BMJ Open

Practical guidance to handle missing values in the 25-question Geriatric Locomotive Function Scale (GLFS-25): a simulation study

Takuya Kawahara,[1] Keiko Yamada ,[2,3] Ryohei Terashima,[4] Ikumi Takashima,[1] Sakae Tanaka ,[2] Toru Ogata,[5] Hirotaka Chikuda,[6] Hiromasa Miura,[7] Kozo Nakamura,[8] Takashi Ohe[9]

To cite: Kawahara T, Yamada K, Terashima R, *et al*. Practical guidance to handle missing values in the 25-question Geriatric Locomotive Function Scale (GLFS-25): a simulation study. *BMJ Open* 2022;**12**:e065607. doi:10.1136/bmjopen-2022-065607

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-065607).

TK and KY contributed equally.

Received 13 June 2022
Accepted 15 December 2022

Check for updates

For numbered affiliations see end of article.

Correspondence to
Dr Keiko Yamada;
yamadak@adm.h.u-tokyo.ac.jp

ABSTRACT

Objectives Despite the possible large number of missing values on the 25-question Geriatric Locomotive Function Scale (GLFS-25), how we should treat them is unknown. In a simulation study, we investigated how to handle missing values in the GLFS-25.

Design, setting and participants We used three datasets with different participant characteristics: community dwellers who could walk by themselves, outpatients of orthopaedics owing to pain, and patients who required surgery for total knee replacement or lumbar spinal canal stenosis.

Outcome measures The missing items of the datasets were artificially created, and four statistical methods, complete case analysis, multiple imputation, single imputation using individual mean, and single imputation using individual domain average, were compared in terms of bias and mean squared error. Simulation studies were conducted to compare them under varying numbers of participants with missing values (5%–40%) and under varying numbers of missing items of GLFS-25 (4–16).

Results Multiple imputation had the lowest root mean squared error. Complete case analysis showed the largest bias, and the performances of the single imputation were between those methods. The relative performances were similar across the three datasets. The absolute bias of the single imputation was<0.1. The bias and mean squared error of multiple imputation and single imputation were comparable when the number of missing items was less than or equal to eight.

Conclusions Multiple imputation is preferable, although single imputation using subject average/subject domain average can be used with practically negligible bias as long as the number of missing items is up to 8 out of 25 items in each individual of the population.

STRENGTHS AND LIMITATIONS OF STUDY

⇒ Three datasets were used with different participant characteristics.

⇒ Four statistical methods were compared in terms of bias and mean squared error.

⇒ However, the simulation study area was limited since the missing patterns considered were limited to the five most frequent patterns in actual datasets.

INTRODUCTION

Locomotive syndrome is a concept proposed by the Japanese orthopaedic Association in 2007 to tackle the issues of mobility decrease in Japan's super-ageing society.[1 2] This concept is defined as the 'mobility decrease leading to disabity' based on musculoskeletal disorders.[1 2] As mobility decreases, people have more difficulty in everyday activities such as walking, standing up or climbing stairs. The quantification of mobility decrease leading to disability was determined by the locomotive syndrome risk test, which comprises two physical tests and one self-administered questionnaire, the 25-question Geriatric Locomotive Function Scale (GLFS-25).[1 2] The GLFS-25 is composed of 25 items correlated with body pain, usual care, cognition, movement-related difficulty and social activities. GLFS-25 includes pain, since pain and functional deterioration are mutually related with.[3] The score of the each GLFS-25 item ranges from 0 to 4 (0; best, 4: worst). The highest possible GLFS-25 score is 100, indicating the worst condition.[4]

The GLFS-25 is an easily usable screening tool, but it takes a longer time to complete in clinical settings because of its 25 questions, resulting in participants omitting some of the questions.[5 6] A previous study reported that only 50%–70% of the patients answered all 25 questions, which leads to many missing values.

The large number of missing values causes bias, which can mislead the interpretation of the results.[7] However, despite the possible large number of missing values for GLFS-25, how we should treat them is unknown.

Table 1 Participant characteristics

			Surgical patients	
	Community dwellers	**Outpatients**	**TKR**	**LSCS**
N	9044	743	110	75
Age (mean, SD)	52.0 (18.3)	67.4 (16.1)	74.6 (5.9)	74.0 (7.6)
Sex (female, %)	58.9	73.0	80.0	34.7
GLFS-25 (mean, SD)	4.5 (5.9)	13.7 (11.8)	38.3 (15.7)	39.0 (17.2)

GLFS-25, 25-question Geriatric Locomotive Function Scale; LSCS, lumbar spinal canal stenosis; TKR, total knee replacement.

Therefore, the aim of this study was to present practical guidance on the choice of handling missing values for GLFS-25 based on the characteristics of the dataset.

METHODS

Data

In this study, we used three observational datasets to assess the impact on the estimates (ie, overall average of GLFS-25) by the difference in the statistical methods, such as multiple imputation and complete case analysis. Three observational datasets had different participant characteristics: community dwellers who could walk by themselves, outpatients of orthopaedic departments owing to pain, and patients who required surgery for total knee replacement (TKR) or lumbar spinal canal stenosis (LSCS).

The first and second datasets were obtained from a nationwide study, as described previously.[8] In brief, this study collected data from 10 444 adults aged 20–89 years in Japan from 2017 to 2019 to define the reference values of GLFS-25 and other physical tests (locomotive syndrome risk test). In this study, we used 9044 participants who could walk by themselves and had data on both sex and age as the first dataset. In the same study, we used 743 patients who were orthopaedic outpatients owing to pain who had data on both sex and age, as the second dataset.

The third dataset was obtained from surgical patient data, of persons aged 60–89 years. This prospective observational study included patients who underwent TKR or LSCS at six hospitals in Japan. This study aimed to investigate the feasibility of using the locomotive syndrome risk test in patients with severe musculoskeletal diseases. In this study, we included 110 and 75 patients with TKR and LSCS, respectively, who had data on both sex and age. All participants provided written informed consent, and the study was approved by the authors' affiliated organizations and institutions.

