Do the methods used to analyse missing data really matter? A case of Intermediate Care patients.

Billingsley Kaambwa[1]; Stirling Bryan[2] and Lucinda Billingham[3]

[1]Health Economics Unit, Public Health Building, University of Birmingham, Edgbaston, Birmingham B15 2TT

[2]Centre for Clinical Epidemiology and Evaluation, University of British Columbia, Research Pavilion, 702-828 West 10th Ave, Vancouver

[3]Cancer Research UK Clinical Trials Unit (CRCTU), University of Birmingham, Edbaston, Birmingham B15 2TT

ABSTRACT

Aims
Missing data is a common statistical problem in healthcare datasets from populations of elderly people. Some argue that arbitrarily assuming the mechanism responsible for the missingness and therefore the method for dealing with this missingness is not the best option - but is this always true? This paper explores what happens when extra information that suggests that a particular mechanism is responsible for missing data is disregarded and methods for dealing with the missing data are chosen arbitrarily.

Methods
Regression models are used to explain variation in costs, EQ-5D and Barthel index. Three methods for dealing with missingness were used: complete case analysis (assuming missing completely at random - MCAR), multiple imputation (assuming missing at random - MAR) and Heckman selection model (assuming missing not at random - MNAR).

Data
2,533 intermediate care patients with up to 42% missingness.

Results & Conclusions
Extra information strongly suggested that missing cost data was MCAR. Preliminary results show that MCAR and MAR-based methods yielded similar results while those based on MNAR-methods were statistically different.

All three mechanisms of missingness were shown to be potential causes of the missing EQ-5D and Barthel data. It was not clear however whether the choice of method had any significant effect on the results for these data.

Questions for discussion
- How large should departures from MAR be to effectively invalidate the results of an MAR-based analysis?
- When all three mechanisms of missingness can be logically considered to be responsible for the missingness, does the method used really matter?

Introduction

Missing data is an unwanted reality in most evaluations as it can lead to threats to the internal and external validity of the results obtained from analysing such data [1-3]. There is a possibility that even under the best of conditions, missing data may result in a significant reduction in sample size leading to threats to external validity as a sample reduced in size may no longer be representative of the target population [2-4]. This is more problematic in circumstances where the likelihood of response is related to observed characteristics. Certain forms of missingness can reduce the statistical power of the analyses of the available data and therefore compromise the internal validity of a study, which is more serious [5, 6]. A situation that can potentially lead to reduced internal validity is when the likelihood of response is related to the values of the variable for which values are only partially observed, which is a possibility in a lot of cases.

The three main reasons or mechanisms that lead to missing data are: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). If data are MAR or MCAR, they can also be referred to as "ignorable" data while those MNAR are "non-ignorable" [7]. Missing data are said to be ignorable if the parameters that are used to model the missing data process are not related to the parameters used to model the observed data while non-ignorability exists if there is a systematic difference between responders and nonresponders even after accounting for all the observed data [8, 9]. There are various methods that have been proposed to deal with missing data with each of these methods premised on a specific missing data mechanism [1, 10, 11]. Croninger and Douglas [9] indicate that the choice of method used for coping with missing data is not crucial if there is not much data missing and/or the sample is big. This is because most methods will yield similar results in such circumstances. But as the level of missingness rises and/or the sample becomes smaller, the choice of method becomes potentially more significant.

Most times, the method chosen to deal with missing data is not based on concrete evidence of the mechanism responsible for this missing data. It is consequently difficult to assess the accuracy of such methods because the data are by definition 'missing' [12]. It is a recognised fact that data often provide little or no information at all to help determine the correct mechanism behind missingness [4, 13]. In many scenarios, therefore, it is difficult, or even impossible, to know what mechanism is responsible for the missingness. Sometimes more than one mechanism may be responsible for different sets of missing data within the same evaluation [9, 14]. This therefore means that choosing among these alternative methods is not an easy task. Curran et al [14] suggest two approaches for determining the missing data mechanism: hypothesis testing and collecting extra information, during the data collection process, about why missing data is missing. In the absence of missing data being recovered and analysed, hypothesis testing can at best only rule out that missing data are MCAR with no way of confirming that data are actually MCAR. Provided enough data has been collected, it therefore seems that, where missing data is irrecoverable, it is only the latter approach that will give some fairly credible indication about whether data are MCAR, MAR or MNAR [14].

This paper explores what happens when extra information that suggests that a particular mechanism is responsible for missing data is disregarded and methods for dealing with the missing data are chosen arbitrarily. Using a dataset that had missing

data on several variables, the factors that explain variation in costs per patient, change in EQ-5D from admission to discharge (ΔEQ-5D) and change in the Barthel index from admission to discharge (ΔBarthel) of intermediate care (IC) patients were explored in a regression modelling framework. Three methods incorporating techniques for dealing with missing data were used: (1) generalised linear models (GLMs) and ordinary least squares (OLS) on complete cases (assuming that missing data were MCAR), (2) GLM and OLS models on data obtained through multiple imputation (assuming missing data were MAR) and (3) Heckman selection models (assuming that missing data were MNAR). We were interested in examining the signs and sizes of coefficients (and associated standard errors) in the regression model results obtained.

