Journal of Epidemiology 27 (2017) 163–171

Contents lists available at ScienceDirect

Journal of Epidemiology

journal homepage: http://www.journals.elsevier.com/journal-of-epidemiology/

Original Article

Handling missing Mini-Mental State Examination (MMSE) values: Results from a cross-sectional long-term-care study

Judith Godin [a, b], Janice Keefe [b], Melissa K. Andrew [a, *]

[a] *Geriatric Medicine Research Unit, Nova Scotia Health Authority and Dalhousie University, Canada*
[b] *Department of Family Studies and Gerontology and the Nova Scotia Centre on Aging, Mount Saint Vincent University, Canada*

ARTICLE INFO

Article history:
Received 22 July 2015
Accepted 17 May 2016
Available online 27 December 2016

Keywords:
MMSE
Missing data
Multiple imputation
Long-term care
Care and construction project

ABSTRACT

Background: Missing values are commonly encountered on the Mini Mental State Examination (MMSE), particularly when administered to frail older people. This presents challenges for MMSE scoring in research settings. We sought to describe missingness in MMSEs administered in long-term-care facilities (LTCF) and to compare and contrast approaches to dealing with missing items.
Methods: As part of the Care and Construction project in Nova Scotia, Canada, LTCF residents completed an MMSE. Different methods of dealing with missing values (e.g., use of raw scores, raw scores/number of items attempted, scale-level multiple imputation [MI], and blended approaches) are compared to item-level MI.
Results: The MMSE was administered to 320 residents living in 23 LTCF. The sample was predominately female (73%), and 38% of participants were aged >85 years. At least one item was missing from 122 (38.2%) of the MMSEs. Data were not Missing Completely at Random (MCAR), χ^2 (1110) = 1,351, $p < 0.001$. Using raw scores for those missing <6 items in combination with scale-level MI resulted in the regression coefficients and standard errors closest to item-level MI.
Conclusions: Patterns of missing items often suggest systematic problems, such as trouble with manual dexterity, literacy, or visual impairment. While these observations may be relatively easy to take into account in clinical settings, non-random missingness presents challenges for research and must be considered in statistical analyses. We present suggestions for dealing with missing MMSE data based on the extent of missingness and the goal of analyses.

Introduction

Study of the use of the Mini Mental State Examination (MMSE), a test of cognitive function in older adults,[1] among long-term-care residents is limited. Missing values for individual items are common, particularly when the MMSE is administered to frail older people. This could be due to participants declining to answer items, the setting in which the test is administered (e.g., ease of sitting upright or the presence of a suitable writing surface), participants' inability to write (e.g., due to hand weakness or tremor) or due to visual deficits or literacy challenges. Because of their training and experience, clinicians are able to interpret test scores with an understanding of why items are missing. In contrast, handling missing MMSE scores in a research setting is challenging: research assistants may administer the test and will not have sufficient training or knowledge of the patients to make these clinically-based decisions.

Multiple imputation (MI) is a highly recommended method of dealing with missing data. Researchers have tested the accuracy of MI in both Monte Carlo simulations[2,3] and using real data.[4,5] Item-level imputation performs better than scale-level imputation,[6,7] and standard errors (SEs) can increase by up to 10% when using scale-level over item-level MI.[6]

Item-level MI is reliable, but not always feasible. In order for MI to produce accurate estimates, all variables that will be included in the analyses must be in the imputation model. Thus, the imputation model can become unwieldy with even moderately sized datasets, especially if other variables are also missing data.[6,8] We compared alternative missing-data techniques to item-level MI, which we considered the "gold standard" method, to test whether simpler more feasible techniques provided accurate estimates and SEs. We

* Corresponding author. c/o Division of Geriatric Medicine, Veterans' Memorial Building, 5955 Veterans' Memorial Lane, Halifax, Nova Scotia, Canada.
E-mail address: mandrew@dal.ca (M.K. Andrew).

Peer review under responsibility of the Japan Epidemiological Association.

http://dx.doi.org/10.1016/j.je.2016.05.001

aimed to provide practical recommendations for researchers dealing with missing values in the MMSE.

We explored patterns of missing MMSE data, compared methods for addressing missing data in research settings, and assessed which alternative techniques best measure up to item-level MI.

Methods

Data

We analyzed data from the Care and Construction Project, which was conducted in long-term-care facilities (LTCF) in Nova Scotia, Canada. This project examined resident quality of life from the perspectives of residents, their families, and staff.[9] We used data from the resident survey, in which residents completed an interview-based questionnaire. Criteria for inclusion were willingness to participate and ability to consent and communicate in English.

Rates of cognitive impairment and dementia in Nova Scotia LTCF are high; recent studies identified a dementia prevalence of 62–64%,[10,11] though under-diagnosis likely remains a significant problem. To encourage participation across a range of abilities, and because capacity to consent is poorly correlated with scores on tests of cognition, the MMSE was not used as an inclusion criterion. Rather, the MMSE was an explanatory variable included to explore the impact of cognition on quality of life.

Standard protocol approvals, registrations, and patient consent

An informed-consent process was used to assess residents' ability and willingness to participate in this study. Ethics Review Boards of all participating universities and, where appropriate, participating LTCF and health authorities approved the research conducted during the project.

Measures

Demographic variables

Age was recorded as 18–64, 65–74, 75–84, or ≥85 years. Participants reported their sex, marital status (never married, married or common law, divorced or separated, or widowed), education (8th grade or less, some high school, completed high school, some college or university, or college or university graduate), and tenure in the LTCF (<6 months, 6 to <12 months, 12 months to 2 years, or >2 years).

Mini Mental State Examination

The MMSE is a standardized cognitive screening test with a possible score of 0–30. Domains assessed include orientation, registration and short-term recall, attention and concentration, language (naming, sentence writing, and comprehension), and visuospatial abilities.[1] Individual items are summed to generate the total score. If individuals decline or are unable to attempt a task, the value on that particular item would be missing. Trained research assistants administered the MMSE as part of the full study interview.