Simulation framework

We performed two simulations to evaluate the statistical properties of frequently used methods to handle missing data under natural missing data patterns with artificial probabilities of participants who had missing data (simulation #1) and varying numbers of missing items of GLFS-25 (simulation #2). The simulation aims were to compare the statistical methods in terms of bias and variability and its relation to the number of participants, and to obtain practical guidelines on how many items of the GLSF-25 can be missing when we use these statistical methods.

Simulation #1

In each observational dataset, we created two datasets: the full dataset comprised patients who had data on both sex and age, and the complete dataset comprised patients who had complete data (no missing value in each dataset) on GLFS-25, sex and age. Namely, the sizes of the full data were 9 044, 743, and 185 for the community dwellers, outpatient and surgical patient datasets, respectively. Our simulation study involved the following five steps for each dataset.

Step 1. Extract the missing patterns using the full dataset.

Step 2. Calculate the true average value using the complete dataset.

Step 3. Artificially create missing data using the complete dataset.

Step 4. Apply each statistical method for the dataset created in step 3.

Step 5. Repeat steps 3 and 4.

To provide practical guidance on the choice of handling missing values for the GLFS-25, we reproduced common missing patterns. To reproduce common missing patterns to find item(s) often left missing, we created a cross-tabulation table of the number of missing items in each item. In this study, we used the five most observed missing patterns for the subsequent simulation study.

For step 2, we calculated the true average value of GLFS-25, and the true value was used to calculate the bias in each statistical method in step 4. In step 3, to create artificial missing data for random datasets, we used the algorithm proposed by van Buuren[9] and used by other scholars.[10–12] As covariates to form the structure of missingness, we considered sex and age, because they are important variables related to the missingness of GLFS-25 questionnaire.[8] Moreover, female sex and older age were related to having at least one missing item. Step 3 is detailed in the online supplemental appendix 1.

Simulation #2

This simulation has steps 2, 4 and 5, which are the same as in simulation #1. In this simulation, we did not use the

Table 2 Missing patterns of GLFS-25

Order	Community-dwellers dataset (n=9044)		Outpatient dataset (n=743)		TKR dataset (n=110)		LSCS dataset (n=75)	
	Missing pattern	N (%)	Missing pattern	N (%)	Missing pattern	N (%)	Missing pattern	N (%)
1	Item 21 missing	40 (0.4)	Item 1 missing	10 (1.3)	Items 23–25 missing	2 (1.8)	Item 1 missing	3 (3.7)
2	Item 15 missing	19 (0.2)	Item 21 missing	7 (0.9)	Several patterns	1 (0.9)	Item 15 missing	2 (2.4)
3	Items 18–25 missing	17 (0.2)	Item 2 missing	3 (0.4)			Several patterns	1 (1.2)
4	Item 1 missing	16 (0.2)	Item 10 missing	3 (0.4)				
5	Item 3 missing	13 (0.1)	Several patterns	2 (0.3)				
–	Any item missing	239 (2.6)	Any item missing	80 (6.6)	Any item missing	17 (15.3)	Any item missing	14 (17.1)

GLFS-25, 25-question Geriatric Locomotive Function Scale; LSCS, lumbar spinal canal stenosis; TKR, total knee replacement.

missing patterns observed in the full dataset; therefore, step 1 was omitted. Step 3 also differs from simulation #1. Briefly, in step 3 of this simulation, participants who had missing data were selected, and their items were randomly selected and missed. In this simulation, the proportion of participants who would have missing items was set at 40%. The number of missing items varied from 4 to 16 in increments of 4.

In each simulation, the statistical methods described in the following section were applied to the created dataset and the resulting estimated average value of the GLFS-25 was compared with the true value.

Comparison of statistical methods

In this study, we compared four frequently used statistical methods that handle missing values differently: complete case analysis, single imputation using subject average (SI1), single imputation using subject domain average (SI2), and multiple imputation using chained equations (MICE).

Complete case analysis was used to calculate the average value of GLFS-25 within patients without any missing items (ie, participants who did not fall within any missing pattern in steps 3–5). SI1 calculated the average value of the GLFS-25 after imputing the average value of the observed items of the participant into the missing items. SI2 calculated the average value of the GLFS-25 after imputing the average value of the observed items within the same domain of the participant into the missing items.[4]

MICE (a.k.a. fully conditional specification[13]) is one of the methods used to generate imputations in multiple imputation literature. The attractive feature of MICE is its ability to handle non-monotone missing data, as seen in the three observational datasets in this study. Another feature of MICE is its ability to handle different variable types (continuous, binary, unordered and ordered categorical) because of its flexibility in imputation models. Each imputation model was regressed on the age, sex and other items of the GLFS-25. We used the linear model for imputing models and imputed values using predictive mean matching, that is, imputed a value randomly from a set of observed values whose predicted values were closest to the predicted value for the missing value from the simulated regression model. Several software packages are available to use MICE: S-Plus, R, SPSS and SAS.[14] A tutorial paper of the MICE is available elsewhere.[15] Note that multiple imputation is expected to be unbiased under a missing random structure, as created in step 3, but this does not apply to the other methods.

We will briefly describe the missingness mechanism of missing completely at random (MCAR) and missing at random (MAR), where the latter mechanism is assumed in the simulation studies described above. Under MCAR, missing values are not correlated to other variables, while under MAR, missing values are correlated only with observed variables. For example, when the reason for missingness was unconscious, unintentionally skipping the item, the missing item should be caused by the MCAR mechanism, while when the missing depends on the demographic variable such as age, the missing item is caused by the MAR mechanism. All four methods, including complete case analysis, are theoretically unbiased under MCAR, while appropriately analysed multiple imputation is only unbiased under MAR. The validity of other methods depends on the correlation between variables, but they are generally biased and underestimate the variability of the parameter estimates.