Methods

Source of data

Data for this study were obtained from five anonymous case study sites in the UK which were part of the National Evaluation of the Costs and Outcomes of IC for Older People (NECOIC) [15]. These sites were 'whole systems' of IC i.e. areas with a specific geographical boundary. IC services aim to prevent admission to acute care or long term care and also aid discharge from hospital for older people [15]. Quantitative data were collected by staff working for the IC services according to protocols set out by the evaluation team. Service staff completed study pro forma, with or on behalf of their patients, at the point of admission to the service, for all IC admissions over a defined period. They completed discharge questions on the day of discharge, transfer to another IC service or as soon as possible following end of service provision. In addition, reasons as to why some data were missing were sought and noted down. Data were collected between January 2003 and January 2004. Ethical approval was granted by the Trent Multicentre Research Ethics Committee.

Missing data in the IC dataset

The variables that were collected in the IC dataset are presented in Table 1. Up to 42% of the data were missing for some variables in that dataset. For purposes of comparing the methods for dealing with missing data, a decision was made to focus on missingness only in the dependent variables i.e. missingness in the cost per patient, ΔEQ-5D and ΔBarthel. As a result, there were no missing data for any of the independent variables. A sample of 717 individuals was therefore used for the cost per patient models and 125 (17.4%) of these individuals had missing observations on the cost variable. For the ΔEQ-5D and ΔBarthel models, a sample of 1105 individuals was utilised. Of this sample, 417 (37.7%) and 392 (35.5%) had missing values on the ΔEQ-5D and ΔBarthel variables, respectively.

The dependent variables

The cost per patient variable was calculated by combining resource data with budget information for the individual IC services.

The EQ-5D is an outcome measure whose construct validity when used on populations of older people has been well documented [16-19]. It is comprised of five dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. There are three levels of impairment in each domain: no,

some/moderate, and extreme problems in the relevant dimension of health. Using these responses, the EQ-5D is able to distinguish between 243 states of health [20, 21]. The UK-specific EQ-5D valuation algorithm was used in order to convert the EQ-5D health description into a valuation. EQ-5D scores have a range of -0.59 to 1: the maximum score of 1 represents perfect health and a score of 0 represents death[21]. Scores less than 0 represent health states that are worse than death [21-24]. Its generic nature makes it comparable across patient populations.

The Barthel Index has been recommended for scientific research purposes and also for use as a proxy for other outcome measures [25]. To measure a person's level of functional independence, the Barthel uses 10 items, with each item carrying different weights [26]. Two items (bathing and grooming) are rated on a two-point scale of 0 and 5, six (feeding, dressing, bowels, bladder, toilet use and stairs) on a three-point scale of 0, 5 and 10 and the last two items (transfers and mobility) are rated on a four-point scale of 0, 5, 10 and 15. The scores on each item are added to produce an overall score which ranges from 0 to 100. To standardise them, the scales used in this paper were divided by 5 and therefore ranged from 0 to 20 [27]. The higher the score recorded for an item, the greater the level of independence. The reliability, sensitivity and suitability for proxy-assessment of the Barthel has been shown elsewhere [27-29].

Reasons for missing data in the IC dataset

When data are MCAR, it implies that the probability of an item missing is unrelated to any measured or unmeasured characteristic for that unit [30] while under MAR, the probability of an item having incomplete data depends on other variables in the dataset [1]. MNAR is when the probability of missingness depends on the values of the unobserved values perhaps in addition to one or more other variables and/or the observed variables [31].

Because of time constraints placed on the data collection process, it was not possible to collect all the cost data. No other reason was established as being responsible for the missing cost data. This suggests that where cost data were missing, it would be reasonable to assume that these data were MCAR.

Information obtained from the IC coordinators about some of the missing EQ-5D and Barthel data indicated that some services did not routinely collect this information [15]. This suggested that it was plausible to assume that the missingness mechanism for such data was MCAR.

The ΔEQ-5D and ΔBarthel scores were calculated by subtracting the scores at admission from those at discharge. A number of individuals had however been transferred to other services before the end of their IC episode. For some of these, it meant that their EQ-5D and Barthel scores at 'discharge' were not collected making it impossible to compute the ΔEQ-5D and ΔBarthel variables. This could be seen as a situation where the missing data were MAR as the reason for the patients transfer was sometimes linked to their health or functional status e.g. the more functionally independent an individual was, the more likely they were to be transferred to a less intensive form of IC.

The mean Barthel scores for some individuals who had missing EQ-5D scores were on average lower than those for individuals who did not have missing EQ-5D information [32]. This suggests that lower Barthel scores could therefore be predictive of missing EQ-5D scores. Further there was a positive relationship between EQ-5D and Barthel scores. Since some individuals with missing EQ-5D data were associated with lower Barthel scores, it means that, by virtual of the positive relationship between the two instruments, there is a possibility that these individuals would also have had lower EQ-5D scores had these been collected. It was therefore reasonable to assume that some of the missing data on the EQ-5D could also have been MNAR i.e. the poorer ones' health status was, the more difficult it was for them to provide data on the EQ-5D. By the same token, some of the missing Barthel data could have been MNAR.

