ELSEVIER

Journal of Clinical Epidemiology 57 (2004) 147–153

Journal of Clinical Epidemiology

Selection bias found in interpreting analyses with missing data for the prehospital index for trauma

Lawrence Joseph[a,b,*], Patrick Bélisle[a], Hala Tamim[c], John S. Sampalis[a]

[a]*Division of Clinical Epidemiology, Montreal General Hospital, Department of Medicine, 1650 Cedar Avenue, Montreal, Quebec, H3G 1A4, Canada*
[b]*Department of Epidemiology and Biostatistics, 1020 Pine Avenue West, McGill University, Montreal, Quebec, H3A 1A2, Canada*
[c]*University of Balamand, Faculty of Health Sciences, Ashrafieh, Youssef Sorsok Street, P.O. Box 166378-6417, Beirut, Lebanon*

Accepted 8 August 2003

Abstract

Objective: To evaluate the effects of missing data on analyses of data from trauma databases, and to verify whether commonly used techniques for handling missing data work well in theses settings.

Study Design and Setting: Measures of trauma severity such as the Pre-Hospital Index (PHI) are used for triage and the evaluation of trauma care. As conditions of trauma patients can rapidly change over time, estimating the change in PHI from the arrival at the emergency room to hospital admission is important. We used both simulated and real data to investigate the estimation of PHI data when some data are missing. Techniques compared include complete case analysis, single imputation, and multiple imputation.

Results: It is well known that complete case analyses and single imputation methods often lead to highly misleading results that can be corrected by multiple imputation, an increasingly popular method for missing data situations. In practice, unverifiable assumptions may not hold, meaning that it may not be possible to draw definitive conclusions from any of the methods.

Conclusion: Great care is required whenever missing data arises. This is especially true in trauma databases, which often have much missing data and where the data may not missing at random.

Keywords: Bayesian methods; Missing data; Multiple Imputation; Polytomous regression; Prehospital Index; Trauma

1. Introduction

Trauma is a major cause of death and disability [1,2]. A crucial component of the immediate post-injury phase is the decision by on-site emergency medical personnel as to the level of care required. Designated Level I Trauma Centers provide a more timely and complete range of trauma care at higher cost than Level II and III Centers.

Various triage protocols have been proposed [3–5]. Measures of trauma severity, such as the Prehospital Index (PHI) [6] and the Revised Trauma Score [7], are used both as triage tools and in trauma research. Research uses include calculating summaries of the types of patients that are transported to the various levels of care, and tracking changes in patient condition over time. Changes over time are important, because they indicate whether the care provided has resulted in improvements to the patients' conditions.

The nature of emergency-scene medical care makes it common for trauma databases to have large amounts of missing data. Data are said to be missing completely at random (MCAR) if the probability that each data item is observed is independent of the missing data values themselves [8]. This often is not the case. For example, in trauma databases, a plausible reason for missing data is the severity of the injury, meaning that the probability that a PHI component is missing is directly related to its value. A weaker condition is that the data are missing at random (MAR). MAR data occur if the probability that an item is missing depends on other observed data values, but not on the values of the unobserved missing data items, given the observed data. This is more plausible than MCAR data in trauma research, for example, because other data collected on the scene may be highly correlated with the missing data, so that the missing values themselves are uninformative (or very nearly so) after accounting for what has been observed. Both MCAR and MAR missing data mechanisms can be termed ignorable, because valid inferences can be derived even if the missing data mechanism is ignored (i.e., not explicitly considered in the analysis).

Non-ignorable data arise when the missing data mechanism depends on unobserved data, even after accounting for the information in the observed data. Incorrectly assuming

* Corresponding author. Montreal General Hospital, Division of Clinical Epidemiology, 1650 Cedar Avenue, Montreal H3G1A4, Canada; Tel.: 514-934-1934, ext. 44713; fax: 514-934-8293.

E-mail address: Lawrence.Joseph@mcgill.ca (L. Joseph).

0895-4356/04/$ – see front matter
doi: 10.1016/j.jclinepi.2003.08.002

ignorability can produce biased estimates, and unverifiable assumptions typically must be made about the missing data mechanism to account for the missing items. In practice, it is often very difficult to determine the missing data mechanism. Various reviews are available [8–12] that define MCAR, MAR, and ignorability in more technical terms.