The number of simulation repetitions was set to 1000, and the number of imputations and burn-in iterations before each imputation for the MICE were 50 and 20,

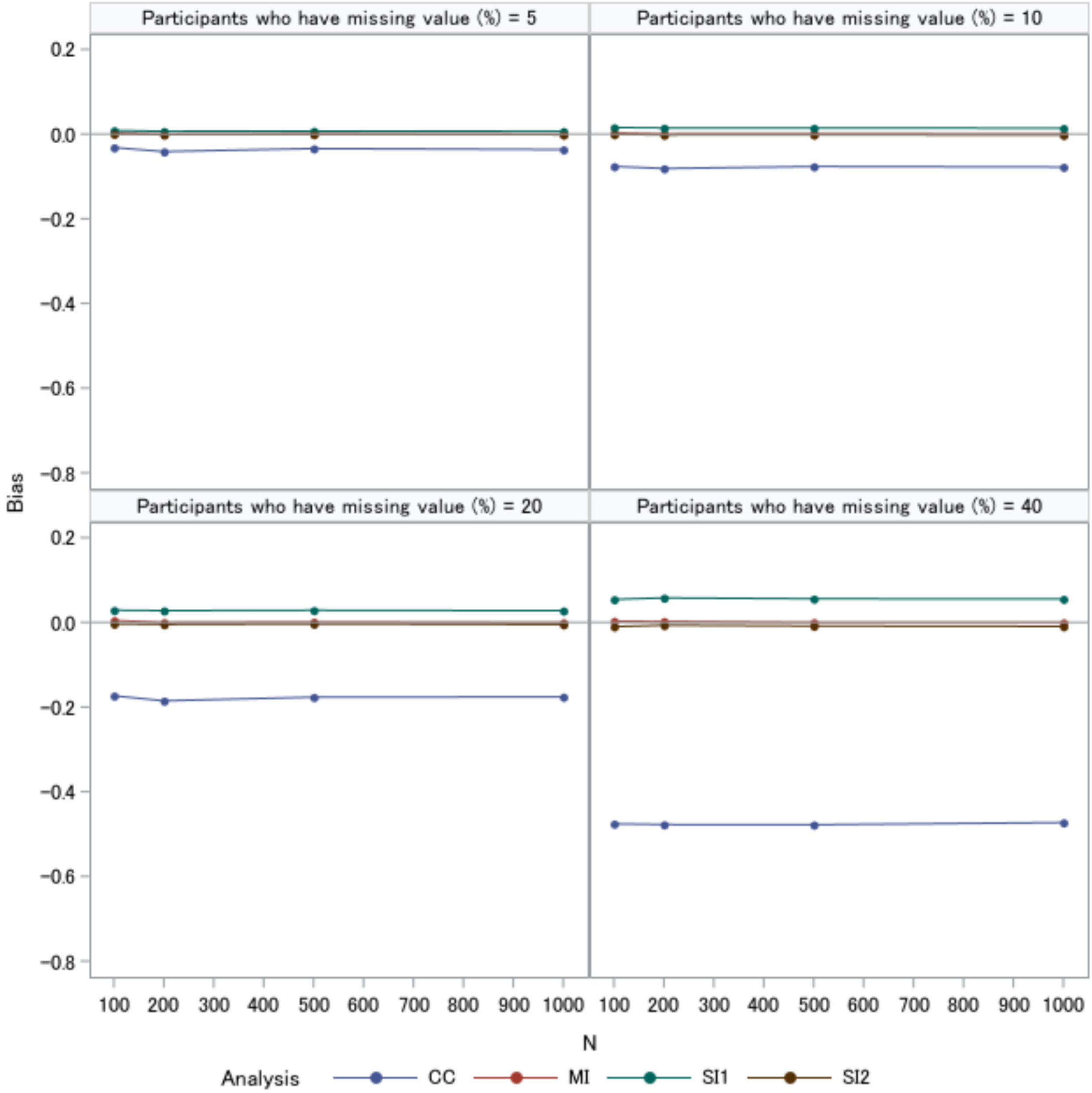

Figure 1 Comparison of the statistical methods for estimating bias in the simulation study using the community-dwellers dataset (simulation #1). CC, complete case; MI, multiple imputation; SI1, single imputation 1; SI2, single imputation 2.

respectively. We varied the marginal probabilities with missing items of 5%, 10%, 20% and 40%, but within each scenario, the relative frequency of missing patterns followed the actual percentage calculated in step 1. To evaluate the performance for various sample sizes, in steps 3–1, we varied the sample size to 100, 200 and 500 by sampling without replacement, except for surgical patient data, which had only approximately 100 participants each. We calculated the average bias and root mean squared error (RMSE) for each statistical method for each scenario. All statistical analyses were conducted using SAS and SAS/STAT V.9.4 software of the SAS System for Windows (SAS Institute). For MICE, we used the MI and MIANALYZE procedures, which perform multiple imputations under numerous settings. PROC MI implements popular methods for creating imputations under monotone and non-monotone (arbitrary) patterns of missing data, and PROC MIANALYZE analyses the results from multiple imputed datasets.

Patient and public involvement

Patients and public were not involved in the development of the research question, outcome measures, the design of this study and conduct of this study.

RESULTS

The participant characteristics are shown in table 1. The average score (SD) of the GLFS-25 in the complete dataset was 4.47 (5.89), 13.72 (11.76), 38.30 (15.74) and 39.00 (17.20) in the community dwellers, outpatients and surgical patients with TKR and LSCS, respectively.

