Multiple Imputation to Salvage Partial Respondents: Analysis of the Forgotten Joint Score-12 After Total Hip Arthroplasty

山手, 智志

https://hdl.handle.net/2324/7329438

出版情報：Kyushu University, 2024, 博士（医学）, 課程博士
バージョン：

Multiple Imputation to Salvage Partial Respondents

Analysis of the Forgotten Joint Score-12 After Total Hip Arthroplasty

Satoshi Yamate, MD, Satoshi Hamai, MD, PhD, Shinya Kawahara, MD, PhD, Daisuke Hara, MD, PhD, Goro Motomura, MD, PhD, Satoshi Ikemura, MD, PhD, Masanori Fujii, MD, PhD, Taishi Sato, MD, PhD, Satoru Harada, MD, Tetsunari Harada, MD, Yasuhiko Kokubu, MD, and Yasuharu Nakashima, MD, PhD

Investigation performed at the Department of Orthopaedic Surgery, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

Background: Missing responses are common when Asian patients complete the Forgotten Joint Score-12 (FJS-12), which is widely used to evaluate total hip arthroplasty (THA). We aimed to provide orthopaedic researchers with a solution for handling missing values in such patient-reported outcome measures (PROMs).

Methods: Patients who had undergone primary THA between 1998 and 2016 (n = 1,021) were investigated in 2020. The FJS-12 and 9 other PROMs, including questions related to Asian lifestyle activities, were administered. Risk factors for missing FJS-12 items were investigated. Partial respondents were matched with complete respondents; then, in each pair, the items not completed by the partial respondent were deleted from the responses of the complete respondent. Predictive mean matching (PMM) was performed in an attempt to recover the deleted items, using 65 sets of imputation models. After the missing values had been imputed, we explored patient characteristics that affected the FJS-12, using data from all complete and partial respondents.

Results: A total of 652 patients responded to the survey (393 complete and 193 partial respondents). Partial respondents were older, more often female, and less active. Older respondents were more likely to skip items involving the bed, while those who reported a better ability to sit in the seiza style (traditional Japanese floor sitting) were more likely to skip items about chair sitting. The imputed FJS-12 value exhibited excellent reliability (intraclass correlation coefficient for agreement with the true scores, 0.985). FJS-12 values of complete respondents were significantly higher than those of respondents with 4 to 11 missing items (51.6 versus 32.8, $p < 0.001$). Older age was associated with higher FJS-12 values, which was revealed only via analysis of the multiply imputed data sets ($p < 0.001$).

Conclusions: Analysis of only complete FJS-12 responses after THA resulted in a nonresponse bias, preferentially excluding older, female, and less active individuals and those with a traditional floor living style. Multiple imputation could provide a solution to scoring and analyzing PROMs with missing responses by permitting the inclusion of partial respondents.

Level of Evidence: Therapeutic Level III. See Instructions for Authors for a complete description of levels of evidence.

In recent years, there has been an increase in the number of studies of total hip arthroplasty (THA) using patient-reported outcome measures (PROMs)[1]. However, PROMs are susceptible to missing data[2]. Most studies have analyzed only patients with complete responses[2], with the unrealistic assumption that the data were missing completely at random (MCAR)[3-5].

The Forgotten Joint Score-12 (FJS-12)[6], a PROM with a wide score distribution and small ceiling effect, has been translated into several languages and validated in countries predominantly speaking those languages[6-11]. Because of its score characteristics, it enables the evaluation of a broad range of clinical outcomes as well as the identification of differences among those with high scores after THA; the latter is useful in efforts to further improve surgical techniques and implants[12]. However, since the FJS-12 was developed for Western populations, there is a potential risk of missing responses for certain items among non-Western populations with traditional lifestyles, such as Japanese who sit on the floor or sleep on tatami mats.

Multiple imputation[13] is the preferred method to handle missing data according to the guidelines of both the European Medicines Agency[14] and a U.S. National Research Council expert panel[15,16] requested by the Food and Drug Administration. This

Disclosure: The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (http://links.lww.com/JBJS/H287).

method can incorporate auxiliary information about missing data into the analysis, by generating multiple complete data sets with different imputed values for missing variables to incorporate any uncertainty[15]. Several models have been proposed to obtain multiple complete data sets[17-19]; one of the most popular is predictive mean matching (PMM)[2,19-21]. Although analysis using multiple imputation is recommended for data with missing values[13-16], analyses of FJS-12 results after THA have not been performed with this method, to our knowledge.

This study aimed to provide a solution for orthopaedic researchers with concerns about scoring and analyzing incomplete PROM data. We have addressed this issue by focusing on the FJS-12, which has been increasingly reported in recent years and is known to be susceptible to missing items[8,22]. We also focused on

