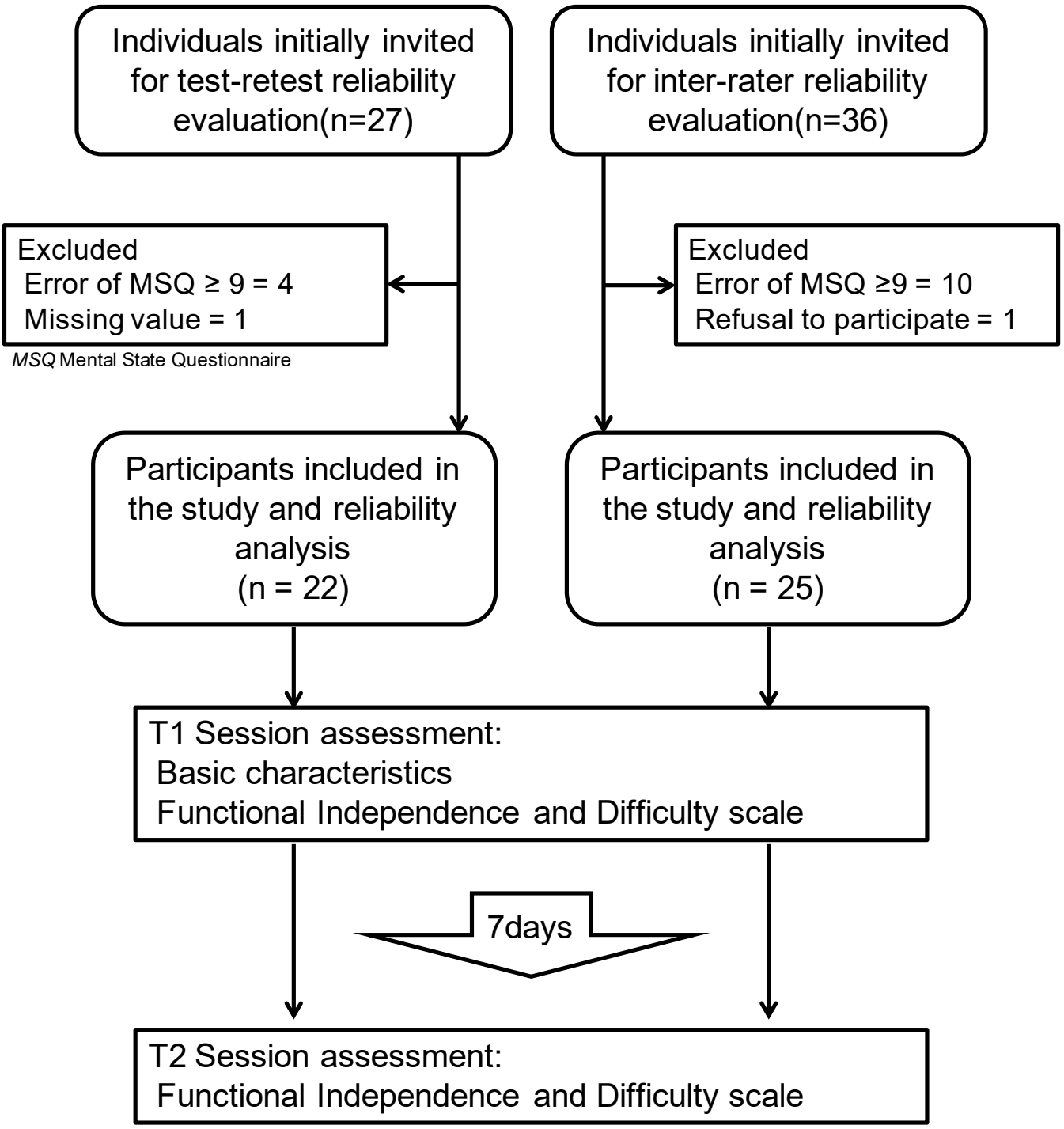
Individuals initially invited for test-retest reliability evaluation(n=27)
Individuals initially invited for inter-rater reliability evaluation(n=36)
Excluded
Error of MSQ ≥ 9 = 4
Missing value = 1
Excluded
Error of MSQ ≥9 = 10
Refusal to participate = 1
MSQ Mental State Questionnaire
Participants included in the study and reliability analysis (n = 22)
Participants included in the study and reliability analysis (n = 25)
T1 Session assessment:
Basic characteristics
Functional Independence and Difficulty scale
7days
T2 Session assessment:
Functional Independence and Difficulty scale

Figure 1.