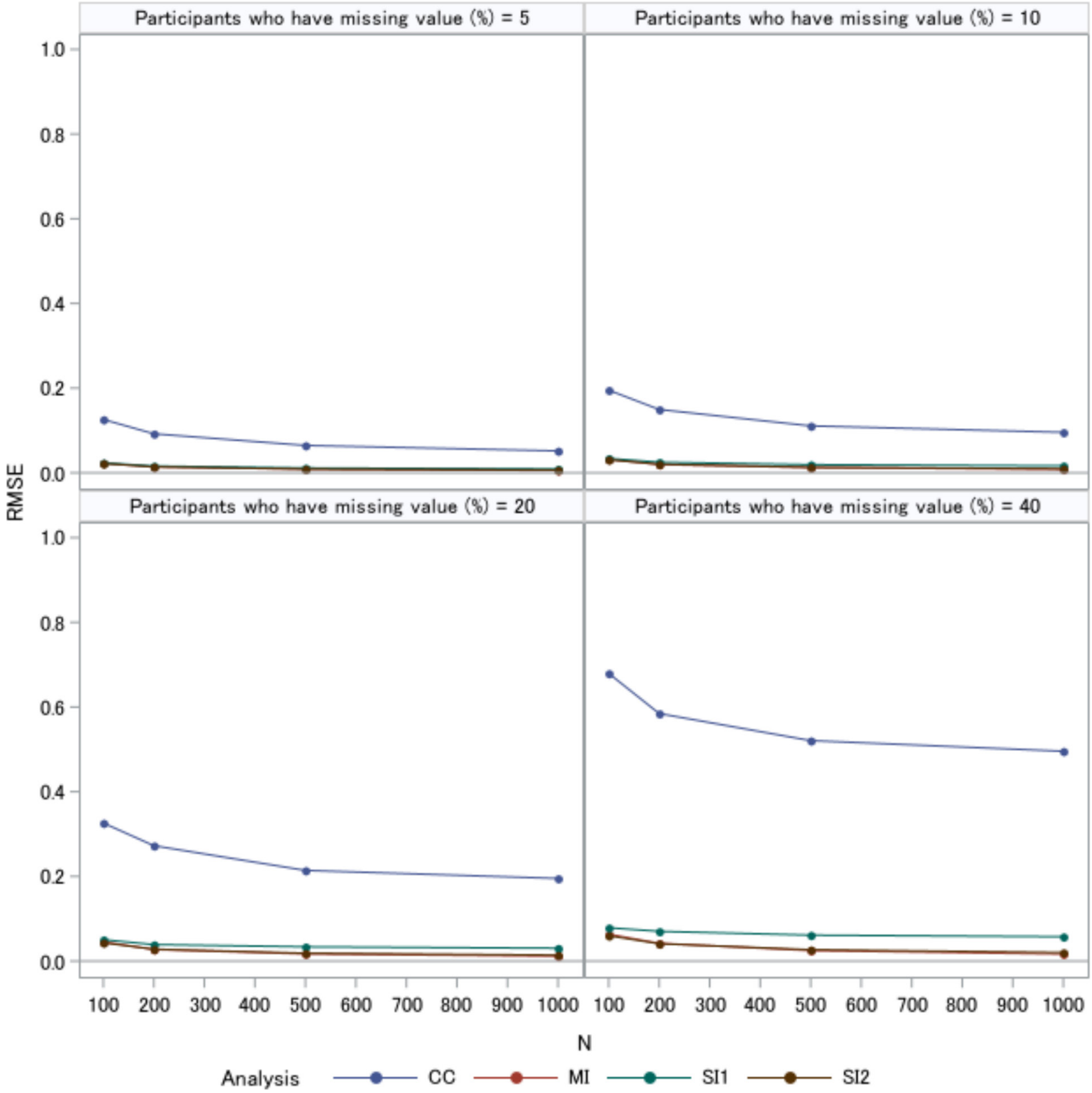

Figure 2 Comparison of the statistical methods for estimating the root mean squared error in the simulation study using the community-dwellers dataset (simulation #1). CC, complete case; MI, multiple imputation; RMSE, root mean squared error; SI1, single imputation 1; SI2, single imputation 2.

Therefore, the GLFS-25 scores increased in the following order: community dwellers, outpatients and surgical patients.

Table 2 lists the missing patterns observed in each dataset. Across the dataset, items 1, 15 and 21 tended to be missing. In the outpatient dataset, the four most frequently missing patterns were used because several were ordered as the fifth frequent pattern. In the surgical patient dataset (TKR and LSCS), we could not find common missing patterns; thus, for the simulation of the surgical patient dataset, we used the five most frequent missing patterns observed in the community-dwellers dataset, as the community-dwellers dataset was larger than that of the other dataset. Presence of the missing items depended on the sex and age; female and/or higher age individuals tended to report missing items (online supplemental tables 1,2).

Simulation #1

Community-dwellers dataset

Figures 1 and 2 show a comparison of statistical methods in the simulation study with varying missing probabilities using the community-dwellers dataset. For all missing probabilities, the bias curve was horizontal and the bias did not depend on the sample size. When the missing probability was 5%, the biases were negligible for each imputation method (single and multiple). As the missing probabilities increased, the bias in the complete case analysis stood out compared with the imputation methods. When the missing probability was 40%, bias was close to