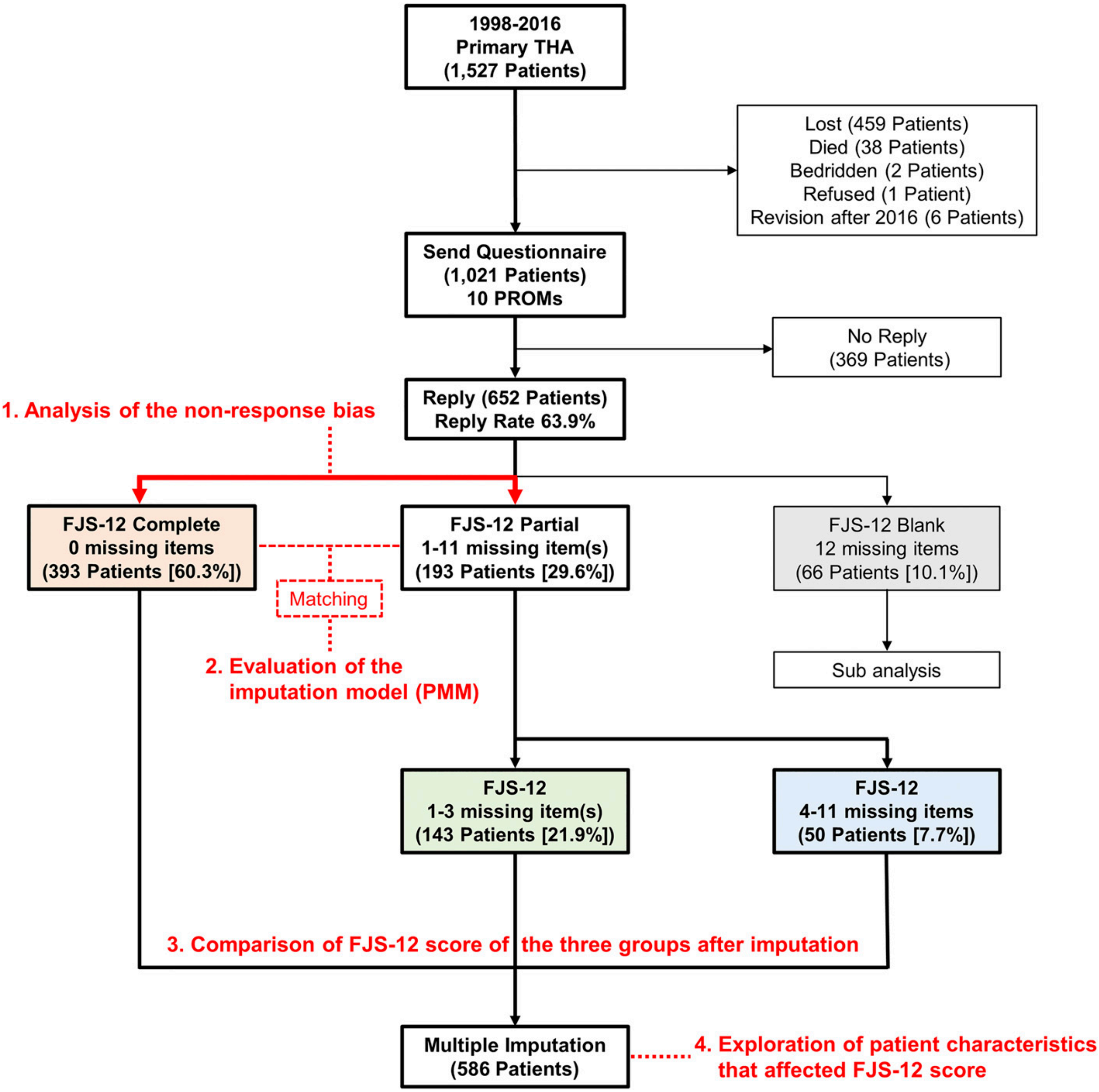

Fig. 1

Flow diagram of patient selection for the analysis. The 4 outcomes described in the Materials and Methods section are shown in red, indicating the populations in which they were analyzed.

TABLE I Patient Demographics*

	Complete FJS-12 (N = 393)	Partial FJS-12 (N = 193)	P Value	Complete and Partial FJS-12 (N = 586)
Age at survey† *(yr)*	70.5 ± 9.2 (41.7-90.7)	73.9 ± 9.9 (35.7-95.7)	<0.001‡	71.6 ± 9.6 (35.7-95.7)
Follow-up† *(yr)*	9.6 ± 4.2 (3.7-21.5)	9.9 ± 4.3 (3.7-22.3)	0.549	9.7 ± 4.3 (3.7-22.3)
Sex *(no. [%])*			0.027‡	
Male	77 (19.6)	23 (11.9)		100 (17.1)
Female	316 (80.4)	170 (88.1)		486 (82.9)
BMI† *(kg/m²)*	23.0 ± 3.3	22.9 ± 3.5	0.729	23.0 ± 3.4
Surgical side *(no. [%])*			0.565	
Unilateral	229 (58.3)	118 (61.1)		347 (59.2)
Bilateral	164 (41.7)	75 (38.9)		239 (40.8)
Diagnosis *(no. [%])*			0.150	
OA	350 (89.1)	167 (86.5)		517 (88.2)
ONFH	33 (8.4)	14 (7.3)		47 (8.0)
RA	3 (0.8)	5 (2.6)		8 (1.4)
Other§	7 (1.8)	7 (3.6)		14 (2.4)

*FJS-12 = Forgotten Joint Score-12, BMI = body mass index, OA = osteoarthritis, ONFH = osteonecrosis of the femoral head, RA = rheumatoid arthritis. †Values are given as the mean and standard deviation, with or without the range in parentheses. ‡Significant at $p < 0.05$. §Fracture, arthrodesis, spondyloepiphyseal dysplasia, posttraumatic osteonecrosis, and rapidly destructive coxopathy.

THA patients at medium to long-term follow-up, who were considered highly susceptible to having missing items[14].

The research questions were as follows: (1) Does exclusion of partial FJS-12 respondents after THA introduce a nonresponse bias? (2) How well does multiple imputation by PMM perform in imputing data for patients with missing FJS-12 items? (3) Do FJS-12 results after imputation differ according to the number of missing items? (4) Are there any patient characteristics that could affect FJS-12 results using multiply imputed data sets? We hypothesized that exclusion of partial FJS-12 respondents would introduce a nonresponse bias and that analysis of scores calculated from data sets imputed using PMM could provide less biased results.

Materials and Methods

Participants

This retrospective cohort study was conducted according to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement[23]. Our institutions approved the human protocol for this investigation, which was conducted in conformity with ethical principles of research. Ethical approval for this study was obtained from the Graduate School of Medical Sciences, Kyushu University (approval number 30-246, 21142-00).