From this extra information collected, it was reasonable to assume that missing cost data were MCAR. On the other hand, MCAR, MAR and MNAR could all be assumed as the reason for the missing data on the EQ-5D and Barthel.

Choice of regression families

In this exercise, it was important to compare both the signs and sizes of coefficients (and associated standard errors) from the different regression models. Both costs per patient and outcome variables were skewed and heteroscedastic in their residuals. We chose the GLM as it is able to simultaneously deal with both problems [33, 34]. We also used log-transformation where the natural log of the dependent variable was obtained [35] as another method for dealing with skewed data despite several limitations associated with this approach [36, 37]. The exponentiated coefficients from the GLM model have been shown to be easily comparable to the exponentiated counterparts obtained from a log-transformed model [38]. For the cost models, therefore, a decision was made for the GLM to be used for both the complete cases and the multiply imputed datasets while a log transformed cost per patient was used in the Heckman regression model. A different approach was taken for the outcome dependent variables (ΔEQ-5D and ΔBarthel). This was because these variables also had negative values. As a result, log transformation of these variables would have required the use of a shift factor and the transformed variables would then have had to be appropriately retransformed once the results of the model had been obtained. However for ease of analysis and comparison, a decision was made to use the raw scale of these variables. As a result, OLS regressions were used for both the ΔEQ-5D and ΔBarthel in the regression on complete cases and on multiply imputed datasets. The raw scale of the two variables was also used in the Heckman selection models.

Approaches for dealing with the missing data

In this paper, we do not provide a detailed discussion of the various methods that can be used to deal with missing data. Interested readers can see Fielding et al [12] for such a discussion. In general though, complete case analysis (both listwise and pairwise deletion) can be performed when data are MCAR [39]. Approaches for use when data are MAR include listwise deletion, various imputation techniques, creating an extra category, propensity adjustment strategy, raw maximum likelihood and expectation maximisation [1, 4, 5, 40-43]. When data are MNAR, panel selection models, including the Heckman, and pattern-mixture approaches can be used [5, 10, 44-46].

For our two samples obtained from the NECOIC dataset, three methods, each assuming either MCAR, MAR or MNAR, were used. A regression framework was employed in the analysis and in general, the regression relationship between the outcomes of interest and the independent variables could be illustrated as [47]:

$$Y_i = f(X_i) + \mu_i$$

where Y_i denotes the outcome of interest (cost per patient, ΔEQ-5D or ΔBarthel) for the ith individual,
X_i is a vector of explanatory variables for the ith individual and
μ_i is the stochastic error term for the ith individual.

A total of six sets of regression models (two for each method) were conducted:

Method 1 involved running regression models on complete cases (assuming that data were MCAR). A GLM was used to explain variation in 'cost per patient' while OLS models were run for cases where the dependent variables were ΔEQ-5D and ΔBarthel. Pairwise deletion was used in arriving at the complete cases resulting in samples of 592, 688 and 713 observations for the cost per patient, ΔEQ-5D and ΔBarthel models, respectively. These results are therefore only for subjects with no missing data in any particular equation.

Method 2 involved running GLM and OLS regression models again to explain variation in costs per patient and outcomes (ΔEQ-5D and ΔBarthel), respectively. Here, however, we used multiply imputed (MI) datasets (assuming that data were MAR). Up to about 38% of the data were missing and multiple imputation datasets were created to account for these missing data before running GLM and OLS regression models. These analyses focussed on imputing values for the dependent variables where the independent variables were not missing. This imputation created complete datasets where there was no missing data i.e. 717 observations for the cost per patient model and 1105 observations for the ΔEQ-5D and ΔBarthel models. The rationale for this particular imputation was to allow for direct comparison between the results obtained using this method and those produced by method 3, which comparison required that the same samples were analysed. Five sets of imputations were created. Since there was up to 38% data missing, these imputations led to point estimates that were at least $(1+0.38/5)^{-1}$ = 93% as efficient as those based on m = ∞ imputations [1].

In method 3, Heckman selection models (assuming that missing data were MNAR) were run on the log of 'cost per patient', on ΔEQ-5D and ΔBarthel using 'complete cases'. Whereas method 1 only considered cases where there was no missing data for both the dependent variable and independent variables, method 3 considers all subjects including those that had missing cost, EQ-5D or Barthel information. The sample selection used a dummy variable equal to 1 if the dependent variable was not missing and equal to 0 if it was. Using this classification, 125 out of 717 observations

were censored (missing) for the cost per patient model while 417 and 392, out of 1105 observations, were censored for the ΔEQ-5D and ΔBarthel models, respectively.

Multiple imputations were conducted in NORM [48] while the rest of the analyses were done in STATA version 8.2 [49]

Results

The results of the above analyses are presented in Table 2 for the costs per patient models and Tables 3 and 4 for the ΔEQ-5D and ΔBarthel models, respectively.