EQ-5D

Participants responded to the EQ-5D, which includes five questions assessing mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, and indicated whether they had no problems, some problems, or extreme problems. Scores were converted into a single index using the method that incorporates country-appropriate value weighting.[12,13] We used value sets derived from a representative American sample,[14] as a value set does not currently exist for Canada. An index score of 1.00 indicates perfect health.

Participants self-reported their health on a visual analogue scale that is part of the EQ-5D but not used in the index. The visual analogue scale ranged from 0 (i.e., worst imaginable health) to 100 (i.e., best imaginable health).

Quality of life and nursing home experience

Two single-item measures were used to assess quality of life and nursing home experience: '*How would you describe your overall quality of life?*' and '*Given your health status today, how would you describe your overall experience of living in this nursing home?*'. Participants indicated their responses on a scale of 1 (very poor) to 5 (very good).

Analyses

First, we described item-specific MMSE missingness in relation to demographic and well-being variables. For MMSE items with more than 5% missing data, independent-samples t-tests were used to compare means of those with and without missing data on the other MMSE items, MMSE total score, the EQ-5D, health, and the two single-item questions. χ^2 cross-tabulations were used to examine relationships between MMSE score missingness and demographic variables. We also examined the bivariate correlations between item scores and correlations between item missingness (Table 2).

Second, we examined different techniques (listwise deletion, scale-level MI, raw scores, and normed totals) for dealing with MMSE missing data and compared these techniques to item-level MI. For many techniques, data are assumed to be missing completely at random (MCAR); that is, the missing values are a random sample of the complete data. In practice, data are rarely MCAR and are usually either missing at random (MAR) or not missing at random (NMAR). Data are MAR when, after controlling for other variables in the data, there are no associations between the missingness and the variable itself. Data are NMAR when the missingness is associated with the variable itself or unmeasured variables.

Listwise deletion leads to a loss of power and, when data are not MCAR, results in biased estimates.[4,6–8] MI is considered one of the best methods for dealing with missing data because it produces estimates that are very close to complete data analysis, retains power, and takes into consideration uncertainty inherent in missing data analyses.[6,8] In MI, missing values are imputed 'm' times based on other variables in the imputation model and random error, thus creating 'm' datasets. Standard analyses are conducted on each of the multiple-imputed data sets and the results are pooled. For estimates, an average across all datasets is taken. SEs are pooled using Rubin's rules,[15] which take into account within- and between-imputation variance.

Here, data were multiple-imputed via chained equations (MICE) using the MICE package[16] in R.[17] Each variable was imputed using predictive mean matching. We imputed 20 datasets, conducted analyses on each dataset individually, and pooled the results. This pooling is important because, rather than focusing on individual datasets or analyses, researchers should rely on the pooled results.[18]

Imputing individual items can lead to a large number of variables in the imputation model, so it is not always a practical option. The dataset must also contain responses for each individual item in the scale (here, the MMSE), which is not always the case. MI

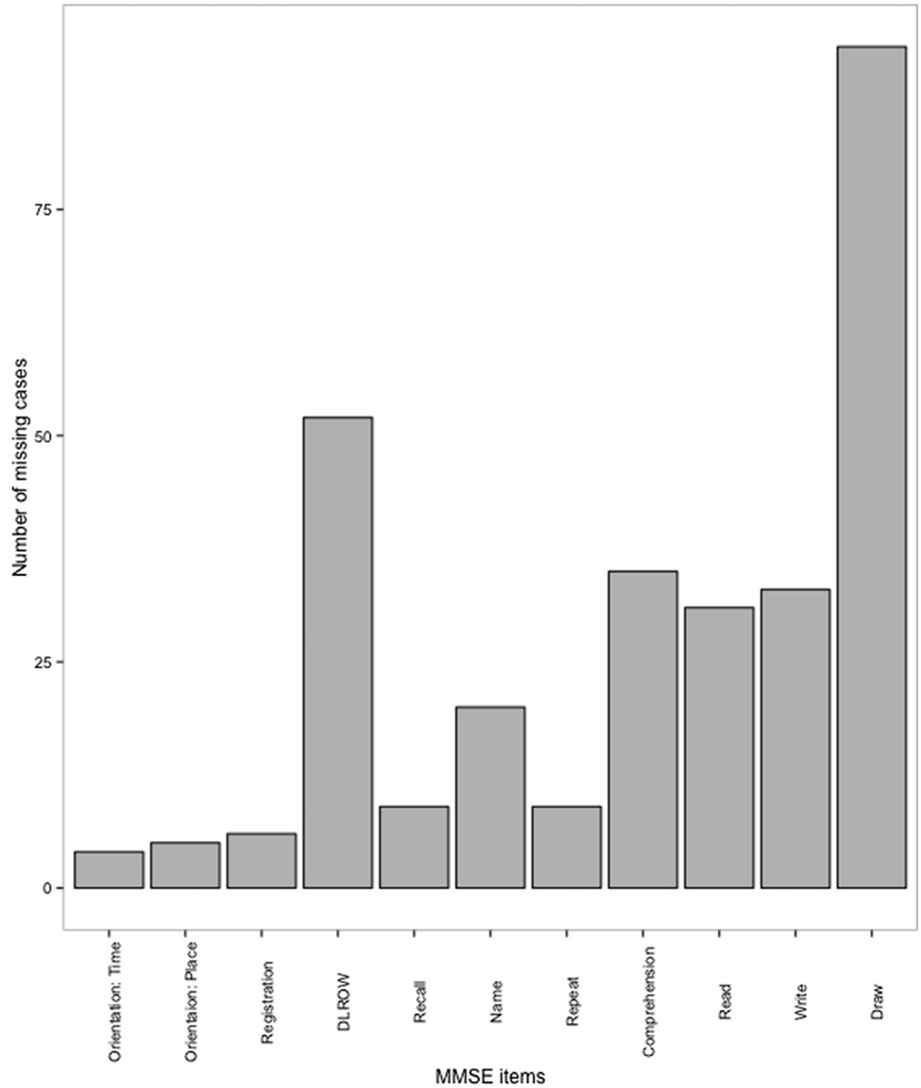

Fig. 1. Distribution of missing data by item on the MMSE in our sample of long-term-care facility residents. MMSE, Mini-Mental State Examination.