The data for 1,527 Asian patients who underwent primary cementless THA for symptomatic osteoarthritis between January 1998 and December 2016 at a single university hospital were reviewed. This period was chosen to provide >1,000 cases with a minimum follow-up of 3 years after THA. All surgical procedures were performed via the posterolateral approach by the same group of 13 senior surgeons. Of these patients, 1,021 met the following inclusion criteria: no reoperation after THA, a valid address, prior agreement to participate in the survey obtained by telephone, and absence of any condition unrelated to the hip that might lead to bed rest. The mail survey was conducted in September 2020 among the 1,021 eligible patients. All participants provided informed consent to participate in the study, and the data were handled according to the ethical standards of the Declaration of Helsinki.

PROMs

The FJS-12[6] is a 12-item questionnaire assessing joint awareness in daily life after arthroplasty. The answers to the 12 items regarding the frequency of joint awareness (never, 0 points; almost never, 1 point; seldom, 2 points; sometimes, 3 points; mostly, 4 points) are summed and converted to a 100-point scale, with higher scores indicating better outcomes. The FJS-12 can be scored using the original rule[6] if no more than 3 items are missing. The University of California Los Angeles (UCLA) activity score[24] allows patients to select 1 answer from among 10 statements describing their current activity state, with higher scores indicating higher activity. The Oxford Hip Score (OHS)[25] is a 12-item disease-specific quality-of-life PROM that assesses function and pain in patients undergoing THA. The total OHS can range from 0 to 48, with higher scores indicating better outcomes. The visual analogue scale (VAS) satisfaction[26] score can range from 0 to 100, with 0 indicating that the patient is "not at all satisfied with the surgery" and 100 indicating that they are "very satisfied with the surgery." The patient satisfaction scale[27] uses a 4-point Likert scale: very dissatisfied, 25 points; somewhat dissatisfied, 50 points; somewhat satisfied, 75 points; and very satisfied, 100 points. Finally, 3 original questions in this study investigated the ease of squatting, sitting in the seiza style

TABLE II Multivariable Logistic Regression Analysis of Forgotten Joint Score-12 Item Missingness*

Question	Awareness of the Artificial Joint	Missing Items *(no. [%])*	Factors Associated with Missingness	Adjusted Odds Ratio (95% CI)	P Value
1	"In bed at night"	33 (5.6)	Age at survey	1.07 (1.01-1.15)	0.038†
			Unilateral THA	2.65 (0.78-12.1)	0.148
2	"When you are sitting on a chair for more than 1 hour"	30 (5.1)	Ease of sitting in the seiza style	2.15 (1.17-5.52)	0.039†
3	"When you are walking for more than fifteen minutes"	27 (4.6)	Age at survey	1.07 (0.99-1.15)	0.077
			Unilateral THA	3.59 (0.90-23.9)	0.108
4	"When you are taking a bath/shower"	33 (5.6)	Age at survey	1.05 (0.99-1.12)	0.152
			Unilateral THA	2.41 (0.74-9.59)	0.166
			OHS	0.94 (0.88-1.01)	0.070
5	"When you are traveling in a car"	44 (7.5)	Follow-up year	1.09 (0.98-1.22)	0.117
			Unilateral THA	2.86 (0.99-9.35)	0.060
			UCLA activity score	0.80 (0.57-1.08)	0.173
			OHS	0.95 (0.89-1.01)	0.083
6	"When you are climbing stairs"	31 (5.3)	OHS	0.90 (0.83-0.98)	0.013†
			Age at survey	1.06 (0.99-1.15)	0.130
			Ease of sitting in the seiza style	1.66 (0.92-3.37)	0.117
7	"When you are walking on uneven ground"	33 (5.6)	OHS	0.91 (0.85-0.98)	0.009†
			Age at survey	1.06 (0.98-1.16)	0.164
8	"When you are standing up from a low-sitting position	29 (4.9)	OHS	0.90 (0.81-0.99)	0.049†
			Age at survey	1.10 (0.98-1.24)	0.121
			VAS satisfaction	1.15 (0.99-1.44)	0.101
			UCLA activity score	0.51 (0.24-1.01)	0.052
			Ease of squatting	0.55 (0.24-1.23)	0.139
			Ease of sitting in the seiza style	1.86 (0.86-4.26)	0.119
9	"When you are standing for long periods of time"	29 (4.9)	OHS	0.93 (0.86-0.99)	0.040†
			Age at survey	1.05 (0.99-1.13)	0.143
			Unilateral THA	3.48 (0.96-17.3)	0.081
10	"When you are doing housework or gardening"	33 (5.6)	OHS	0.90 (0.83-0.98)	0.013†
			Unilateral THA	3.89 (0.97-21.3)	0.075
			Satisfaction with pain	1.06 (0.99-1.19)	0.146
			UCLA activity score	0.71 (0.43-1.06)	0.128
11	"When you are taking a walk/hiking"	64 (10.9)	Age at survey	1.07 (1.01-1.13)	0.016†
			VAS satisfaction	1.03 (0.99-1.08)	0.138
			UCLA activity score	0.77 (0.56-1.02)	0.088
			OHS	0.95 (0.90-1.01)	0.079
12	"When you are doing your favorite sport"	173 (29.5)	UCLA activity score	0.79 (0.66-0.94)	0.012†
			Follow-up year	1.07 (0.99-1.14)	0.057
			VAS satisfaction	1.01 (0.99-1.03)	0.164
			Ease of squatting	0.84 (0.67-1.07)	0.152

*Procedure for the model selection: 10 variables (age at survey, follow-up year after total hip arthroplasty [THA], sex, body mass index, surgical side, visual analogue scale [VAS] for satisfaction, satisfaction with pain, University of California Los Angeles [UCLA] activity score, Oxford Hip Score [OHS], ease of squatting) were respectively fitted as the initial dependent variables, except for the following adjustments. "Ease of squatting" was not used for question 1 because it was not relevant. "Ease of sitting in the seiza style (Japanese traditional sitting style)" was added for questions 4, 8, and 10, and "Ease of getting in and out of a car" was added for question 5, because they could be related to the questions. Finally, the variables that had an effect were selected in the forward and backward stepwise selection process to minimize the Akaike information criterion. CI = confidence interval. †Significant at $p < 0.05$.