Cost per patient models

The results of the GLM regression model on complete cases (method 1) and GLM regression model on multiply imputed datasets (method 2) are similar. As shown in Table 2, all of the variables that were found to be significant in method (2) were also significant in method (1) with the exception of one (acute admission avoidance service) which was significant in model (2) only. Also, the size of coefficients for nearly all of these variables differed by less than 3.4% except the one for 'completed IC episode' which differed by about 14.4%. The sizes of the standard errors were also similar. Further, the variables significant in both models had the same direction of influence on costs per patient. On the other hand, the results obtained from the Heckman selection regression model (method 3) were much more different. A lot more variables were found to be insignificant with only two variables shown to significantly influence costs per patient. The sizes of the coefficients in the Heckman model were also different from those of the other two methods. For instance, the coefficient for 'acute admission avoidance service' was about 730 times bigger than that in obtained in method (2). The mills ratios were -3.402 and -4.506 for the Heckman selection models with and without interactions, respectively. These were both statistically significant at 95% level of significance.

Change in EQ-5D models

Here, the results from all three models/methods were broadly similar (Table 3). Nearly all of the variables that were significant in one model were also significant in the other models. The only exception were the 'duration of service provision' and 'Alternative to IC-Other*Type of IC' (both only significant in method 2), 'acute admission avoidance service' (only significant in method 3) and 'alternative to IC-Other' (significant only in models 1 and 3). The sizes of the coefficients of variables commonly significant in all models differed at most by about 22% with the standard errors differing at most by 42% (Table 3). Further, the variables significant in all three models had the same direction of influence on the change in EQ-5D. The mills ratios were -0.284 and -0.143 for the Heckman selection models with and without interactions, respectively. These were both statistically significant at 95% level of significance.

Change in Barthel models

As in the 'change in EQ-5D' models, the results obtained from all three models/methods for the change in Barthel were broadly similar (Table 4). All of the variables that were significant in one model were also significant in the other models

with the exception of 'acute admission avoidance service' (only significant in method 3) and 'Other IC Outcome' variable only significant in both method (1) and method (2). However, the differences in terms of the sizes of coefficients and standard errors of variables significant in all methods were slightly bigger in these models than in the 'change in EQ-5D' models. They differed at most by about 54% and 322% for coefficients and standard errors, respectively. The variables significant in all three models had the same direction of influence on the change in Barthel. The mills ratios were -1.662 and -0.101 for the Heckman selection models with and without interactions, respectively. These were both statistically significant at 95% level of significance

Discussion

The NECOIC dataset had up to 42% and 38% of the data on EQ-5D and Barthel scores, respectively, missing while 31% of the sample had missing cost data. Further, all but one variable in the dataset (Type of IC) had missing data ranging from 3 to 18%. This situation is common to a vast number of health service research datasets. If these missing data are simply ignored, then there is a chance that biased and underpowered results may be obtained [1, 50]. The most appropriate method of dealing with this amount of missingness therefore had to be determined [14, 51].

The evidence gathered concerning the missing cost data strongly suggested MCAR as the reason for this missingness. When MAR-based methods were used for these data, the results obtained were not significantly different from those based on the MCAR assumption. These findings seems to bear out the position held by Schafer et al [52] and David et al [53] that in many realistic applications, departures from MAR are not big enough to effectively invalidate the results of an MAR-based analysis. The use of an MNAR-based method in the costs per patient model yielded results that were so different to those obtained when either MCAR or MAR were assumed. In particular, fewer significant variables were obtained in the MNAR-based method while the sizes of the coefficients were larger. Different conclusions would therefore be reached if the MNAR assumption was made for the missing cost data. Care must therefore be taken not to apply MNAR-based methods when it is not absolutely clear that the missing data are MNAR as MNAR approaches often require assumptions that cannot be validated from the data at hand [54]. MNAR-based approaches are best implemented as sensitivity analyses so as to assess how robust results are across different analytic approaches [55].

All three mechanisms of missingness were shown to be potential causes of the missing EQ-5D and Barthel data. The results from the ΔEQ-5D and ΔBarthel models show that the choice of mechanism did not have a very significant effect on the results. Despite the sizes of the coefficients and standard errors being somewhat different, the results from all three methods were broadly similar and therefore similar conclusions could have been reached. A possible explanation for this may have been the fact that the reason for missing data could be ascribed to any one of the three mechanisms of missingness or indeed a combination of these mechanisms. While the extra information gathered during the data collection process supported the assertion that the missing data were either MCAR, MAR or MNAR, the significant mills ratios lent further support to the MNAR assumption as its significance in the selection

models indicated the presence of significant selection bias. However, selection models, even though identifiable, should be treated with caution especially when data are not MNAR [56]. Pattern mixture models would be another alternative [44, 57].