Because trauma data are not likely to be missing completely at random, serious biases may result if an analysis is performed using only those cases with no missing data, known as complete case analysis, which is currently common in trauma (and other) research [13,14]. Furthermore, the typically large volume of missing data means that complete case analysis can be statistically inefficient, because data collected on cases with one or more missing items are not used. Other simple methods of handling missing data, such as a single imputation of the overall mode, mean, or median value in the population or sample, can also result in biased estimates, and can provide confidence intervals that are too narrow.

More sophisticated techniques developed to handle missing data include observed data maximum likelihood [12,15], generalized estimating equations [16,17], and latent data models, for example, through data augmentation [18]. Multiple imputation [8] is an attractive alternative in missing data situations, because it is easy to use, can be applied to simple or complex sampling designs, and provides valid inferences provided the assumptions behind the model are correct. Directly checking these assumptions is usually impossible, however, because the verification depends on the data that are, in fact, missing. One way to proceed is to obtain the missing data on a random subset of non-responders [19], but this is often not practical. For example, in trauma research, one can never obtain missing PHI data, because the patients' condition continually evolves through time. Therefore, one might carry out indirect verifications, perhaps based on other similar but more complete data sets, with the hope that the data set of interest is sufficiently similar for the assumptions to hold there as well.

The Quebec Trauma Registry [20] was established in 1993, and includes information on patient demographics, injury description, injury severity scores, prehospital and in-hospital care, complications and survival. Injury severity measures can be divided into anatomical [21,22] and physiologic scales. Anatomic scales are obtained from physical examination, investigative procedures, surgical interventions, and, in fatal cases, postmortem examination, so cannot typically be used for triage. Physiologic scales [7,23], including the PHI [6,24], measure the acute response to injury. Physiologic scales can be used for triage, because, in principle, all data may be quickly collected at the scene of the injury.

The PHI, which we focus on in this article, consists of only four components: systolic blood pressure, pulse rate, respiratory status, and level of consciousness. The PHI is calculated by assigning a value between 0 and 5 to each of the above vital signs, with 0 indicating normal functioning, and 5 indicating maximum physiologic deterioration. An additional four points are added in the presence of penetrating abdominal or thoracic trauma. The PHI ranges from 0 to 24 (see Table 1). Values of the PHI between 0 and 3 are usually considered to indicate minor injuries, values of 9 or greater indicate major trauma, while intermediate values point to moderate injuries [22]. When the PHI has missing components, it is most common in trauma research to simply delete the case from any subsequent analyses [13,14].

An important component of the evaluation of trauma care is the change in severity from arrival at the emergency room to hospital admission. We use estimation of this parameter to compare three techniques for missing data: complete case analysis, and single and multiple imputation. There are large amounts of missing data in the Quebec Trauma Registry, and we do not know whether the missing data are ignorable or not. The validity of the ignorability assumption is unknown, as verification depends on the missing data components themselves. Instead, we will carry out an empirical simulation based on a complete PHI data to gauge whether it is reasonable to expect near-ignorability or not. Briefly, we will randomly delete values in our complete data set by three different missing data mechanisms, and use multiple imputation to see if we can derive inferences similar to those available from the complete data set. The main question is whether there is sufficient information in the observed data to recapture enough of the missing information for reasonably valid inferences, despite nonignorability of the missing data mechanisms. If yes, one might hope that multiple imputation could be useful in adjusting for bias due to missing data in other trauma databases, in particular, the subset of the Quebec Trauma Registry of interest here.