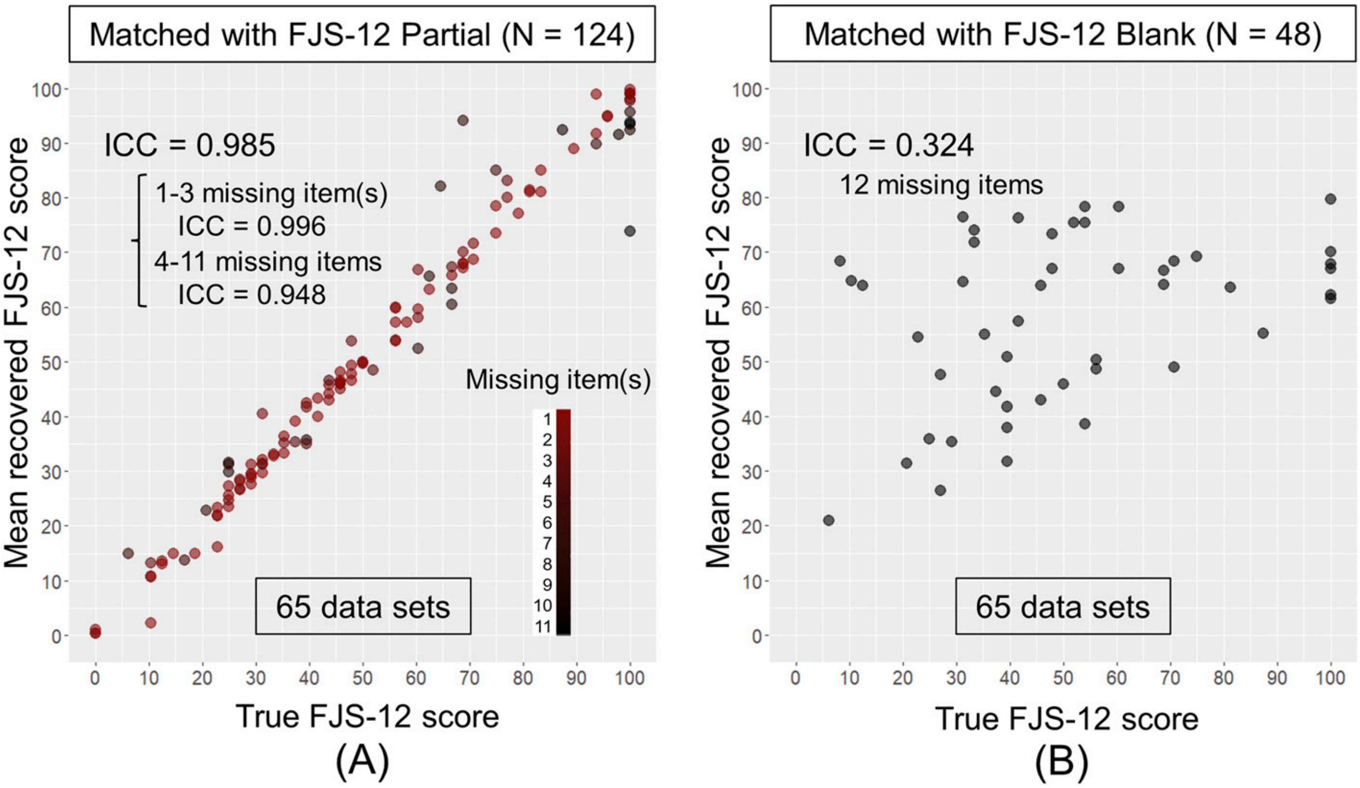

Fig. 2

Intraclass correlation coefficient (ICC) for agreement between the true Forgotten Joint Score (FJS)-12 and the mean FJS-12 calculated from 65 data sets after imputing the missing items using predictive mean matching (mean recovered FJS-12). **Fig. 2-A** Of the 193 partial respondents with 1 to 11 missing items, 124 (64.2%) could be matched. The reliability of the mean recovered FJS-12 in the corresponding full respondents was excellent (ICC for agreement between the recovered and true FJS-12 = 0.985). **Fig. 2-B** Of the 66 blank respondents (all 12 items were missing), 48 (72.7%) could be matched. The reliability of the mean recovered FJS-12 was poor (ICC for agreement between the recovered and true FJS-12 = 0.324), indicating that if none of the 12 FJS-12 items were available, PMM could not recover the true value of the individual's FJS-12.

(traditional Japanese floor sitting), and getting in and out of a car. These had 4 options: no, "have not tried" (1 point); no, "impossible" (2 points); yes, "possible with some support" (3 points); and yes, "easily possible" (4 points). These 3 movements were considered to have a considerable impact on joint awareness in daily life, particularly in Asia[28-30], and they were therefore included in the analysis of factors that might have an effect on the FJS-12.

Nonresponse Bias with Selection of Only Complete FJS-12 Respondents

FJS-12 respondents were divided into 3 groups: those with responses to all 12 questions (complete respondents), those with completion of only some questions (partial respondents), and those with all 12 questions unanswered (blank respondents). First, FJS-12 complete and partial respondents were analyzed. For each item, multivariable[31] analysis with a binary outcome variable of missing or not missing was performed to identify risk factors for the absence of a response to that item. These results were used to determine whether the pattern of missing responses was MCAR or not.

Reliability of the FJS-12 Value Calculated Using Imputation of Missing Items with PMM

PMM[2,19-21] was performed with the mice package[32] in R (The R Foundation for Statistical Computing), which is widely used for analyses involving PMM. In PMM, imputation of the missing values is performed by randomly selecting from the values observed in other individuals who are similar, as defined using the distance in a normal-theory linear regression model.

The following analysis was conducted to evaluate the reliability of the recovered FJS-12 value calculated using data sets in which missing item(s) were imputed using PMM. Complete and partial respondents were matched in a 1:1 ratio based on age at the time of the survey (±2 years), follow-up duration (±1 year), surgical side, and sex.