assumes that the data are at least MAR; including auxiliary variables in the imputation model increases the tenability of the MAR assumption.[18,19] Good auxiliary variables predict the values of the variable with missing data and the missingness. Imputing at the item-level provides a number of variables that meet these criteria. With item-level imputation, individual items are imputed based on other MMSE items and all demographic and well-being variables. Subsequently, the scale total is computed based on the imputed items. All variables that are included in the main analyses must be included in the imputation model.[6] Excluding variables from the imputation model can lead to extremely biased results [2]; therefore, we included both predictors and outcomes in our imputation model. Although assessing missing-data techniques using analyses that included variables from the imputation model may raise concerns regarding overfitting, this method has been consistently used in missing-data studies.[2,4,20–22]

We compared the following missing-data techniques to item-level MI:

1. Excluding participants with any missing data (listwise deletion).
2. Using raw scores (i.e., correct items÷30) for participants who were missing up to 5, 10, or 15 items and using raw scores regardless of how many missing points.
3. Using normed scores (i.e., [correct items÷complete items] × 30) for participants who were missing up to 5, 10, or 15 items and using normed scores regardless of how many missing points.
4. Using scale-level MI only, using scale-level MI in combination with raw scores and normed scores, and using scale-level MI with a few key items included in the imputation model. With scale-level imputation, the scale total is imputed using other variables in the dataset.

Assessing which missing-data techniques are best for descriptive versus regression analyses.

Considering the goal of the analyses is important, and we anticipated the possibility that different missing-data techniques would perform best for descriptive versus regression analyses. Hence, we present our results in two sections: Section A relates to descriptive analyses and Section B relates to regression analyses. In Section A, mean MMSE scores obtained from different missing-data techniques were compared to the mean obtained from the gold standard technique, item-level MI.

In Section B, for each missing-data technique, five regression models were tested: MMSE score was regressed on sex, age, marital status, nursing-home tenure, and education in separated models to determine which alternative missing-data technique came closest to the profile of estimates for our gold standard comparison. Each demographic variable was dummy coded into C-1 dichotomous variables (where C = the number of categories). R^2 and regression coefficients obtained from these regressions were compared to those obtained through item-level MI.

Results

Description of the sample

The sample included 320 residents, of whom 72.5% were women. Regarding age categories, 18.1% were younger than 65, 44.1% were aged 65–84 years, and 37.8% were 85 years or older.

Table 1
Means of MMSE items and demographic and well-being variables by data present (P) and data missing (M) for MMSE items missing more than 5% of cases.

	DLROW		Name		Comp 1		Read		Write		Draw		MMSE	
	P	M	P	M	P	M	P	M	P	M	P	M	P	M
World: Spell world backwards	3.45	–	3.42	4.00	3.42	3.76	3.42	4.00	3.48	3.14	3.48	3.36	3.45	3.44
Recall apple: What are the 3 objects I asked you to remember	0.72	0.70	0.72	0.71	0.73	0.66	0.73	0.64	0.73	0.63	0.74	0.68	0.74	0.69
Recall penny: What are the 3 objects I asked you to remember	0.55	0.59	0.55	0.64	**0.53**	**0.79**	0.55	0.64	0.56	0.56	0.55	0.57	0.53	0.60
Recall table: What are the 3 objects I asked you to remember (Table)	0.43	0.41	0.43	0.36	**0.40**	**0.62**	0.43	0.40	0.43	0.37	0.43	0.42	0.41	0.44
Name pen: What is this called? (Pencil/Pen)	**0.98**	**1.00**	0.98	–	**0.98**	**1.00**	0.98	0.92	**0.98**	**1.00**	0.98	0.99	0.97	0.99
Name watch: What is this called?	0.99	0.95	0.98	–	**0.98**	**1.00**	0.99	0.92	**0.98**	**1.00**	0.99	0.96	0.99	0.97
Repetition: Please repeat the following: No ifs, ands or buts.	0.82	0.68	0.80	0.64	0.80	0.78	0.81	0.61	**0.78**	**0.96**	0.80	0.80	0.81	0.77
Comp. Took: Please take this piece of paper in your right hand.	0.90	0.86	0.89	0.91	0.89	–	0.89	0.84	0.89	0.94	0.89	0.90	0.90	0.87
Comp. fold: fold the paper in half	0.96	0.95	**0.96**	**1.00**	0.96	10.00	0.96	0.95	0.96	0.94	0.96	0.97	0.96	0.95
Comp. Put: Put the paper on the table.	0.97	0.90	**0.96**	**1.00**	0.96	10.00	0.96	0.95	0.96	0.94	0.96	0.98	0.97	0.94
Read: Please read the following	0.93	0.86	0.93	0.00	0.92	0.96	0.92	–	0.92	0.92	0.94	0.88	0.94	0.89
Write: Write any sentence on this piece of paper.	0.89	0.90	**0.89**	**1.00**	**0.88**	**1.00**	0.89	0.95	0.89	–	0.88	0.93	0.87	0.93
Draw: Please copy the drawing on the same piece of paper.	0.55	0.35	**0.54**	**0.00**	0.53	0.50	0.54	0.20	0.53	–	0.53	–	**0.56**	**0.34**
Year: What is the year?	**0.74**	**0.54**	0.71	0.69	0.69	0.81	0.71	0.67	0.70	0.76	0.71	0.69	0.73	0.66
Season: What is the season?	0.90	0.79	0.88	0.94	0.88	0.87	0.88	0.89	0.88	0.86	0.89	0.84	0.89	0.86
Month: What is the month?	**0.84**	**0.60**	0.80	0.81	0.80	0.77	0.80	0.81	0.81	0.72	0.81	0.76	**0.84**	**0.74**
Week: What day of the week is it?	0.73	0.67	0.73	0.56	0.72	0.74	0.72	0.70	0.72	0.69	0.71	0.74	0.71	0.73
Date: What is the date?	**0.50**	**0.33**	0.48	0.50	0.47	0.55	0.47	0.52	0.47	0.55	0.48	0.48	0.47	0.49
Province: What province are we in?	**0.97**	**1.00**	**0.97**	**1.00**	**0.97**	**1.00**	**0.97**	**1.00**	0.98	0.93	0.97	0.98	0.97	0.97
City: What city/town are we in?	0.94	0.87	**0.93**	**10.00**	0.93	0.97	0.93	0.96	0.93	0.90	0.93	0.92	0.93	0.92
Building: What is the building we are in?	**0.82**	**0.66**	0.81	0.67	0.81	0.73	0.80	0.77	0.79	0.89	0.81	0.77	**0.84**	**0.74**
Floor: What floor are we on?	0.79	0.66	0.77	0.80	**0.75**	**0.90**	0.76	0.85	**0.75**	**0.93**	0.75	0.81	0.75	0.79
Room: What is your room number?	**0.68**	**0.51**	0.66	0.60	0.65	0.73	0.65	0.69	0.65	0.68	0.67	0.61	0.68	0.61
Reg. 1: Can you repeat the 3 items for me (Apple)	0.99	0.96	**0.98**	**1.00**	0.99	0.97	0.99	0.96	**0.98**	**1.00**	0.98	0.99	0.98	0.98
Reg. 2: Can you repeat the 3 items for me (Penny)	0.99	0.96	0.99	0.93	**0.98**	**1.00**	0.99	0.96	**0.98**	**1.00**	0.98	0.99	0.99	0.97
Reg. 3: Can you repeat the 3 items for me (Table)	0.98	0.94	0.97	0.93	**0.97**	**1.00**	0.98	0.92	**0.97**	**1.00**	0.97	0.97	0.98	0.95
	0.55	0.35	**0.54**	**0.00**	0.53	0.50	0.54	0.20	0.53	–	0.53	–	**0.56**	**0.34**
EQ-5D index score	0.59	0.64	0.60	0.61	**0.62**	**0.46**	0.60	0.60	0.60	0.62	**0.62**	**0.55**	0.61	0.59
Health: Visual analogue health scale	65.25	67.41	65.67	63.57	66.19	59.15	65.53	66.02	66.05	60.39	66.93	61.88	66.45	64.01
QOL: Single item quality of life	3.89	3.87	3.90	3.76	3.90	3.84	3.92	3.57	3.90	3.80	3.90	3.85	3.90	3.87
NH Exp: Single item nursing home experience	4.15	4.15	4.14	4.31	4.16	4.06	4.16	4.04	4.14	4.24	4.14	4.18	4.14	4.18