Conclusion

Many studies have emphasised the importance of determining the mechanism behind missing data before deciding on the technique to use [14, 51, 58]. This paper considered three different mechanisms that may be responsible for missing data and then discussed approaches that can be used to deal with the missing data. The results from this analysis suggest that the methods used to analyse missing data really do matter especially when one is considering whether or not to use MNAR-based methods. MAR-based methods seem to be more robust for use even in cases where data are strictly not MAR. Dealing with missing data is not easy especially as the hypothesis-based techniques for detecting the pattern of missingness are limited in that they can only be used to rule out MCAR but can not confirm this mechanism. Further, there are no hypothesis-test-based techniques available for determining if data are MAR or MNAR in cases where the missing data are irrecoverable. This therefore means that unless there is extra information gathered during the data collection exercise about the cause of missingness, there should not be any arbitrary selection of assumptions behind data missing mechanisms.

Table 1: Variables for use in economic analysis (with level of completeness)

Variable	Description	Missing (%)
Episode Characteristics		
Age	Age on 01/01/03	3
Gender	1 = female , 0 = Male	2
Live alone	1 = Individual lives alone, 0 = Otherwise	9
Barthel – Start	Barthel Score at start of IC episode	31
Barthel – End	Barthel Score at end of IC episode	38
EQ5D – Start	EQ-5D at start of IC episode	40
EQ5D – End	EQ-5D at end of IC episode	41
Change in ED-5D	Difference between EQ-5D score at end and at start of IC episode	42
Change in Barthel	Difference between Barthel score at end and at start of IC episode	41
Cost	Cost per patient	38
Descriptors of IC Services		
	Type of service required	3
Admission Avoidance service	1 = Acute Admission Avoidance service, 0 = Otherwise	
Supported Discharge service	1 = Supported discharge service, 0 = Otherwise	
Other Service	1 = Other IC Services, 0 = Otherwise	
Type of IC	1 = Residential IC, 0 = Non-Residential IC	0
	Outcome of IC episode	13
Transfer	1 = Transferred before end of IC episode, 0 = Other outcome	
Complete	1 = Completed IC episode, 0 = Otherwise	
Died	1 = Patient Died, 0 = Otherwise	
Other Outcome	1 = Alternative Outcome, 0 = Other outcome	
Stay Duration	Duration of service provision (number of days)	17
Descriptors of IC related services		
	Source of referral	3
Referral – primary	0 = Otherwise, 1 = Primary Care	
Referral – hospital	0 = Otherwise, 1 = Hospital	
Referral – social	0 = Otherwise, 1 = Social Services	
Referral – other	0 = Otherwise, 1 = Other Sources	
	Alternatives to IC services	18
Alternative – Home	0 = Else, 1 = Home	

Alternative – Hospital	0 = Else, 1 = Hospital	
Alternative – other	0 = Else 1 = Other alternative	

Table 2: Comparison of Results from three methods of Regression Analysis of costs per patient

		GLM on complete cases n = 592 [1]		*GLM on MI dataset n = 717 [2]*		*Heckman on complete cases n = 717, 125 obs censored [3]*	
	Variables	**Exp (Coeff)**	**S.E.**	**Exp (Coeff)**	**S.E.**	**Exp (Coeff)**	**S.E.**
Episode Characteristics	Age in 2003	0.996	0.003	0.997	0.002	1.000	0.012
	Gender	0.982	0.063	1.009	0.060	1.085	0.281
	Lives alone	1.052	0.059	1.047	0.056	1.106	0.275
	Barthel score at admission	0.973	0.009**	0.984	0.008*	0.884	0.061*
	EQ5D score at admission	0.973	0.090	0.935	0.087	1.400	0.435
Descriptors of IC Service	Acute Admission Avoidance Service	0.930	0.129	0.812	0.092*	6.723	0.960*
	Type of IC	3.181	0.079**	3.150	0.070**	5.146	1.274
	Transferred before end of IC episode	1.144	0.310	1.259	0.258	1.316	1.422
	Completed IC episode	2.094	0.300*	2.396	0.248**	4.611	1.318
	Other IC Outcome	2.703	0.337**	2.796	0.287**	4.374	1.475
	Patient Died (Reference. Group)						
Descriptors of IC-related Services	Referral – Primary	0.777	0.123*	0.764	0.121*	0.936	0.576
	Referral – Hospital	0.914	0.158	0.777	0.134	4.523	0.930
	Referral – Other	1.001	0.212	0.935	0.195	2.240	0.984
	Referral – Social Workers (Reference Group)						
	Alternative to IC – Other	1.053	0.079	1.058	0.077	0.508	0.451
	Alternative to IC – Home	1.121	0.074	1.058	0.070	1.112	0.329
	Alternative to IC – Hospital (Reference Group)						
Interactions	Barthel score at admission*Type of IC	1.031	0.018	1.017	0.097	1.131	0.092
	Acute Admission Avoidance Service* Type	1.214	0.163	1.217	0.136	0.579	0.752

	of IC						
	Transfer before IC end*Type of IC	1.145	0.185	1.176	0.169	1.145	0.825
	Completed Episode*Type of IC	1.152	0.195	1.112	0.162	0.240	0.952
	Other IC Outcome*Type of IC	0.717	0.708	0.583	0.534	0.773	2.846
	Patient died*Type of IC (Reference group)						
	_constant	1140.3	0.421**	951.5	0.360**	345.3	1.866**
	N		592	717			717
	Censored obs						125
	R-Squared		0.359				0.634
	Rho						0.950