For each partial respondent, the pattern of missing responses was replicated by deleting the responses of the matched complete respondent to the same items, and multiply imputed data sets were created using PMM. We compared the mean recovered FJS-12 of each complete respondent with their true FJS-12 and calculated the intraclass correlation coefficient (ICC) for agreement, specifically the ICC(2,1)[33,34].

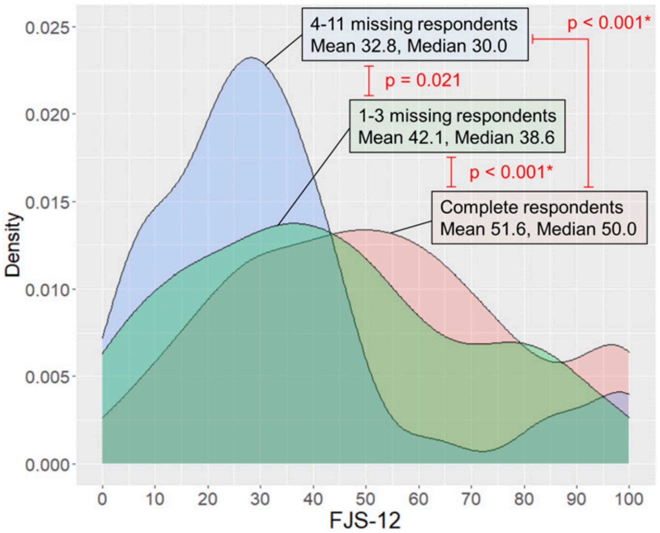

Fig. 3

FJS-12 distribution in the study sample, estimated using kernel densities, according to the number of missing items. The area of each histogram was standardized to 1. The mean FJS-12 was 51.6 ± 26.7 for complete respondents, 42.1 ± 26.9 for those with 1 to 3 missing responses (calculated with the original scoring rule), and 32.8 ± 24.5 for those with 4 to 11 missing responses (calculated from multiply imputed data sets using predictive mean matching). *The complete respondents differed significantly from those missing 1 to 3 items and those missing 4 to 11 items ($p < 0.001$ for both) according to the Bonferroni method.

Comparison of FJS-12 Groups by the Number of Missing Items

Partial FJS-12 respondents were divided into 2 groups: those with 1 to 3 missing items and those with 4 to 11 missing items; the original scoring rule calls for the exclusion of the latter from analyses[6]. FJS-12 respondents with up to 3 missing items were scored according to the survey instructions.

In addition, multiply imputed data sets were created using PMM for respondents with missing items, and the mean FJS-12 was obtained. These scores were used to assess the ability of PMM to approximate the true FJS-12 in participants with 1 to 3 and 4 to 11 missing responses.

Patient Characteristics That Affected FJS-12 Values Calculated from Imputed Data Sets Using PMM

We explored patient characteristics that affected the FJS-12 using data from all complete and partial respondents after imputing missing values. Multivariable[31] regression analysis was performed on each of the 65 data sets that had been completed using imputation. The results were combined using the Rubin rules[13] to identify patient characteristics that contributed to the FJS score.

Statistical Analysis

All statistical analyses were conducted using R version 4.0.5 and RStudio version 1.4.1106. The chi-square test was used to analyze categorical variables, and the Student t test was used for continuous variables. The Kruskal-Wallis test was used to compare the FJS-12 among the 3 groups, with significance adjusted for multiple comparisons by the Bonferroni method. For each FJS-12 item, a multivariable logistic regression analysis was conducted with a binary outcome variable indicating whether the response for that item was missing or not missing. Multiple imputation was conducted in the mice 3.13.0 package[32] using PMM. A total of 37 variables (see Appendix Supplementary data 1) were applied in the imputation model[35,36]. Multiple complete data sets were generated by repeating the imputation process the same number of times as the number of respondents who were missing at least one of the variables used in the multivariable analysis[18,36]. Multivariable regression analysis of FJS-12 values (treated as a continuous variable) was separately applied to each imputed data set. These estimates and their standard errors were combined using the Rubin rules[13]. The variables for the multivariable analyses were selected using forward and backward stepwise selection to minimize the Akaike information criterion[37]. $P < 0.05$ was defined as significant.

Source of Funding

No external funding was received for any aspect of this work.

Results

Of the 652 respondents, 393 (60.3%) completed all 12 FJS-12 items, 193 (29.6%) had 1 to 11 missing item(s), and 66 (10.1%) had 12 missing items (Fig. 1; see also Appendix Supplementary data 2, 3, and 4). Among the partial respondents, the median number of missing items was 1, and 50 patients (7.7% of the total) had 4 to 11 missing items. The mean follow-up duration was 9.6 years for complete respondents and 9.9 years for partial respondents, with no significant difference ($p = 0.549$). The partial respondents were significantly older than the complete respondents at the time of the survey ($p < 0.001$) and more likely to be female ($p = 0.027$) (Table I). Partial respondents had significantly lower UCLA activity scores ($p = 0.002$) and more unfavorable scores for each FJS-12 item (see Appendix Supplementary data 5 and 6).

Older age at the time of the survey was a risk factor for skipping FJS-12 item 1 (related to joint awareness in bed at night) and item 11 (related to walking/hiking). High scorers for sitting in the seiza style were more likely to skip FJS-12 item 2 (related to sitting in a chair), while skipping items 6 to 10 (which are related to moving states) was associated with lower OHS (function and pain) scores. High UCLA activity scores were associated with lower rates of missing values for the sport-related FJS-12 item 12 (Table II). Thus, we concluded that the missing FJS-12 items did not follow the MCAR pattern and that the partial respondents should be included in the analysis using multiple imputation.