Bolded means are significantly different ($p < 0.05$).

Table 2
Items with 2% or more missing data: correlations between missingness (top), correlations between item scores (bottom), and percent of missing data (diagonal).

	World backwards	Recall: penny	Recall: table	Name: pen	Name: watch	Repetition	Comp: took	Comp: fold	Comp: put	Read	Write	Draw
World backwards	**16.2%**	0.06	0.07	0.11	0.11	0.07	0.00	0.00	0.00	0.10	−0.04	0.05
Recall: penny	0.15	**2.2%**	0.88	0.49	0.49	0.75	0.36	0.36	0.36	0.38	0.37	0.23
Recall: table	0.11	0.44	**2.8%**	0.42	0.42	0.66	0.30	0.30	0.31	0.33	0.32	0.18
Name: pen	−0.03	0.16	0.08	**6.3%**	1.00	0.42	0.28	0.28	0.29	0.75	0.21	0.35
Name: watch	−0.05	−0.01	0.06	0.35	**6.3%**	0.42	0.28	0.28	0.29	0.75	0.21	0.35
Repetition	0.10	−0.07	0.08	−0.07	0.07	**2.8%**	0.42	0.42	0.43	0.46	0.38	0.22
Comp: took	0.00	−0.03	−0.03	0.03	−0.05	0.14	**10.9%**	0.97	0.98	0.29	0.38	0.50
Comp: fold	−0.08	−0.08	−0.06	−0.03	−0.03	0.04	0.22	**10.9%**	0.98	0.29	0.38	0.50
Comp: put	−0.08	0.03	−0.04	0.10	−0.03	0.04	0.28	0.62	**10.9%**	0.30	0.38	0.52
Read	0.11	−0.02	0.12	0.06	0.31	0.03	0.08	0.02	0.02	**9.7%**	0.20	0.40
Write	0.22	0.10	0.03	0.03	0.12	0.09	0.13	0.05	0.05	0.08	**10.3%**	0.53
Draw	0.21	0.15	0.07	0.04	−0.09	0.12	0.04	0.13	0.19	−0.05	0.20	**29.1%**

Nearly half (48.1%) were widowed, and 17.8% reported being married or in a common-law relationship. For education, 22.5% had achieved 8th grade or lower, 32.5% had some high school, 18.1% graduated high school, 10.3% had some college or university, and 15.3% graduated university or college. A minority (14.4%) had been living in the LTCF for less than 6 months, whereas 45.0% had resided in LTC for more than 2 years. Sex and marital status had no missing data. Age was missing in 1.6%, education was missing in 1.3%, and length of time living in the LTCF was missing in 0.9% of cases.

The mean for the visual analogue health scale was 65.57 (standard deviation [SD] 21.17). The mean quality of life rating was 3.89 (SD 0.98) and the mean level of nursing home experience was 4.15 (SD 0.90). The mean EQ-5D index was 0.60 (SD 0.26). The EQ-5D index and quality of life were missing in 3.4% of cases. Nursing home experience was missing in 3.7% and the health scale was missing in 5.0% of the sample.

Patterns and mechanisms of missing data

Only 198 (61.9%) of the 320 participants completed all items. The frequency of missingness by item varied substantially. Orientation-to-time items were least likely to be missing (1.2%), while the pentagon-copying task was most frequently missing (29.1%) (Fig. 1).

Little's MCAR test was statistically significant, indicating the data were not MCAR ($\chi^2(1110) = 1{,}351$, $p < 0.001$). Missingness on particular items was sometimes associated with the values of other items; however, there was no discernible overarching pattern in these associations (Table 1 and 2).