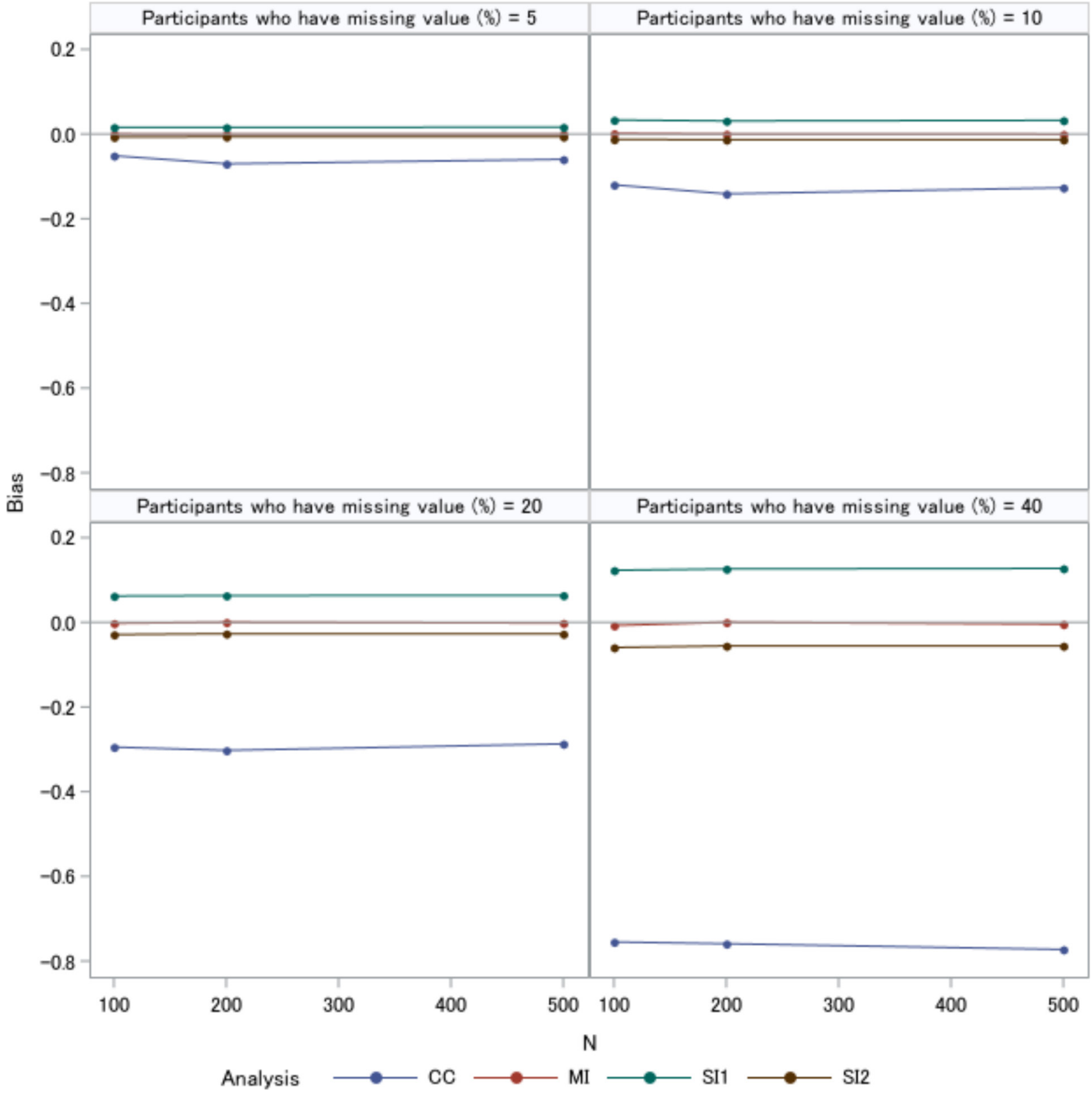

Figure 3 Comparison of the statistical methods for estimating bias in the simulation study using the outpatient study dataset (simulation #1). CC, complete case; MI, multiple imputation; SI1, single imputation 1; SI2, single imputation 2.

0.5. As expected, multiple imputations were unbiased. In contrast, other simple imputation methods showed a higher bias as the missing probability increased. However, the bias was much smaller than that in the complete case analysis.

Similar to bias, the RMSE of each statistical method (figure 2) increased as the missing probability increased. As expected, the RMSE decreased with an increase in sample size. All imputation methods showed a similar curve for RMSE, and the multiple imputation methods showed the lowest RMSE.

Outpatient dataset

Figures 3 and 4 show a comparison of statistical methods in the simulation study with varying missing probabilities using the outpatient dataset. The methods' relative characteristics were similar to those of the community-dwellers dataset; multiple imputations were unbiased, although other single imputation methods showed bias that was smaller than the complete case analysis. Compared with the simulation results from the community dwellers dataset, the bias was larger in the outpatient dataset; the bias in the complete dataset was as much as −0.3 and −0.8 when the missing probability was 20% and 40%, respectively.

Surgical patient dataset (TKR and LSCS)

Table 3 presents a comparison of the statistical methods in the simulation study using the surgical patient dataset. The sample sizes were the same as those of the original complete dataset (n=110 and 75 for TKR and LSCS, respectively). Because we found similar relative characteristics