Of the 193 partial respondents, 124 (64.2%) were matched (see Appendix Supplementary data 7). The ICC for agreement between the true and PMM-recovered FJS-12, using 65 data sets, exhibited excellent reliability (ICC for agreement = 0.985, Fig. 2). However, the same method indicated poor reliability for the estimated scores for the blank-response group (ICC for agreement = 0.324). Therefore, blank respondents were not included in the subsequent analyses. Scoring reliability for FJS-12 respondents with up to 9 missing items was excellent (ICC for agreement = 0.939 to 0.993; see Appendix Supplementary data 8 and 9).

TABLE III Multivariable Regression Analyses of the Forgotten Joint Score (FJS)-12 Value*

Factors	Complete and All Partial Respondents (N = 586)†		Complete Respondents and Those with 1-3 Missing Responses (N = 261)‡		Complete Respondents Only (N = 203)§	
	β (95% CI)	P Value	β (95% CI)	P Value	β (95% CI)	P Value
Age at survey	0.37 (0.17 to 0.57)	<0.001**	0.27 (−0.03 to 0.57)	0.077		
Follow-up, in years					1.15 (0.31 to 1.98)	0.007**
Male (reference: female)					5.24 (−2.21 to 12.69)	0.167
BMI						
Unilateral THA (reference: bilateral)	−1.08 (−4.95 to 2.79)	0.583	−3.86 (−9.24 to 1.52)	0.159		
ONFH (reference: OA)						
Other diagnosis†† (reference: OA)						
History of DDH (reference: no history)					5.70 (−1.11 to 12.51)	0.101
History of any osteotomy (reference: no history)						
UCLA activity score (range, 1-10)	1.10 (−0.22 to 2.43)	0.103	1.42 (−0.20 to 3.04)	0.085		
OHS (range, 0-48)	0.76 (0.42 to 1.11)	<0.001**	1.00 (0.50 to 1.50)	<0.001**	1.30 (0.77 to 1.82)	<0.001**
VAS satisfaction (range, 0-100)	0.19 (0.07 to 0.31)	0.002**	0.23 (0.04 to 0.42)	0.020**	0.21 (−0.01 to 0.43)	0.063
Satisfaction with pain (range, 25-100)						
Satisfaction with housework (range, 25-100)						
Satisfaction with recreation (range, 25-100)	0.42 (0.27 to 0.57)	<0.001**	0.36 (0.15 to 0.58)	<0.001**	0.42 (0.18 to 0.67)	0.001**
Ease of getting in and out of car (range, 1-4)	4.96 (2.64 to 7.28)	<0.001**	6.84 (3.67 to 10.01)	<0.001**	5.79 (2.29 to 9.29)	0.001**
Ease of squatting (range, 1-4)	3.23 (1.44 to 5.02)	<0.001**	2.60 (0.25 to 4.95)	0.031**	2.84 (0.25 to 5.42)	0.032**
Ease of sitting in the seiza style (range, 1-4)						

*The regression coefficient, β, indicates the estimated increase in FJS-12 per unit increase in the dependent variable. CI = confidence interval, BMI = body mass index, THA = total hip arthroplasty, ONFH = osteonecrosis of the femoral head, OA = osteoarthritis, DDH = developmental dysplasia of the hip, UCLA = University of California Los Angeles, OHS = Oxford Hip Score, VAS = visual analogue scale. Blank cells indicate variables that were not selected in the stepwise selection process based on the Akaike information criterion. Because the variables selected for each multiple complete data set could be different, the same variable selected in sensitivity analysis 1 was used for the primary analysis. †Primary analysis. Multiple imputation using 65 data sets, whose results were combined using the Rubin rules. ‡Sensitivity analysis 1. A single data set. For respondents with 1 to 3 missing items, each missing score was replaced with the mean of the scores for the items completed by the patient; respondents with 4 or more missing items were excluded (per the original scoring rule). Respondents with missing values for ≥1 of the 17 non-FJS variables used in the multivariable regression analysis were also excluded. §Sensitivity analysis 2. A single data set. Respondents with missing values for ≥1 of the 17 non-FJS variables used in the multivariable regression analysis were excluded. **Significant at $p < 0.05$. ††Rheumatoid arthritis, fracture, arthrodesis, spondyloepiphyseal dysplasia, posttraumatic osteonecrosis, and rapidly destructive coxopathy.

The mean FJS-12 (and standard deviation) was 51.6 ± 26.7 for complete respondents, 42.1 ± 26.9 using the original scoring rule for those with 1 to 3 missing responses, and 32.8 ± 24.5 using PMM for those with 4 to 11 missing responses. The complete respondents differed significantly from those missing 1 to 3 items and those missing 4 to 11 items ($p < 0.001$ for both, Fig. 3).

When all partial and complete respondents were included and the multivariable regression analysis results for the 65 data sets were combined, the mean FJS-12 was 47.5 ± 27.1, and a higher age at the time of the survey was associated with a higher FJS-12 value ($p < 0.001$) (Table III). In the sensitivity analyses that did not use multiple imputation, age at the time of the survey was not associated with the FJS-12.

Discussion

This is the first study, to our knowledge, to use multiple imputation to create complete data sets for partial respondents in order to permit their inclusion in analyses of FJS-

12 data. We found that the pattern of missing FJS-12 items in this Asian cohort was not MCAR, and that exclusion of partial respondents resulted in a nonresponse bias that might be associated with several factors such as sex differences in sports participation after THA[38] and differences in activity levels[8]. This finding is clinically important for orthopaedic research using PROMs with missing outcome values, an issue that has not gained much attention[2]. The difference in FJS-12 values between complete respondents and those with 4 to 11 missing items, whom the FJS-12 developers recommended should be excluded[6], exceeded 17.5 points, which is the previously reported minimal clinically important difference[39]. Because worldwide use of the FJS-12 as an outcome is anticipated, it is necessary to discuss the validity of applying the original scoring rule to questionnaires with missing values, essentially imputing a value based on the mean of the other individual items if 1 to 3 items were not completed but not scoring questionnaires with ≥4 missing items. This study of medium to long-term FJS-12 results after THA indicated that the rate of missing values increased with lower patient activity and greater age. Therefore, analysis using multiply imputed data sets may be recommended.