An MMSE score was calculated for participants who completed all items. Participants who were missing at least one MMSE item scored significantly lower on identifying the month, identifying the building, and the drawing task. There were no associations between missing data on the MMSE and any of the well-being or demographic variables, with the exception of education

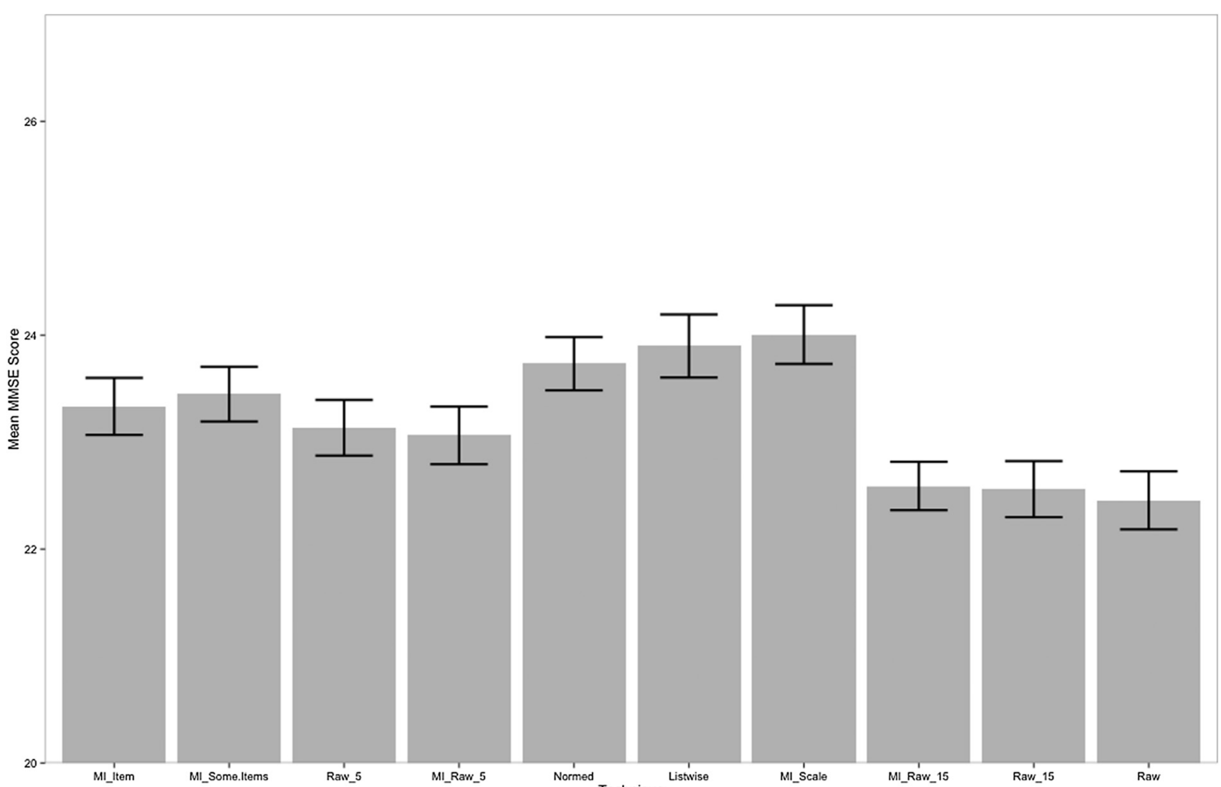

Fig. 2. Estimated means and standard error bars for MMSE scores across techniques. MMSE, Mini-Mental State Examination.

($\chi^2(4) = 12.56$, $p < 0.05$). However, there was no association between educational attainment and missing values, with the highest and lowest education groups having essentially equivalent missingness.

Section A. Comparison of other missing-data techniques to item-level MI for the purpose of descriptive analyses.

Using raw scores, either alone or in combination with MI, underestimated the mean MMSE score, except when a raw score was tabulated only for those participants missing five items or fewer. Using a normed total or listwise deletion overestimated the mean MMSE score. The technique that came closest to reproducing the mean obtained through item-level MI was scale-level MI with select items (based on correlations with MMSE total scores and MMSE missingness) included in the imputation model (Fig. 2).

Section B. Comparison of other missing-data techniques to item-level MI for the purpose of regression analyses.

Most techniques produced accurate estimates of R^2 for MMSE scores in relation to demographic variables (i.e., nursing home tenure, marital status, and sex). All techniques overestimated the R^2 value for education to some degree. Using a normed total shrunk the error variance for age, which led to a discrepancy in statistical significance (Table 3).

In examining the 14 regression coefficients for MMSE scores in relation to demographic variables (age [3], sex [1], marital status [3], time in LTCF [3], and education [4]), three techniques performed noticeably better than others and had an SE that changed less than a 5% from the gold standard item-level MI (Table 4). Scale-level MI in combination with raw scores for those missing five or fewer points produced 10/14 regression coefficients that were within 0.5 SEs of the item-level MI estimate and 12/14 estimates that were within 1 SE. Using raw scores without MI produced eight estimates within 0.5 SEs of the item level MI estimate and 12 estimates within 1 SE. Scale-level MI in combination with normed

Table 3
R^2s, F statistics, and Ns for the regressions of the MMSE on each demographic variable for each missing data technique.