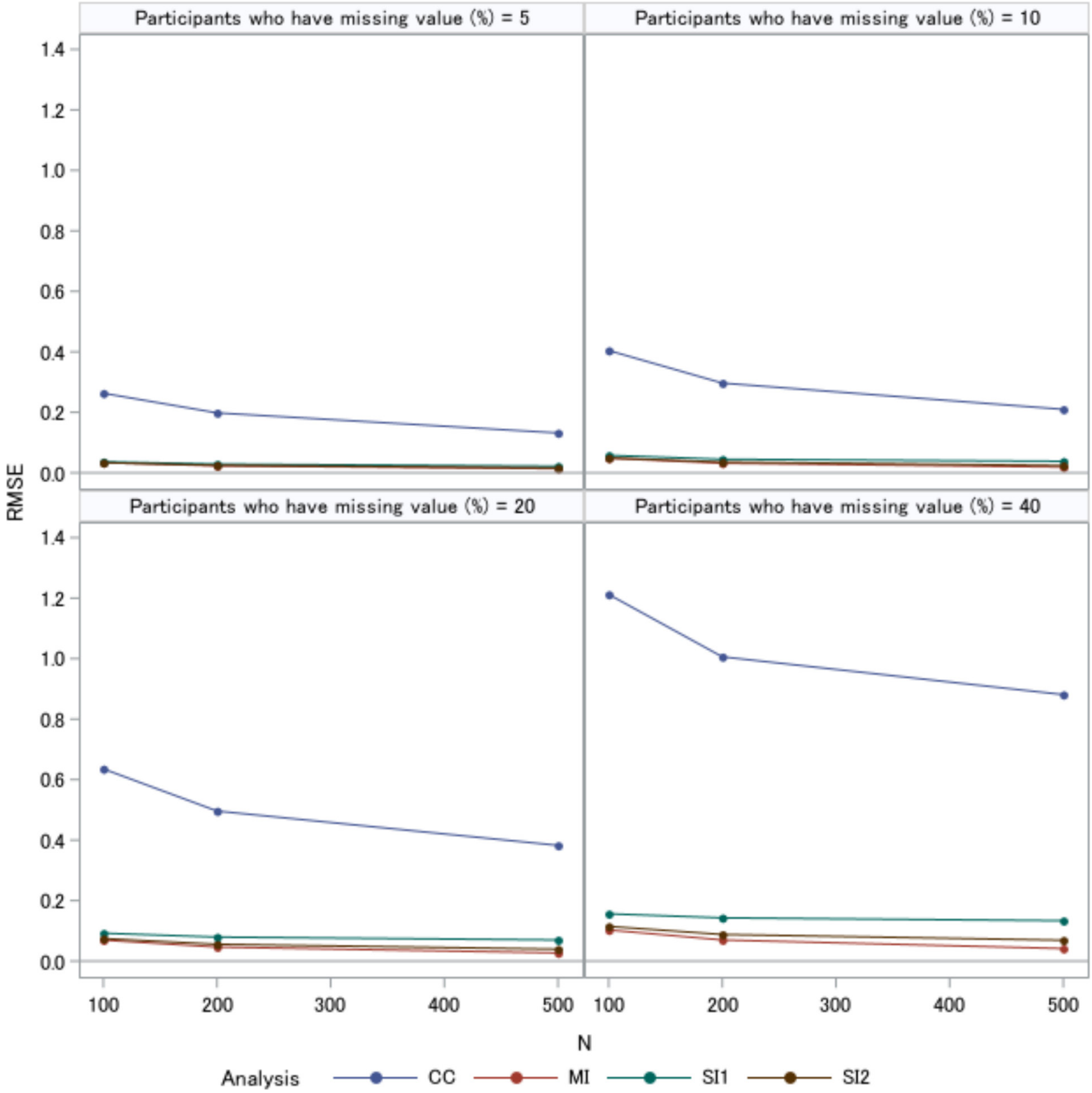

Figure 4 Comparison of the statistical methods for estimating the root mean squared error in the simulation study using the outpatient study dataset (simulation #1). CC, complete case; MI, multiple imputation; RMSE, root mean squared error; SI1, single imputation 1; SI2, single imputation 2.

of the methods, we fixed the missing probability at 20% for the surgical patient dataset.

Multiple imputation showed the least bias among the methods, but SI2 also showed low bias. Even SI1 exhibited a bias of approximately 0.2. Regarding the RMSE, there was no marked difference between multiple imputation and SI2.

Simulation #2

Figures 5 and 6 show a comparison of statistical methods in the simulation study with varying numbers of missing items using the outpatient and surgical (TKR) datasets. As the number of missing items increased, both the bias and RMSE of all imputation methods increased. In the simulation, SI 1 showed the least bias. Multiple imputation and SI 2 showed higher bias when the number of missing items exceed eight. However, when the number of missing items was about less than or equal to eight, their performances were comparable in practice. In terms of RMSE, multiple imputation was the best among these methods.

DISCUSSION

In this study, we evaluated the statistical performance of complete case analysis and common imputation strategies (single imputation and multiple imputation) via simulation with missing artificial missing data at random structures. We aimed to present practical guidance on the choice of handling missing values for the GLFS-25

Table 3 Simulation results for LSCS and TKR datasets

	Bias	RMSE
LSCS dataset		
Complete case	0.78	1.31
Multiple imputation	0.02	0.13
Single imputation 1	–0.23	0.27
Single imputation 2	–0.09	0.15
TKR dataset		
Complete case	–0.33	0.91
Multiple imputation	–0.01	0.10
Single imputation 1	–0.23	0.25
Single imputation 2	–0.05	0.11

LSCS, lumbar spinal canal stenosis; RMSE, root mean squared error; TKR, total knee replacement.

based on the characteristics of the dataset. Although the GLFS-25 scores differed among the three datasets (the community-dwelling, outpatient and surgical patient datasets), the relative performances were similar across the datasets. Multiple imputation was unbiased and had the lowest RMSE, complete case analysis showed the largest bias, and the performances of the single imputation were between those methods.

As practical guidance, we recommend multiple imputations to impute missing values of the GLFS-25 for the population average as a target parameter because the multiple imputation was theoretically unbiased when the models to impute missing values were correctly specified. Additionally, multiple impuatation could result in better performance than single imputation when the aim of the analysis is not the 'simple' population average, for example, a prognostic factor analysis. This is because multiple imputation method imputes predicted values based on the relationship among missing values and factors included in the analysis. The comparative performance among imputation techniques is beyond our scope, but we believe it is an interesting research area. The simulation study using actual datasets showed the unbiased estimates from multiple imputation, regardless of the missing probabilities. The drawbacks of multiple imputations include possible implementation difficulties. Although many software packages are currently available for multiple imputation,[15] some clinicians might find them complex to implement. We used SAS/STAT and PROC MI, which create imputations under monotone and non-monotone patterns of missing data, respectively. Because of their ability to handle various settings, users must specify several options. Moreover, when the user wants to consider additional restrictions, such as resampling, if the value falls outside the possible value (eg, less than zero), additional codes might be necessary. Without appropriate code management, the results of multiple imputations could not be reproduced.