* 5 % level of significance, ** 1 % level of significance; Dependent variable: cost per patient for GLM and log of cost per patient for Heckman Selection model,
IC = Intermediate care

Table 3: Comparison of Results from three methods of Regression Analysis (Change in EQ5D)

		OLS on complete cases n = 688 *[1]*		*OLS on MI dataset n = 1105 cases* *[2]*		*Heckman on complete cases n = 1105, 417 obs censored* *[3]*	
	Variables	**Coeff**	**S.E.**	**Coeff**	**S.E.**	**Coeff**	**S.E.**
Episode Characteristics	Age in 2003	0.000	0.001	0.000	0.001	0.000	0.001
	Gender	0.046	0.022*	0.051	0.018**	0.054	0.024*
	Lives alone	0.020	0.020	0.015	0.017	0.029	0.023
	Barthel score at admission	0.017	0.003**	0.017	0.002**	0.016	0.003**
	EQ5D score at admission	-0.495	0.033**	-0.479	0.026**	-0.484	0.037**
Descriptors of IC Service	Acute Admission Avoidance Service	-0.038	0.027	-0.017	0.021	0.156	0.042**
	Duration of Service Provision	0.000	0.000	0.001	0.000*	0.000	0.000

Descriptors of IC-related Services	Referral – Primary	-0.031	0.052	-0.044	0.043	-0.020	0.058
	Referral – Hospital	-0.098	0.051	-0.053	0.042	0.020	0.059
	Referral – Other	-0.003	0.078	0.059	0.065	0.013	0.086
	Referral – Social Workers (Reference Group)						
	Alternative to IC – Other	-0.063	0.031*	-0.077	0.025**	-0.077	0.030*
	Alternative to IC – Home	-0.045	0.023*	-0.028	0.019	-0.046	0.022*
	Alternative to IC – Hospital (Reference Group)						
Interactions	Gender*Type of IC	-0.048	0.053	-0.027	0.037	-0.057	0.053
	Barthel score at admission*Type of IC	0.003	0.004	-0.002	0.003	0.004	0.004
	EQ5D score at admission *Type of IC	-0.098	0.083	0.061	0.057	-0.118	0.082
	Acute Admission Avoidance Service*Type of IC	0.110	0.064	0.039	0.039	0.086	0.063
	Alternative to IC – Other *Type of IC	0.137	0.084	0.133	0.059*	0.140	0.082
	Alternative to IC – Home*Type of IC	0.086	0.106	-0.027	0.049	0.070	0.104
	Alternative to IC – Hospital *Type of IC (Reference Group)						
	_constant	0.157	0.101	0.093	0.084	0.100	0.105
	N		688	1,105			688
	Censored obs						417
	R-Squared		0.284	0.266			0.634
	Rho						0.950

* 5 % level of significance, ** 1 % level of significance;

Dependent variable: change in EQ-5D, IC = Intermediate care

Table 4: Comparison of Results from three methods of Regression Analysis (Change in Barthel)

		OLS on complete cases n = 712 [1]		*OLS on MI dataset n = 1105 cases* [2]		*Heckman on complete cases n = 1105, 392 obs censored* [3]	
	Variables	**Coeff**	**S.E.**	**Coeff**	**S.E.**	**Coeff**	**S.E.**
Episode Characteristics	Age in 2003	-0.010	0.009	-0.009	0.007	-0.011	0.009
	Gender	-0.007	0.208	0.097	0.164	0.037	0.218
	Lives alone	0.225	0.190	0.181	0.150	0.320	0.202
	Barthel score at admission	-0.318	0.028**	-0.325	0.022**	-0.305	0.030**
	EQ5D score at admission	-0.343	0.312	-0.428	0.239	-0.216	0.328
Descriptors of IC Service	Acute Admission Avoidance Service	0.103	0.218	0.060	0.167	0.728	0.337*
	Duration of Service Provision	0.008	0.003*	0.011	0.003**	0.006	0.003*
Descriptors of IC-related Services	Transfer before IC end	4.084	2.452	0.559	0.607	2.713	2.348
	Completed Episode	7.438	2.440**	3.443	0.587**	4.926	2.478*
	Other IC Outcome	6.640	2.477**	2.921	0.656**	4.727	2.432
	Patient died (Reference group)						
	Alternative to IC – Other	-1.130	0.291**	-1.076	0.221**	-1.267	0.291**
	Alternative to IC – Home	-0.709	0.223**	-0.667	0.169**	-0.669	0.219**
	Alternative to IC – Hospital (Reference Group)						
Interactions	Barthel score at admission*Type of IC	-0.071	0.050	-0.035	0.027	-0.072	0.051
	Acute Admission Avoidance Service* Type of IC	0.592	0.575	0.131	0.354	0.599	0.589
	Duration of Service Provision*Type of IC	-0.006	0.008	0.001	0.006	-0.006	0.008
	Transfer before IC end*Type of IC	-0.299	0.979	-0.160	0.411	-0.300	0.962
	Completed Episode*Type of IC	1.053	0.830	0.374	0.424	1.055	0.816
	Other IC Outcome*Type of IC	0.189	2.000	0.072	0.889	0.200	1.980
	Patient died*Type of IC (Reference group)						