Method	Statistic	Variables				
		Age	Education	Time in NH	Marital	Sex
Item-level MI	R^2	0.02	.05[b]	0.02	.04[b]	0.01
	F	1.48	3.63	1.47	3.81	2.46
Listwise	R^2	0.02	.09[b]	0.02	.04[a]	0.01
	F	1.56	4.47	1.37	2.99	2.21
	N	195	196	196	198	198
Raw scores	R^2	0.02	.07[c]	0.01	.04[b]	0.00
	F	1.67	6.06	1.32	4.32	1.15
	N	311	312	313	316	316
Raw scores (<6 missing)	R^2	0.02	.09[c]	0.02	.03[b]	0.01
	F	1.45	6.53	1.72	4.25	2.29
	N	279	280	280	283	283
Raw scores (<11 missing)	R^2	0.02	.09[c]	0.02	.04[b]	0.01
	F	1.86	7.25	1.80	4.43	1.48
	N	308	309	310	313	313
Raw scores (<16 missing)	R^2	0.02	.09[c]	0.02	.04[b]	0.01
	F	1.68	7.63	1.80	4.45	1.85
	N	309	310	311	314	314
Normed total	R^2	.02[a]	.06[b]	0.02	.04[b]	0.00
	F	2.76	4.43	1.74	4.53	0.77
	N	311	312	313	316	316
Normed total (<6 missing)	R^2	0.03	.09[c]	0.02	.05[b]	0.01
	F	2.61	6.61	1.77	4.57	1.83
	N	279	280	280	283	283
Normed total (<11 missing)	R^2	.03[a]	.07[c]	0.02	.04[b]	0.00
	F	2.67	5.48	1.95	4.50	1.10
	N	308	309	310	313	313
Normed total (<16 missing)	R^2	.03[a]	.07[c]	0.02	.04[b]	0.00
	F	2.70	5.38	1.96	4.53	1.04
	N	309	310	311	314	314
Scale MI	R^2	0.02	.09[b]	0.02	0.03	0.01
	F	1.28	4.02	1.09	1.93	1.79
Scale MI key items	R^2	.03[b]	.10[c]	0.01	.03[a]	0.01
	F	2.93	7.18	1.24	3.20	3.03
Scale MI -Raw scores (<6)	R^2	0.02	.10[c]	0.02	.04[a]	0.01
	F	1.42	7.57	1.48	3.57	1.09
Scale MI -Raw scores (<11)	R^2	0.02	.09[c]	0.02	.04[b]	0.01
	F	1.76	7.36	1.83	4.36	1.43
Scale MI -Raw scores (<16)	R^2	0.01	.10[c]	0.02	.04[b]	0.01
	F	1.44	7.63	1.68	4.33	1.67
Scale MI-Normed (<6)	R^2	0.03	.10[c]	0.02	.04[a]	0.01
	F	2.26	7.05	1.55	3.54	1.24
Scale MI Normed (<11)	R^2	0.02	.07[c]	0.02	.04[b]	0.00
	F	2.16	5.34	1.90	4.34	1.10
Scale MI Normed (<16)	R^2	0.02	.07[c]	0.02	.04[b]	0.00
	F	2.35	5.56	1.86	4.19	0.84

MI, multiple imputation; MMSE, Mini-Mental State Examination; NH, nursing home.
N for all MI techniques is 320.
[a] $p < 0.05$.
[b] $p < 0.01$.
[c] $p < 0.001$.

totals for those missing five or fewer points produced nine estimates within 0.5 SEs of the item level MI estimates and 10 estimates within 1 SE. We examined these analyses by variable and found that, for the four dummy-coded education variables, there were fewer techniques producing estimates within 0.5–1 SE of the item-level MI estimates compared to the other demographic variables.

Listwise deletion, normed total, normed total missing 10 or fewer points, normed total missing 15 or fewer points, and MI in combination with the latter two normed total techniques each produced 3 or fewer SEs that fell within a 5% change of item-level MI. Listwise deletion inflated SEs, whereas the other techniques had a tendency to produce smaller SEs.

Discussion

We found that missing MMSE items were not Missing Completely At Random; tasks requiring writing or sustained effort were more likely to be missing. Participants who were missing data on one or more items had significantly lower scores on identifying the month, identifying the building, and the drawing task, suggesting that those who were missing at least one item had lower levels of cognitive function compared to those who completed all items. For those with five or fewer missing items, use of raw scores with or without scale-level MI performed well in comparison to the gold standard of item-level MI. Other techniques, such as listwise deletion and normed scores, fared less well.

The patterns of missingness we identified suggest possible underlying reasons for the incomplete data. Residents missing one "orientation to time" item (which come up early in the MMSE test administration sequence) were missing all MMSE items, suggesting they were disinclined to participate in the MMSE. Missing data for naming a pen and a watch was strongly correlated with missing values for reading the "Close your eyes" sentence ($r = 0.75$), suggesting that visual difficulties could be contributing. Missingness on sentence writing was moderately correlated with missing values for the three-step command ($r = 0.38$), indicating a possible influence of trouble with manual dexterity. Not attempting the interlocking pentagons drawing task was strongly correlated with missing values on the three-step command ($r = 0.50$) and sentence writing ($r = 0.53$), and was moderately correlated with missingness on "Close your eyes" ($r = 0.40$), which may also be associated with manual dexterity or vision difficulties.

Surprisingly, missing values for spelling "WORLD" backwards were not correlated with missingness in writing a sentence ($r = -0.04$) or reading the "Close your eyes" command ($r = 0.10$), though these items all arguably draw on literacy skills. In fact, missing values for spelling "WORLD" backwards were not strongly correlated with any other item, suggesting that a phenomenon unique to this item may be at play, such as just "giving up" on a more challenging task.

We identified small but meaningful differences among the different missing-data techniques we tested. Using listwise deletion led to unacceptable reductions in sample size (38.2% of cases). This would be expected to be the case whenever the MMSE is administered in settings with frail participants, for whom fatigue or trouble with manual dexterity or vision may limit completion of some items.

Saunders et al considered a number of missing-data techniques, including MI.[5] They provided a worked example (hospitalized older adults with depression), though only 2% were missing MMSE scores. Thus, the differences between techniques were minor, and no specific recommendations were made.

Burns et al used item-level MI in a large dataset; however, the data were from community samples and had little missing data.[22] Burns et al had low numbers of the oldest old and found that MI inflated their scores more than for younger participants. They suggested that MI was less suited to this age group; however, this could be due to the small number of participants.[22] Here, our results are more generalizable to the oldest old as, due to the LTC setting, over a third of our sample was aged 85 years or older. This is particularly relevant for research in LTC settings, given the advanced age of residents (e.g., the mean age of LTC residents was 83 in another study[11]). We expand on the existing research by providing a worked example with a substantial amount of missing data and provide guidelines to help researchers choose missing-data techniques.