Our analysis of FJS-12 values calculated from multiple imputation of data sets using PMM showed that they had excellent[34] reliability when up to 9 items were missing. However, the reliability of the model might depend on the sample size[40], and this will need to be assessed in future studies. Furthermore, since the actual values of the missing items are not known, sensitivity analyses based on various assumptions regarding those values are recommended[15,36].

In our study, older age was a risk factor for skipping item 1, which asks about "awareness in bed at night." While this item seems to assess a normal experience, rates of bed use were only 14.9% in an assessment in 1965 and 60.5% in 2004 according to a Japanese government survey [41], and are believed to be lowest among older adults. As lifestyles and functional needs vary even among different generations within the same country, the creation of questions that could be answered by all patients is unrealistic. However, in the present study, the multiple imputation method allowed the inclusion of older individuals who used traditional Japanese tatami mats instead of beds, did not sit on chairs, and did not hike or play sports.

The FJS-12 has been reported to be higher in older patients than in younger ones after THA[42], after total knee arthroplasty[43], and in the general male population in the U.S.[7]. This observation has been attributed to the naturally reduced activity levels with aging, reducing the opportunity for patients to be aware of their joints and thus resulting in higher FJS-12 values[43]. Our analysis, in which we included partial respondents by using multiple imputation, confirmed that older age was associated with higher FJS-12 values, as previously reported.

This study has some limitations. First, only 63.9% of the patients who received the FJS-12 questionnaire replied and were analyzed. Second, radiographic evaluations were not performed. Third, multiple imputation models other than PMM were not investigated.

Previous reports[44,45] indicated that if >10% of study participants had missing data, maximal effort was required to avoid bias during analysis[2,15]. While there are approaches to develop PROMs tailored to each population to avoid missing values, the use of a myriad of PROMs makes comparisons between studies difficult[1]. It is necessary to establish a core set of PROMs for comprehensive evaluation and surgical decision-making, including PROMs for those with different lifestyles, by considering missing values (i.e., why patients may fail to respond to particular items). Furthermore, we need to recognize the types of movements required for different patient populations and provide patients with information on the extent to which these movements could be achieved after surgery.

In conclusion, analyzing only complete FJS-12 responses after THA in an Asian cohort resulted in a nonresponse bias that preferentially excluded older, female, and less active individuals and those who preferred the traditional floor living style. Missing values, which are not just gaps in the data set, may have clinical importance, and that should be illuminated. As indicated by our findings, patients who do not complete all questions might have worse outcomes in general. Multiple imputation could be used as one solution for including partial respondents, who may have different lifestyle or functional needs, in the scoring and analysis of PROMs.

Appendix

eA Supporting material provided by the authors is posted with the online version of this article as a data supplement at jbjs.org (http://links.lww.com/JBJS/H288). ■

NOTE: The authors thank the medical assistants from the Department of Orthopaedic Surgery, Kyushu University, for their assistance with mailing the questionnaire.

Satoshi Yamate, MD[1]
Satoshi Hamai, MD, PhD[1,2]
Shinya Kawahara, MD, PhD[1]
Daisuke Hara, MD, PhD[1]
Goro Motomura, MD, PhD[1]
Satoshi Ikemura, MD, PhD[1]
Masanori Fujii, MD, PhD[3]
Taishi Sato, MD, PhD[1]
Satoru Harada, MD[1]
Tetsunari Harada, MD[1]
Yasuhiko Kokubu, MD[1]
Yasuharu Nakashima, MD, PhD[1]

[1]Department of Orthopaedic Surgery, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

[2]Department of Medical-Engineering Collaboration for Healthy Longevity, Kyushu University, Fukuoka, Japan

[3]Department of Orthopaedic Surgery, Faculty of Medicine, Saga University, Saga, Japan

Email for corresponding author: shamai0220@gmail.com

References

1. Vajapey SP, Morris J, Li D, Greco NG, Li M, Spitzer AI. Outcome reporting patterns in total hip arthroplasty: A systematic review of randomized clinical trials. JBJS Rev. 2020 Apr;8(4):e0197.

2. Rombach I, Rivero-Arias O, Gray AM, Jenkinson C, Burke Ó. The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature. Qual Life Res. 2016 Jul;25(7):1613-23.

3. Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? Int J Epidemiol. 2014 Aug;43(4):1336-9.

4. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. BMC Med Res Methodol. 2014 Nov 19;14:118.

5. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, Petersen I. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017 Mar 15;9:157-66.

6. Behrend H, Giesinger K, Giesinger JM, Kuster MS. The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. J Arthroplasty. 2012 Mar;27(3):430-436.e1.

7. Giesinger JM, Behrend H, Hamilton DF, Kuster MS, Giesinger K. Normative Values for the Forgotten Joint Score-12 for the US General Population. J Arthroplasty. 2019 Apr;34(4):650-5.

8. Cao S, Liu N, Han W, Zi Y, Peng F, Li L, Fu Q, Chen Y, Zheng W, Qian Q. Simplified Chinese version of the Forgotten Joint Score (FJS) for patients who underwent joint arthroplasty: cross-cultural adaptation and validation. J Orthop Surg Res. 2017 Jan 14;12(1):6.

9. Hamilton DF, Loth FL, Giesinger JM, Giesinger K, MacDonald DJ, Patton JT, Simpson AH, Howie CR. Validation of the English language Forgotten Joint Score-12 as an outcome measure for total hip and knee arthroplasty in a British population. Bone Joint J. 2017 Feb;99-B(2):218-24.

10. Sethy SS, Goyal T, Paul S, Das SL, Choudhury AK, Kalia RB. Translation and Validation of Forgotten Joint Score for Total Hip Arthroplasty for Indian Population. Indian J Orthop. 2020 Aug 17;54(6):892-900.