2. Methods

For the empirical simulations, we used a data set containing complete PHI data on 2,800 subjects. This included all trauma subjects transported via Montreal's ambulance

Table 1
Scoring system of the Prehospital Index

Parameter	Value	Score
Systolic blood pressure	>100	0
	86–100	1
	75–85	2
	<75	5
Pulse rate	51–120	0
	>120	3
	<50	5
Respiratory status	Normal	0
	Laboured/shallow	3
	<10 needs intubation	5
Level of consciousness	Normal	0
	Confused/combative	3
	No intelligible words	5

Final score is the sum of the individual scores in each of the four categories. Add four points for penetrating abdominal or thoracic injuries

service, Urgence Santé, to a hospital between the years 1993 to 1995. We simulated missing data according to two mechanisms:

1. Missing completely at random mechanism: we randomly and independently treated each of the five PHI components from each subject as missing with a rate of 20%.
2. Data-dependent mechanism: values of the PHI components indicating more severe trauma were assigned higher probabilities of being treated as missing. In particular, we used the following two sets of probabilities:

   Probability Set #1:

   $P\{$ a value of 0 or 1 is coded as missing $\} = .175$,

   $P\{$ a value of 2 or 3 is coded as missing $\} = .25$, and

   $P\{$ a value of 4 or 5 is coded as missing $\} = .35$.

   Probability Set #2:

   $P\{$ a value of 0 or 1 is coded as missing $\} = .15$,

   $P\{$ a value of 2 or 3 is coded as missing $\} = .25$, and

   $P\{$ a value of 4 or 5 is coded as missing $\} = .50$.

In both cases, these numbers were selected such that the overall rate of missing data components would again be 20%. This rate was selected to be typical of trauma databases, where rates up to 35% are not unusual.

We compared the true value of the mean PHI total score to the estimates produced by the three missing data methods. For the data missing completely at random, we expect that complete case analysis will produce unbiased but inefficient estimates. Because the rates of missing data were applied independently to each component, and because there are five components per subject in the PHI, the proportion of subjects with one or more missing data components would be expected to be approximately $1 - 0.8^5$, or 67% of all cases.

Because our data-dependent mechanism above directly defines the probability of a given component being missing conditional on the missing value itself, our mechanism is by definition non-ignorable. If, however, there are high correlations among the five PHI components within each subject, then the dependence of the missing data components on their unobserved values given the non-missing components can be considerably reduced, producing "near-ignorability" in practice. If our complete data model (described below) incorporates these correlations, then much of the missingness probability associated with the missing data component will be accounted for by the relationship between the observed and missing data, so that little of the missingness probability is left to be explained by the missing value itself. Looking at the Spearman correlation matrix of the five components from the PHI, we find that all 10 pairwise correlations range from a low of 0.66 to a maximum of 0.86, providing preliminary evidence that this may be the case for our data.

We applied the following methods of analysis to each of the data sets created by the missing data mechanisms described above.

2.1. Complete case analyses

For the complete case analyses, we deleted from consideration all subjects with one or more missing data components. The average PHI score, then, was simply the average score among subjects with no missing data. This is the default type of analysis in the vast majority of statistical packages, although some packages have recently added a multiple imputation option [25,26].

2.2. Single imputation of the mode

One simple way to correct for missing data that has often been used in trauma research is to fill in each missing data item with the modal value of that component. One then has a "complete" data set, from which inferences can be drawn in the usual fashion. It is particularly tempting to use this method in trauma research, because the default complete case analysis is very inefficient, and single imputation of the mode seems plausible, as there is usually one value in each PHI component that occurs much more frequently than the others, typically the value indicating the "normal" state. For example, the median value of each of the five PHI components was 0, indicating that over 50% of all PHI components were simply scored as 0. Therefore, one could expect to impute a correct value a relatively high percentage of the time. Nevertheless, as Rubin [17] points out, a high "hit rate" (i.e., a high percentage of correctly imputed items) does not necessarily ensure valid inferences.

2.3. Multiple imputation

The multiple imputation method we used was embedded into a data augmentation or Gibbs sampler algorithm [27]. The algorithm can be conceptualized as proceeding iteratively by alternating between steps that estimate the values of polytomous regression parameters [28] for models that predict individual PHI components from the other components, and using these models to impute the missing PHI data. Technically, this is roughly equivalent to integrating out the missing data [8,11,18]. The details are as follows.