Table 4
Comparison of techniques for estimating 14 regression coefficient (one for sex, three each for age, marital status, and time in nursing home, and four for education).

Technique	±0.5 SE[a]	±1 SE[b]	Below estimate[c]	Above estimate[d]	<5% SE change[e]	SE −5%[f]	SE + 5%[g]
Listwise deletion	1	2	6	8	2	0	12
Raw scores	8	12	5	9	14	0	0
Raw scores (<6 missing)	7	9	7	7	9	5	0
Raw scores (<11 missing)	5	9	6	8	10	4	0
Raw scores (<16 missing)	5	9	6	8	10	2	2
Normed total	0	0	5	9	1	13	0
Normed total (<6)	8	8	7	7	8	3	3
Normed total (<11)	0	0	6	8	0	14	0
Normed total (<16)	0	0	6	8	0	14	0
Scale level MI	4	6	8	6	7	1	6
Scale MI - key items included	7	11	6	7	11	3	0
Scale MI - raw scores (<6)	10	12	7	7	12	2	0
Scale MI – raw scores (<11)	5	11	7	7	13	0	1
Scale MI - raw scores (<16)	4	10	6	8	13	1	0
Scale MI – normed total (<6)	9	10	8	6	10	4	0
Scale MI – Normed total (<11)	3	3	6	8	3	9	2
Scale MI – normed total (<16)	0	0	7	7	0	12	2

MI, multiple imputation; SE, standard error.
[a] Number of regression coefficients (out of 14) within 0.5 SE of item-level multiple imputation with less than 5% change in SE.
[b] Number of regression coefficients (out of 14) within 1 SE of item-level multiple imputation with less than 5% change in SE.
[c] Number of regression coefficients below item-level MI estimate.
[d] Number of regression coefficients above item-level MI estimate.
[e] Number of SEs that changed less than 5% from the item-level MI SE.
[f] Number of SEs that were negatively biased (too small).
[g] Number of SEs that were positively biased (too large).

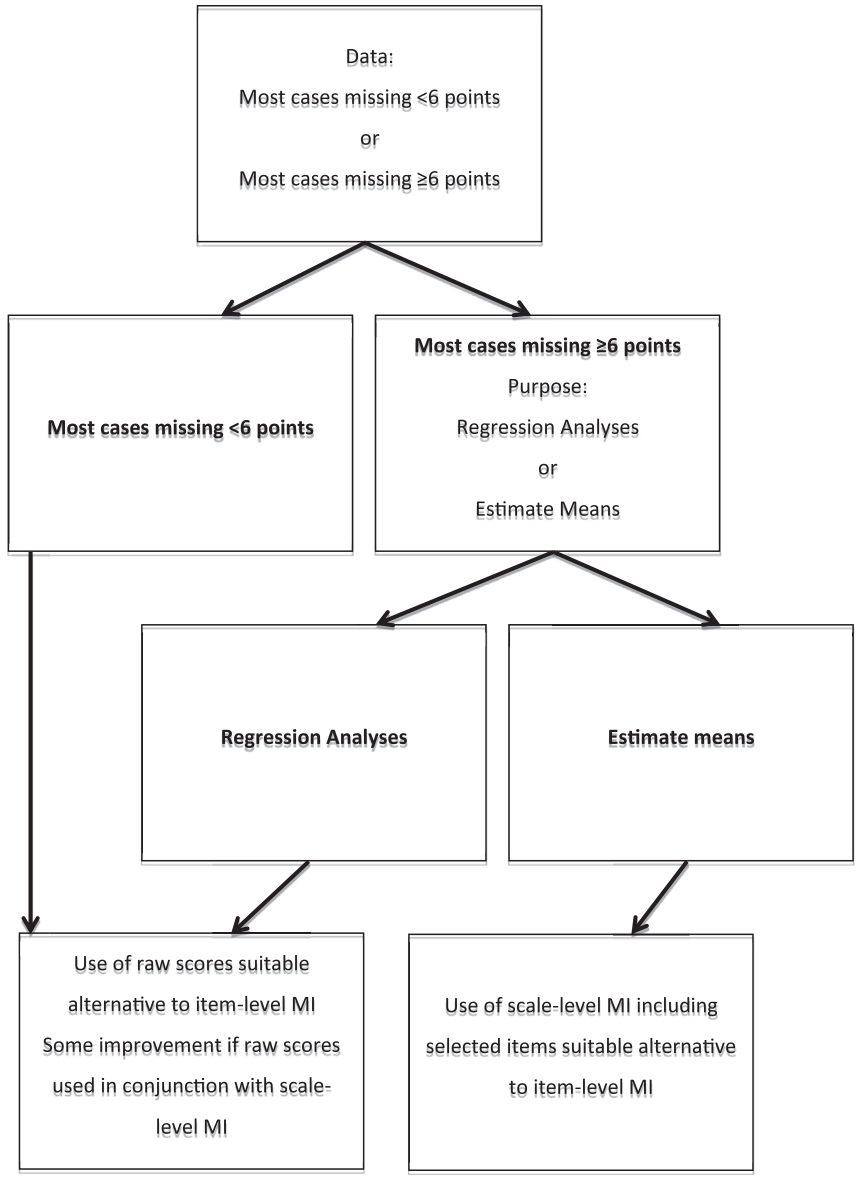

Fig. 3. A decision tree for choosing an appropriate missing-data technique Use of scale-level MI including selected items suitable alternative to item-level MI Use of raw scores suitable alternative to item-level MI Some improvement if raw scores used in conjunction with scale-level MI.

Item-level MI is well documented as a gold standard for dealing with missing data when the data are comprised of scales[6,18]; however, item-level MI is not always feasible. We found that scale-level MI on its own performed poorly, which may be due to low correlations between MMSE scores and demographic and well-being variables. For other techniques, the accuracy and precision depended on the statistic being estimated. Specifically, adding key items to the imputation model improved scale-level MI when estimating mean MMSE scores, but did not produce good estimates of regression coefficients. Using scale-level MI with raw scores for participants missing five or fewer points was the most consistent technique for producing accurate estimates and reasonable SEs for regression coefficients. The fact that the findings varied depending on which statistic was estimated suggests that the purpose of the analysis is an important consideration when choosing which missing-data technique to use.