	Alternative to IC – Other *Type of IC	0.796	0.793	0.968	0.543	0.795	0.778
	Alternative to IC – Home*Type of IC	2.261	1.025*	0.124	0.447	2.261	1.006*
	Alternative to IC – Hospital *Type of IC (Reference Group)						
	_constant	0.046	2.536	3.888	0.843	2.687	2.592
	N		713	1,105			713
	Censored obs						392
	R-Squared		0.278				0.634
	Rho						0.950

* 5 % level of significance, ** 1 % level of significance;

Dependent variable: change in Barthel, IC = Intermediate care

References

1. Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data.* (London: Chapman & Hall)

2. Kline, R. (1998). *Principles and practice of structural equation modeling*. (New York: The Guilford Press)

3. Chen, J. & Shao, J. (2000). Nearest neighbor imputation for survey data. *J Off Stat, 16*, 113-131.

4. Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* (New York: John Wiley & Sons)

5. Little, R. J. A. & Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* (New York: John Wiley.)

6. Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* (New York: John Wiley)

7. Foster, E. M. & Fang, G. Y. (2004). Alternative methods for handling attrition: an illustration using data from the Fast Track evaluation. *Eval Rev, 28*, 434-464.

8. Kmetic, A., Joseph, L., Berger, C., & Tenenhouse, A. (2002). Multiple imputation to account for missing data in a survey: estimating the prevalence of osteoporosis. *Epidemiology, 13*, 437-444.

9. Croninger, R. G. & Douglas, K. M. (2005). Missing Data and Institutional Research . (In P. D. Umbach (Ed.) , *Survey research. Emerging issues. New directions for institutional research #127.* (pp. 33-50) . San Fransisco: Jossey-Bass.)

10. Allison, P. (2000). *Missing data*. (Thousand Oaks, CA: Sage)

11. Schafer, J. L. (1999). Multiple imputation: a primer. *Stat Methods Med Res, 8*, 3-15.

12. Fielding, S., Fayers, P., & Ramsay, C. (2010). Predicting missing quality of life data that were later recovered: an empirical comparison of approaches. *Clin Trials, 7*, 333-342.

13. Heitjan, D. F. (1997). Annotation: what can be done about missing data? Approaches to imputation. *Am J Public Health, 87*, 548-550.

14. Curran, D., Bacchi, M., Schmitz, S. F., Molenberghs, G., & Sylvester, R. J. (1998). Identifying the types of missingness in quality of life data from clinical trials. *Stat Med, 17*, 739-756.

15. ICNET . (2005). *A National Evaluation of the Costs and Outcomes of Intermediate Care for Older People: Final Report*. . ,) . Leicester: The University of Leicester.

16. Brazier, J. E., Walters, S. J., Nicholl, J. P., & Kohler, B. (1996). Using the SF-36 and Euroqol on an elderly population. *Qual Life Res, 5*, 195-204.

17. Coast, J., Peters, T. J., Richards, S. H., & Gunnell, D. J. (1998). Use of the EuroQoL among elderly acute care patients. *Qual Life Res, 7*, 1-10.

18. Lyons, R. A., Crome, P., Monaghan, S., Killalea, D., & Daley, J. A. (1997). Health status and disability among elderly people in three UK districts. *Age Ageing, 26*, 203-209.

19. van den Bos, G. A. & Triemstra, A. H. (1999). Quality of life as an instrument for need assessment and outcome assessment of health care in chronic patients. *Qual Health Care, 8*, 247-252.

20. Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ, 13*, 873-884.

21. Dolan, P. (1997). Modeling valuations for EuroQol health states. *Med Care, 35*, 1095-1108.

22. Kind, P., Hardman, G., & Macran, S. (1999). *UK population norms for EQ-5D. Discussion paper 172*. University of York: Centre for Health Economics.

23. Murphy, R., Sackley, C. M., Miller, P., & Harwood, R. H. (2001). Effect of experience of severe stroke on subjective valuations of quality of life after stroke. *J Neurol Neurosurg Psychiatry, 70*, 679-681.

24. Post, P. N., Stiggelbout, A. M., & Wakker, P. P. (2001). The utility of health states after stroke: a systematic review of the literature. *Stroke, 32*, 1425-1429.

25. Wilkinson, P. R., Wolfe, C. D., Warburton, F. G., Rudd, A. G., Howard, R. S., Ross-Russell, R. W. et al . (1997). Longer term quality of life and outcome in stroke patients: is the Barthel index alone an adequate measure of outcome? *Qual Health Care, 6*, 125-130.