Start by arbitrarily filling in a value for each missing component, providing an initial "complete" data set. We used the modal value of each component for this purpose. Using the above "complete" data set, estimate parameters in a polytomous regression model that predicts one component of the PHI from the remaining components and other covariates. Thus, for example, one could start with a model that uses pulse rate, respiratory status, level of consciousness, and penetrating injury data as independent variables to predict systolic blood pressure. Covariates such as age and sex,

typically available in trauma databases, could also be added to the models. Letting Y generically denote the dependent variable (e.g., systolic blood pressure) in these models, and X denote the vector of independent variables (we used the other four PHI variables, here coded using dummy variables for each possible value of each variable for one model), the model can be represented as the set of equations:

$$\ln\left[\frac{P(Y=j|X)}{P(Y=0|X)}\right] = \beta_{0j} + \beta_j \times X,\ j = 1,\ldots,\ k-1,$$

where β_{0j} and β_j are the intercept and vector of regression coefficients for the jth category of the outcome Y, and where k is the total number of categories for Y. Use this new model to reimpute the missing systolic blood pressure PHI components. Repeat the above steps for each of the five PHI components in turn, so that five polytomous regression models are fit. Using these models provides a second "complete" data set. In Gibbs sampler terminology, the above is termed one cycle through the full conditional distributions of the unknown parameters, and this set of full conditional distributions defines a unique multivariate model for the data [29]. A large number of cycles are run, and final inferences are based on the resulting random values. Inference about the mean PHI value in the population could be calculated by using the mean across sets of "complete" data imputations as a point estimate, whose variance can be conceptualized as the sum of the average within imputation variance and the between imputation variance [8]. Some care is required to ensure that the model converges. We used the method of Raftery and Lewis [30] to determine the number of iterations and burn-in required, and ran each analysis several times from different starting values to ensure convergence to the correct posterior distributions. All Gibbs samplers were run with burn-in of 100 iterations followed by 5,000 or 10,000 iterations for inferences, using BUGS software [31]. Diffuse prior distributions were used for all parameters, so that posterior distributions are influenced almost exclusively by the data. The final inferences were calculated as summaries (means and equal-tailed 95% credible intervals) of the Gibbs output.

For our analysis of the change in PHI from emergency room to hospital, our first multiple imputation model was similar to that used in the empirical simulation, with separate models applied to fill in the missing data from the emergency room and that collected at admission to the hospital. Because data from two time points were available, we added the corresponding PHI value at the other time as a covariate. For example, in imputing missing blood pressure data for the PHI in the emergency room, we added blood pressure at admission as a possible predictor. The second imputation model added the covariates age and sex to the model as well. A parsimonious model is desirable, given the large number of possible cross categories across all imputed variables and their predictors. Therefore, we used approximate Bayes Factors as calculated by the Bayesian Information Criterion [32] to determine final models to use for each submodel in the full imputation program. We also collapsed some very small categories into neighboring larger categories when this resulted in little or no loss of predictive power according to the BIC.

3. Results

3.1. Empirical simulations

Of the 2,800 cases with complete PHI data, the average age was 35.8 (SD = 20.2) years, and the subjects were 64.6% male. The true observed average PHI value among these 2,800 cases was 3.68, with 95% credible interval (CI) of (3.40, 3.96). Table 2 summarizes the results under the various scenarios considered within our simulations.

As expected, when the data are missing completely at random, using a complete case analysis to estimate the mean PHI is valid but inefficient, in that the estimated mean is close to the mean of the full 2,800 cases, but the CI width almost doubles compared to the analysis that includes full data from all 2,800 cases. When the missing mechanism

Table 2
Results from the empirical simulation

Method	Mean PHI	SD	95% CI	CI width	Bias	% Correct
No missing data (n = 2800)	3.68	7.65	3.40, 3.96	0.56	—	—
Missing completely at random mechanism						
Complete case (n = 956)	3.63	7.53	3.15, 4.11	0.96	0.05	—
Single imputation	2.95	6.32	2.72, 3.19	0.47	0.73	83.2
Multiple imputation	3.83	7.71	3.54, 4.12	0.57	0.15	87.7
Data dependent missing mechanism #1						
Complete case (n = 940)	1.82	5.16	1.49, 2.15	0.66	1.86	—
Single imputation	2.46	5.45	2.26, 2.67	0.41	1.22	72.3
Multiple imputation	3.59	7.62	3.31, 3.88	0.57	0.08	86.1
Data dependent missing mechanism #2						
Complete case (n = 1053)	1.01	3.25	0.81, 1.20	0.39	2.67	—
Single imputation	2.01	4.48	1.84, 2.17	0.33	1.67	61.8
Multiple imputation	3.48	7.45	3.20, 3.76	0.56	0.20	83.0