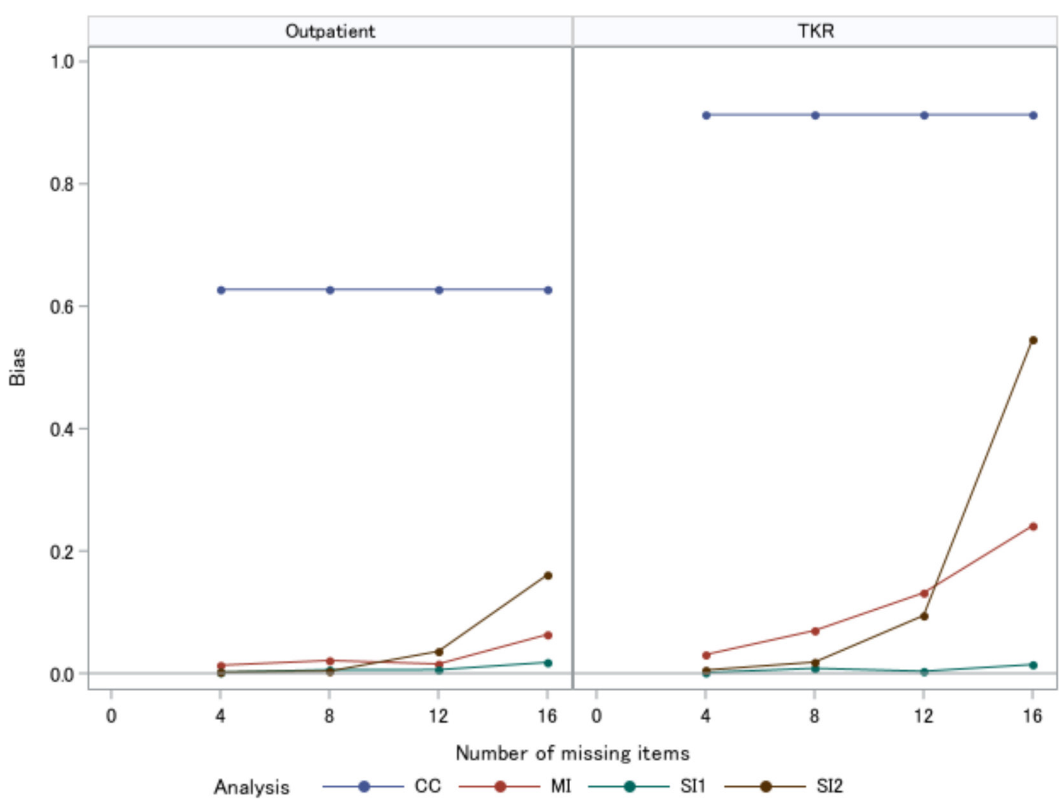

Figure 5 Comparison of statistical methods for estimating bias in the simulation study under varying assumptions regarding the number of missing items (simulation #2). CC, complete case; MI, multiple imputation; SI1, single imputation 1; SI2, single imputation 2; TKR, total knee replacement.

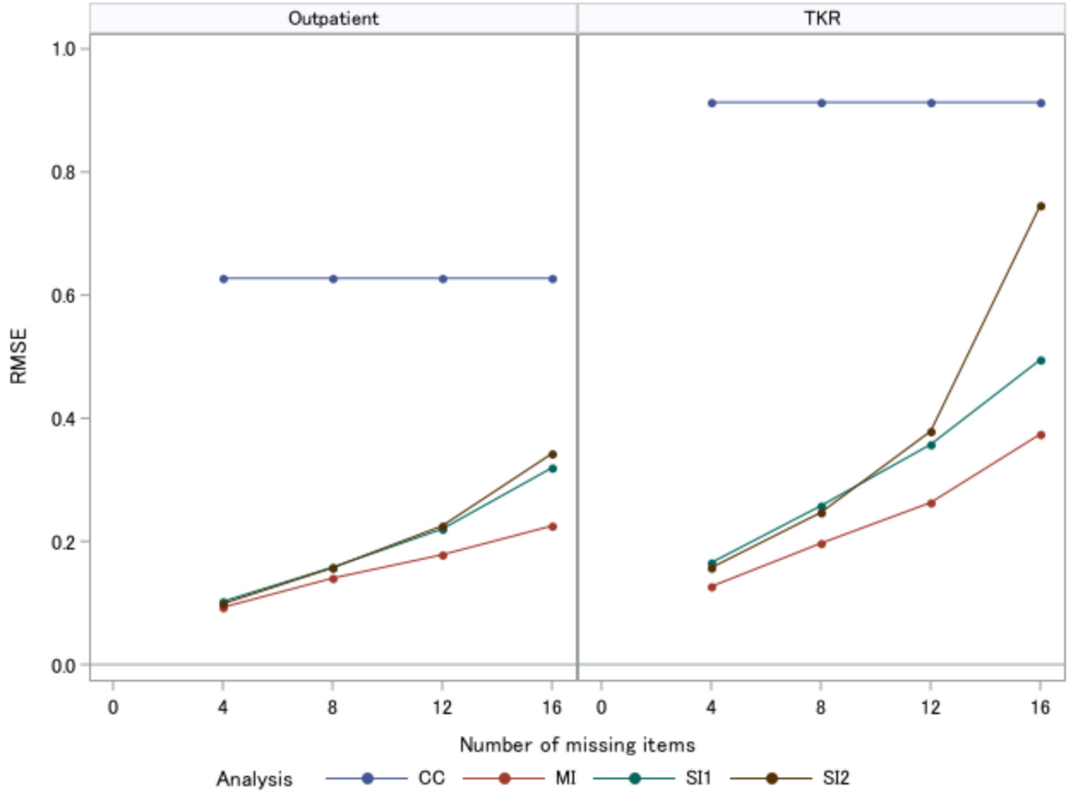

Figure 6 Comparison of statistical methods for estimating the root mean squared error in the simulation study under varying assumptions regarding the number of missing items (simulation #2). CC, complete case; MI, multiple imputation; out, orthopaedic outpatients; RMSE, root mean squared error; SI1, single imputation 1; SI2, single imputation 2; TKR, total knee replacement.

We consider that a single imputation might be an alternative to multiple imputations. The simulation study showed that the bias of single imputation in the community dwellers' study and outpatient datasets increased as the missing probability increased; however, the absolute biases were less than 0.1, even when the missing probabilities were as large as 40%. In a nationwide study of community dwellers, where many participants did not have apparent mobility problems, the median scores of the GLFS-25 were 1–4 in the age categories<75.[8] Comparing the absolute bias of single imputation to the observed median scores, we found that the relative bias ranged from 3% to 10%. Although the 10% bias is not small, the magnitude of the result should be small because the absolute values of GLFS-25 were small in those populations. In the outpatient datasets, where the median score of the GLFS-25 was 10, the relative bias of the single imputation was less than 0.1%. The relative biases in the surgical datasets were comparable to those in the other datasets. Considering these relative biases in all kinds of datasets and the simplicity of implementation, we consider that single imputation can be an alternative to multiple imputation, as long as the number of missing items is up to 8 out of 25 items in each individual of the population. Therefore, practically, one might include individual who completed at least 17 items, and exclude others.

Complete case analysis after exclusion of a large amount of missing data can cause bias if the missing values are not MCAR. However, MCAR is often not the case.[16] The most plausible explanation for the missing value in GLFS-25 is MAR, depending on the other observed data, since each item in GLFS-25 is correlated with one another, as each item representing one aspect of mobility decreases at some time.