26. Mahoney, F. I. & Barthel, D. (1965). Functional Evaluation: The Barthel Index. *Md State Med J, 14:61-5.*, 61-65.

27. Wolfe, C. D., Taub, N. A., Woodrow, E. J., & Burney, P. G. (1991). Assessment of scales of disability and handicap for stroke patients. *Stroke, 22*, 1242-1244.

28. Mahoney, F. I. & Barthel, D. (1965). Functional Evaluation: The Barthel Index. *Md State Med J, 14:61-5.*, 61-65.

29. Shah, S., Vanclay, F., & Cooper, B. (1989). Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol, 42*, 703-709.

30. Musil, C. M., Warner, C. B., Yobas, P. K., & Jones, S. L. (2002). A comparison of imputation techniques for handling missing data. *West J Nurs Res, 24*, 815-829.

31. Fielding, S., Fayers, P. M., & Ramsay, C. R. (2009). Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. *Health Qual Life Outcomes, 7*, 57.

32. Kaambwa, B., Bryan, S., Barton, P., Parker, H., Martin, G., Hewitt, G. et al . (2008). Costs and health outcomes of intermediate care: results from five UK case study sites. *Health Soc Care Community, 16*, 573-581.

33. McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models*. (London: Chapman & Hall)

34. Manning, W. G. & Mullahy, J. (2001). Estimating log models: to transform or not to transform?
173. *J Health Econ, 20*, 461-494.

35. Altman, D. (1991). *Practical statistics for medical research*. (London: Chapman & Hall)

36. Cantoni, E. & Ronchetti, E. (2006). A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures. *J Health Econ, 25*, 198-213.

37. Duan, N. (1983) Smearing estimate a nonparametric retransformation method. J.Amer.Statist.Assoc 78: 605-610.
Ref Type: Journal (Full)

38. Kilian, R., Matschinger, H., Loeffler, W., Roick, C., & Angermeyer, M. C. (2002). A comparison of methods to handle skew distributed cost variables in the analysis of

the resource consumption in schizophrenia treatment. *J Ment Health Policy Econ, 5*, 21-31.

39. Raymond, M. R. & Roberts, D. M. (1987). A comparison of methods for treating incomplete data in selection research. *Educational and PsychologicalMeasurement, 47*, 13-26.

40. Raboud, J. M., Montaner, J. S., Thorne, A., Singer, J., & Schechter, M. T. (1996). Impact of missing data due to dropouts on estimates of the treatment effect in a randomized trial of antiretroviral therapy for HIV-infected individuals. Canadian HIV Trials Network A002 Study Group. *J Acquir Immune Defic Syndr Hum Retrovirol, 12*, 46-55.

41. Wothke, W. (1998). Longitudinal and multi-group modeling with missing data. (In T. D. Little, K. U. Schnabel, & J. Baumert (Eds) , *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples* Mahwah, NJ: Lawrence Erlbaum Publishers.)

42. Allison, P. D. (2000). Multiple imputation for missing data: a cautionary tale. *Sociological methods and Research, 28*, 301-309.

43. Barnard, J. & Meng, X. L. (1999). Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res, 8*, 17-36.

44. Hedeker, D. & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*, 64-78.

45. Schafer, J. L. & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*, 545-571.

46. Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement, 5*, 475-492.

47. Gujarati, D. (1995). *Basic Econometrics*. (New York: McGraw-Hill, Inv)

48. Schafer, J. L. (1999). *NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2. Software for Windows 95/98/NT*. Retrieved May 10, 2011 from http://www.stat.psu.edu/jls/misoftwa.html.

49. StataCorp L.P. (2004) Intercooled Stata 8.2 for Windows. TX, College Station.

50. Roderick, P., Low, J., Day, R., Peasgood, T., Mullee, M. A., Turnbull, J. C. et al . (2001). Stroke rehabilitation after hospital discharge: a randomized trial comparing domiciliary and day-hospital care. *Age Ageing, 30*, 303-310.

51. Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* . (Hillsdale, NJ: Erlbaum)

52. Schafer, J. L. & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*, 147-177.

53. David, M., Little, R. J. A., Samuhel, M. E., & Triest, R. K. (1986). Alternative Methods for CPS Income Imputation. *Journal of the American StatisticalAssociation, 81*, 29-41.

54. Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. (New York: Springer)

55. Mallinckrodt, C. H., Sanger, T. M., Dube, S., DeBrota, D. J., Molenberghs, G., Carroll, R. J. et al . (2003). Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biol Psychiatry, 53*, 754-760.

56. Glynn, R. J., Laird, N. M., & Rubin, D. B. (1986). Drawing Inferences from Self-selected Samples. (In H. Wainer (Ed.) , *Selection modelling versus mixture modelling with nonignorable nonresponse* (pp. 115-142) . New York: Springer-Verlag.)

57. Hogan, J. W. & Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Stat Med, 16*, 239-257.

58. Orme, J. G. & Reis, J. (1991). Multiple regression with missing data. *Journal of Social Service Research,* 61-91.