% Correct refers to the percentage of correctly imputed items.

is data dependent, however, complete case analysis performs very poorly, with the true mean value located very far from the upper limit of the 95% CI. Therefore, not surprisingly, complete case analysis cannot be recommended when analyzing PHI data, even when the data are completely missing at random.

Single imputation of the modal value also performed poorly under all missing data mechanisms, with a bias of at least 0.73 points on the PHI scale, and with none of the 95% CI's capturing the true value. This occurred despite very high percentages of correctly imputed components.

Multiple imputation performed very well under both types of missing data mechanisms. The correct mean was well within the 95% CI in all cases, and the increase in CI width over the full ($n = 2800$) analysis was minimal. Note that even under an MCAR mechanism, the inferences were better for multiple compared to single imputation despite similar percentages of correct imputations. Because our multiple imputation model used only the information contained in the non-missing PHI components, it is clear that within each patient, the non-missing components are predictive of the missing components in a meaningful way. The good performance of multiple imputation was to be expected, because the high correlations between individual PHI components (ranging from a low of 0.66 to a high of 0.86) implied "near non-ignorability" of our model, as discussed above.

We next apply the above three techniques to estimating the change in PHI scores recorded in the emergency room to those measured upon admission to the hospital. The above simulations suggest that only multiple imputation performs adequately among the methods tested, but because the non-simulated missing data analyzed below are again likely to be non-ignorable, there is no absolute guarantee that multiple imputation will again perform as well as it did in the above simulations. We therefore tried two different multiple imputation models, to further investigate robustness of our final results to bias from missing data.

3.2. Estimating the emergency room to hospital admission change in PHI

Data from 3,194 subjects were analyzed. The average age was 49.8 (SD = 22.4) years, with 62.5% male. Table 3 summarizes the results of the analyses of these data.

The complete case analysis for the change in PHI values from the emergency room to hospital admission used only 374 (11.7%) of the subjects' data, and both complete case and single imputation methods estimated both the emergency room and admission average PHI scores to be approximately one (to the nearest integer), thereby estimating very small changes from emergency room to admission. The multiple imputation models both estimated substantially higher scores at both time points, with the model that omitted age and sex estimating higher values compared to the model that included these additional covariates.

Table 3
Results from the analyses of the change in PHI from admission to the emergency room

	Emergency room	Admission	Change
Complete case analysis			
n	569	672	374
Mean PHI	0.87	1.14	0.09
SD	2.54	2.90	1.43
95% CI	0.66, 1.08	0.92, 1.36	−0.05, 0.24
Single imputation ($n = 3194$)			
Mean PHI	0.75	0.95	0.20
SD	1.96	2.32	1.65
95% CI	0.68, 0.82	0.87, 1.03	0.14, 0.26
Multiple imputation (age and sex omitted) ($n = 3194$)			
Mean PHI	1.53	1.96	0.43
SD	3.20	3.73	2.65
95% CI	1.41, 1.66	1.85, 2.09	0.29, 0.585
Multiple imputation (age and sex included) ($n = 3194$)			
Mean PHI	1.36	1.54	0.18
SD	2.90	3.22	1.66
95% CI	1.30, 1.44	1.47, 1.62	0.12, 0.23

The mean PHI change for the complete case analyses does not correspond to the observed difference between the mean at admission and the mean recorded at the emergency room, because each column is based on different numbers of cases.