Limitations

Our data should be interpreted with caution. Trained research assistants, rather than clinicians, administered the MMSE, and their scoring of items would likely not be as nuanced as an MMSE done as part of a clinical evaluation. The research assistants received standard training in MMSE administration and were instructed to encourage participants to complete as many items as they could, but participants were free to decline to answer any question. Further, our sample was accrued as part of a research study, as opposed to a clinical series. LTC residents who could not give informed consent to participate in the study were not included; as such, our sample and findings are not necessarily representative of patterns that might be seen with a more cognitively impaired sample.

We used item-level MI as our gold standard comparison. A more ideal comparison would be a complete data analysis (i.e., no missing data). Item-level MI, however, is a well-researched technique that is known to produce unbiased estimates and accurate standard errors.[6,18]

Recommendations and conclusions

Previous research has provided ample evidence in support of item-level MI. Including the individual items in the imputation model provides a number of variables associated with the other items and the item missingness, which makes the MAR assumption more tenable. However, if item-level MI is not feasible, there are appropriate alternatives that approximate results obtained through item-level MI.

To assist with choosing an appropriate missing-data technique, we have created a decision tree (Fig. 3) based on our results; however, further research is needed to evaluate this tool. When most cases with missing data are missing 5 or fewer points, using raw scores is a suitable and feasible alternative to item-level MI. If many cases have more than 5 missing points, the goal of the analyses should also be considered: for descriptive analyses, we suggest use of scale-level MI including selected items; for regression analyses, raw scores can be used on their own or in conjunction with scale-level MI.

Conflicts of interest

None declared.

Acknowledgements

The Care and Construction Project was supported by a Partnerships for Health System Improvement Grant, which was funded by the Canadian Institutes of Health Research (FRN # 114120) and the Nova Scotia Health Research Foundation (Matching-2011-7173). Judith Godin was supported by a Nova Scotia Health Research Foundation 2013 Scotia Support Grant (PSO-Research Programs-2013-9039).

This work was supported by the Canadian Consortium on Neurodegeneration in Aging, which receives funding from the Canadian Institutes of Health Research (CNA-137794) and partner organizations (www.ccna-ccnv.ca). This study is part of a Canadian Consortium on Neurodegeneration in Aging investigation into how multi-morbidity modifies the risk of dementia and the patterns of disease expression (Team 14). The funders had no role in conducting or approving the study for publication.

References

1. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189–198.
2. Moons KGM, Donders RART, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092–1101.
3. Newman DA. Longitudinal modeling with randomly and systematically missing data: a simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organ Res Methods*. 2003;6(3):328–362.
4. Fichman M, Cummings JN. Multiple imputation for missing data: making the most of what you know. *Organ Res Methods*. 2003;6(3):282–308.
5. Saunders JA, Morrow-Howell N, Spitznagel E, Dore P, Proctor EK, Pescarino R. Imputing missing data: a comparison of methods for social work researchers. *Soc Work Res*. 2006;30:19–31.
6. Enders CK. *Applied Missing Data Analysis*. New York: Guilford Press; 2010.
7. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods*. 2002;7(2):147–177.
8. Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol*. 2009 Jan;60:549–576 [Internet].[cited 2014 Mar 20];Available from: http://www.ncbi.nlm.nih.gov/pubmed/18652544.
9. Godin J, Keefe J, Kelloway EK, Hirdes JP. Nursing home resident quality of life: testing for measurement equivalence across resident, family, and staff perspectives. *Qual Life Res*. 2015;24(10):2365–2374 [Internet]. Springer International Publishing Available from: http://link.springer.com/10.1007/s11136-015-0989-4.
10. Smetanin P, Kobak P, Briante C, Stiff D, Sherman G, Ahmad S. Rising Tide: the impact of dementia in Canada 2008 to 2038. *RiskAnalytica*. 2009:1–345.
11. Marshall EG, Clarke BS, Varatharasan N, Andrew MK. Long-term care — comprehensive geriatric assessment (LTC-CGA) Tool: improving care for frail older adults? *Can Geriatr J*. 2015;18(1):2–10.
12. Oemar M, Oppe M. *EQ-5D-3L User Guide - Basic Information on How to Use the EQ-5D-3L Instrument. Version 5*. 2013.
13. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
14. Shaw JW, Johnson JA, Coons SJ, Shaw JW, Johnson JA, Coons SJ. EQ-5D model of the D1 valuation and testing development. *Med Care*. 2015;43(3):203–220.
15. Rubin DB. *Multiple Imputation for Nonresponse in Surveys. Wiley Clas*. Hoboken, NJ: John Wiley & Sons, Inc; 2004.
16. Buuren S, Groothuis-Ousdhoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1–67.
17. R Core Team. *R: A Language and Environment for Statistical Computing*; 2014 [Internet]. Austria Vienna; Available from: http://www.r-project.org/.
18. Graham JW. *Missing Data: Analysis and Design*. New York: Springer Science+-Business Media; 2012.
19. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
20. Dong Y, Peng C-YJ. In: . *Principled Missing Data Methods for Researchers*. vol. 2. Springerplus; 2013:222 [Internet].Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3701793&tool=pmcentrez&rendertype=abstract.
21. Collins LM, Schafer JL, Kam C-M. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods*. 2014 May 14;6(4):330–351. http://dx.doi.org/10.1037/1082-989X.6.4.330 [Internet]. 2001 cited Available from:.
22. Burns RA, Butterworth P, Kiely KM, et al. Multiple imputation was an efficient method for harmonizing the Mini-Mental State Examination with missing item-level data. *J Clin Epidemiol*. 2011;64(7):787–793. http://dx.doi.org/10.1016/j.jclinepi.2010.10.011 [Internet]. Elsevier Inc. Available from:.