The strength of this study is that we used three datasets for the simulation; the consistent results across the datasets provide a robust conclusion. Despite its strengths, this study has several limitations. First, the simulation study area was limited. For example, the missing patterns considered were limited to the five most frequent patterns in actual datasets. Therefore, we did not assess how single and multiple imputations behave when most items were missing (eg, when the number of missing items was larger than 8, the maximum number of missing items in table 1). Second, we only considered missing data in a random structure; that is, the missing probability depends only on the observed covariates (age and sex). Therefore, our results do not extend to other situations such as missing not at random or MCAR.[17] In the missing not-at-random structure, in general, none of the statistical methods in this paper provide an unbiased estimate, while all methods, including complete case analysis, are expected to provide an unbiased estimate in the MCAR structure. We cannot verify whether the observed missing data have a type of missing structure. However, the key takeaway is that investigators need to use their expert knowledge to decide on a set of covariates related to missing probabilities to minimise bias. Third and finally, our results are demonstrated in terms of the estimation of population mean of GLFS-25; thus, we cannot generalise them to each patient nor other analysis than calculating the population mean of GLFS-25. For example, we do not know whether single imputation or multiple imputation will work for the calculation of each individuals. However, as the bias of population mean is the average of each participant's bias, so we can say that the average of each participant's bias is small by the imputation methods.

In conclusion, the simulation study showed that multiple imputation is the best method to adjust for missing values in the GLFS-25. Simultaneously, single imputation methods are comparable under common missing patterns observed in real datasets and MAR structures. The results were consistent across a broad population, including community dwellers and surgical patients.

Author affiliations
[1]Clinical Research Promotion Center, The University of Tokyo Hospital, Tokyo, Japan
[2]Department of Sensory & Motor System Medicine, Faculty of Medicine, University of Tokyo, Tokyo, Japan
[3]Department of Liberal Arts, Faculty of healthcare and welfare, Saitama Prefectural University, Saitama, Japan
[4]Clinical and Translational Research Center, Niigata University Medical and Dental Hospital, Niigata, Japan
[5]Department of Rehabilitation Medicine, Faculty of Medicine, University of Tokyo, Tokyo, Japan
[6]Department of Orthopaedic Surgery, Graduate School of Medicine, Gunma University, Gunma, Japan
[7]Department of Bone and Joint Surgery, Ehime University, Ehime, Japan
[8]Department of Orthopaedic Surgery, Towa Hospital, Tokyo, Japan
[9]Department of Orthopaedic Surgery, NTT Medical Center Tokyo, Tokyo, Japan

Contributors Conceptualisation: TK and KY. Data curation: KY, TO, ST, HC, HM, KN and TO. Formal analysis: TK, RT and IT. Funding acquisition: ST, KN, TO. Methodology: TK, KY, RT and IT. Project administration: TK and KY. Software: TK, RT and IT. Supervision: TO. Visualisation: TK, RT, and IT. Writing – original draft: TK and KY. Writing - review & editing: RT, IT, TO, ST, HC, HM, KN and TO. Guarantor: KY

Funding This study was supported by JOA-Subsidized Science Project Research (grant 2015- 1) and "Locomo Chanllenge!" Promotion Council, Tokyo, Japan (grant N/A). The Locomo Challenge! Council is a non-profit organization launched by the Japanese Orthopaedic Association, to educate the public about the importance of maintaining adequate mobility in their lives in a super-aging Japanese society.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study was approved by the Ethics Committee of the Japanese Orthopaedic Association (06242017). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

ORCID iDs
Keiko Yamada http://orcid.org/0000-0003-0035-8750
Sakae Tanaka http://orcid.org/0000-0001-9210-9414

REFERENCES

1 Nakamura K, Ogata T. Locomotive syndrome: definition and management. *Clin Rev Bone Miner Metab* 2016;14:56–67.
2 Japanese Orthopaedic Association. Locomotive syndrome, 2015. Available: https://locomo-joa.jp/assets/pdf/index_english.pdf [Accessed 14 Feb 2022].
3 Giovannini S, Coraci D, Brau F, *et al*. Neuropathic pain in the elderly. *Diagnostics* 2021;11:613.
4 Seichi A, Hoshino Y, Doi T, *et al*. Development of a screening tool for risk of locomotive syndrome in the elderly: the 25-question geriatric locomotive function scale. *J Orthop Sci* 2012;17:163–72.
5 Endo K, Saito Y, Tamura A. Physical performance disorder (locomotive syndrome) closely relates to both hypertension and diabetes mellitus which are components of metabolic syndrome. *Shikoku Acta Medica [in Japanese]* 2018;74:107–12.
6 Tanabe Y, Suehara Y, Kim Y, *et al*. The Development of the Short-form of "25-Question Geriatric Locomotive Function Scale". *J Adv Med Med Res* 2018;25:1–13.
7 Little R, Rubin D. *Statistical analysis with missing data*. 3rd edn. Wiley, 2019.
8 Yamada K, Ito YM, Akagi M, *et al*. Reference values for the locomotive syndrome risk test quantifying mobility of 8681 adults aged 20-89 years: a cross-sectional nationwide study in Japan. *J Orthop Sci* 2020;25:1084–92.
9 Van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, *et al*. Fully conditional specification in multivariate imputation. *J Stat Comput Simul* 2006;76:1049–64.
10 Yu L-M, Burton A, Rivero-Arias O. Evaluation of software for multiple imputation of semi-continuous data. *Stat Methods Med Res* 2007;16:243–58.
11 Simons CL, Rivero-Arias O, Yu L-M, *et al*. Multiple imputation to deal with missing EQ-5D-3L data: should we impute individual domains or the actual index? *Qual Life Res* 2015;24:805–15.
12 Rombach I, Gray AM, Jenkinson C, *et al*. Multiple imputation for patient reported outcome measures in randomised controlled trials: advantages and disadvantages of imputing at the item, subscale or composite score level. *BMC Med Res Methodol* 2018;18:87.
13 van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007;16:219–42.
14 Azur MJ, Stuart EA, Frangakis C, *et al*. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011;20:40–9.
15 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;30:377–99.
16 Joseph L, Bélisle P, Tamim H, *et al*. Selection bias found in interpreting analyses with missing data for the prehospital index for trauma. *J Clin Epidemiol* 2004;57:147–53.
17 Hughes RA, Heron J, Sterne JAC, *et al*. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epidemiol* 2019;48:1294–304.