11. Matsumoto M, Baba T, Homma Y, Kobayashi H, Ochi H, Yuasa T, Behrend H, Kaneko K. Validation study of the Forgotten Joint Score-12 as a universal patient-reported outcome measure. Eur J Orthop Surg Traumatol. 2015 Oct;25(7):1141-5.

12. Larsson A, Rolfson O, Kärrholm J. Evaluation of Forgotten Joint Score in total hip arthroplasty with Oxford Hip Score as reference standard. Acta Orthop. 2019 Jun;90(3):253-7.

13. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581-92.

14. European Medicines Agency. Committee for Medicinal Products for Human Use guideline on missing data in confirmatory clinical trials. 2010 Jul 2. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf

15. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP, Stern H. The prevention and treatment of missing data in clinical trials. N Engl J Med. 2012 Oct 4;367(14):1355-60.

16. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: National Academies Press; 2010.

17. Van Ginkel JR, Andries Van der Ark L, Sijtsma K, Vermunt JK. Two-way imputation: A bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation. Comput Stat Data Anal. 2007;51(8):4013-27.

18. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011 Feb 20;30(4):377-99.

19. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. BMC Med Res Methodol. 2010 Dec 31;10:112.

20. Bailey BE, Andridge R, Shoben AB. Multiple imputation by predictive mean matching in cluster-randomized trials. BMC Med Res Methodol. 2020 Mar 30;20(1):72.

21. Little RJA. Missing-data adjustments in large surveys. J Bus Econ Stat. 1988;6(3):287-96.

22. Robinson PG, Rankin CS, Lavery J, Anthony I, Blyth M, Jones B. The validity and reliability of the modified forgotten joint score. J Orthop. 2018 Mar 27;15(2):480-5.

23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007 Oct 20;370(9596):1453-7.

24. Amstutz HC, Thomas BJ, Jinnah R, Kim W, Grogan T, Yale C. Treatment of primary osteoarthritis of the hip. A comparison of total joint and surface replacement arthroplasty. J Bone Joint Surg Am. 1984 Feb;66(2):228-41.

25. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br. 1996 Mar;78(2):185-90.

26. Brokelman RB, Haverkamp D, van Loon C, Hol A, van Kampen A, Veth R. The validation of the visual analogue scale for patient satisfaction after total hip arthroplasty. Eur Orthop Traumatol. 2012 Jun;3(2):101-5.

27. Mahomed N, Gandhi R, Daltroy L, Katz JN. The self-administered patient satisfaction scale for primary hip and knee arthroplasty. Arthritis. 2011;2011:591253.

28. Sugano N, Tsuda K, Miki H, Takao M, Suzuki N, Nakamura N. Dynamic measurements of hip movement in deep bending activities after total hip arthroplasty using a 4-dimensional motion analysis system. J Arthroplasty. 2012 Sep;27(8):1562-8.

29. Shiomoto K, Hamai S, Motomura G, Ikemura S, Fujii M, Nakashima Y. Influencing factors for joint perception after total hip arthroplasty: Asian cohort study. J Arthroplasty. 2020 May;35(5):1307-14.

30. Harada S, Hamai S, Gondo H, Higaki H, Ikebe S, Nakashima Y. Squatting After Total Hip Arthroplasty: Patient-Reported Outcomes and In Vivo Three-Dimensional Kinematic Study. J Arthroplasty. 2022 Apr;37(4):734-41.

31. Hidalgo B, Goodman M. Multivariate or multivariable regression? Am J Public Health. 2013 Jan;103(1):39-40.

32. Buuren SV, Groothuis-Oudshoorn K. Mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1-67.

33. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34-42.

34. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016 Jun;15(2):155-63.

35. Furukawa K, Preston DL, Misumi M, Cullings HM. Handling incomplete smoking history data in survival analysis. Stat Methods Med Res. 2017 Apr;26(2):707-23.

36. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol. 2019 Jun;110:63-73.

37. Chaurasia A, Harel O. Using AIC in Multiple Linear Regression framework with Multiply Imputed Data. Health Serv Outcomes Res Methodol. 2012 Jun;12(2-3):219-33.

38. Hara D, Hamai S, Fukushi JI, Kawaguchi KI, Motomura G, Ikemura S, Komiyama K, Nakashima Y. Does Participation in Sports Affect Osteoarthritic Progression After Periacetabular Osteotomy? Am J Sports Med. 2017 Sep;45(11):2468-75.

39. Longo UG, De Salvatore S, Piergentili I, Indiveri A, Di Naro C, Santamaria G, Marchetti A, Marinis MG, Denaro V. Total Hip Arthroplasty: Minimal Clinically Important Difference and Patient Acceptable Symptom State for the Forgotten Joint Score 12. Int J Environ Res Public Health. 2021 Feb 25;18(5):2267.

40. Simons CL, Rivero-Arias O, Yu LM, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index? Qual Life Res. 2015 Apr;24(4):805-15.

41. Cabinet Office, Government of Japan. Consumer confidence survey. Accessed 2004 Mar. https://www.esri.cao.go.jp/jp/stat/shouhi/0403fukyuritsu.xls

42. Cech A, Kase M, Kobayashi H, Pagenstert G, Carrillon Y, O'Loughlin PF, Aït-Si-Selmi T, Bothorel H, Bonnin MP. Pre-operative planning in THA. Part III: do implant size prediction and offset restoration influence functional outcomes after THA? Arch Orthop Trauma Surg. 2020 Apr;140(4):563-73.

43. Behrend H, Zdravkovic V, Giesinger J, Giesinger K. Factors Predicting the Forgotten Joint Score After Total Knee Arthroplasty. J Arthroplasty. 2016 Sep;31(9):1927-32.

44. Bennett DA. How can I deal with missing data in my study? Aust N Z J Public Health. 2001 Oct;25(5):464-9.

45. Rombach I, Gray AM, Jenkinson C, Murray DW, Rivero-Arias O. 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 Aug 28;18(1):87.