The estimate of change in PHI was larger in the model that excluded age and sex compared to the model that included these covariates, with point estimates of 0.18 and 0.43 points, respectively. Interestingly, the 95% credible intervals for the mean change from the two different multiple imputation models did not overlap. As age and sex appeared to be good predictors of most of the PHI components at both time points, one might tend to prefer results from this model, but the fact that two reasonable imputation models provide somewhat different results raises the issue of how to interpret the results overall, and whether either model is providing valid inferences. Looking at the pairwise correlations between the PHI components among the complete cases in this second data set, we were surprised to find that they were much lower than in the complete data set. Across the 20 possible bivariate correlations of the five components within each of the admission and emergency times, the range of the Spearman correlation coefficients was from 0.03 to 0.72, with all but three being less than 0.25. This is in contrast with the much higher values we found in the first data set. The 2,800 complete cases analyzed in the empirical simulation were also younger, and therefore possibly represent a different class of trauma patients compared to the 3,194 subjects analyzed here. In other words, patients admitted to hospital following emergency care may differ from patients treated and then discharged from emergency.

4. Discussion

Confronted with a data set with much missing data, we decided to use multiple imputation to adjust for possible bias.

In the absence of any direct evidence that our data were MAR, however, we wondered how certain we could be that our final inferences were correct. We decided to use two additional checks to help verify whether multiple imputation would likely work well. Using a similar but complete PHI data set, the empirical simulations we performed confirmed the well-established result that multiple imputation can work well when other techniques fail, giving us initial confidence to proceed with our main analyses. Our assumption was that if the mean PHI score could be well estimated at each of two time points, then the difference in PHI between these time points would also be well estimated.

The dependence of our results upon which imputation model is selected raises questions about which set of final inferences is to be preferred. This is especially true because a case could be made for either model to be chosen. The first model used PHI data alone, so was a closer replication of our simulation model, which worked well. The second model added reasonable covariates, and a simulation study has suggested that larger models tend to outperform smaller models [33], at least in the context of data missing at random. This analysis, however, carries with it the extra assumption that the relation between these covariates is the same in those subjects with and without missing data, which is not guaranteed.

We conclude the following:

1. Different reasonable multiple imputation models can sometimes lead to different inferences. Robustness to model choice should be considered. Although in non-missing data situations a best model (or set of best models) can be chosen, the choice is less certain here, owing to the unverifiable extra assumption that the model fit using the non-missing data also adequately models the missing data items.
2. Because the assumptions are difficult to directly verify, indirect verifications based on similar data sets may be useful. This, too, can have drawbacks, as a model that works well in one database may work less well in another.
3. If different reasonable models provide different inferences, one needs to use clinical judgement as to whether the differences are important enough to invalidate the analyses or not.

In our example, even though the confidence intervals for the PHI difference from the two models did not overlap, taking the extreme range from the two intervals given us an "inclusive interval" of (0.12, 0.585). If this interval is sufficiently narrow to draw conclusions about changes in the PHI, then reasonable confidence can be placed in the final conclusions. If, however, there is a different conclusion based on whether the PHI difference is above or below 0.25, say, then definitive conclusions cannot be drawn from this data set.

Of course, these conclusions do not apply uniquely to trauma databases, but may carry over to other types of databases with missing data. Although multiple imputation may be helpful in some cases, it is clearly not always sufficient. This is an important point, because multiple imputation will be increasingly used, as evidenced by its recent addition to popular standard statistical packages [25,26] and articles in the epidemiologic literature recommending its use [34,35].

In summary, complete case analyses require data that are completely missing at random, often not the case in epidemiologic studies, and almost never true for trauma research. Single imputation models are often no better, and provide confidence intervals that are too narrow. Multiple imputation models can provide improved confidence interval coverage, but their validity depends on unverifiable ignorability conditions, which may not hold even if preliminary investigations in similar databases suggests that they are plausible. Recent work on models that do not assume MAR generally require explicit models for the missing data mechanism, but may be of help in some cases. See Schafer and Graham [19] for a recent summary.

Acknowledgments

This work was partially supported by a grant from The Natural Sciences and Engineering Council of Canada. Lawrence Joseph and John Sampalis are both Senior Scientists of the Canadian Institute of Health Research.

References

[1] Uzych L. Trauma care systems. Am J Emerg Med 1990;8:71–5.
[2] Max W, Rice DP, MacKenzie EJ. The lifetime cost of injury. Inq 1990;27:332–43.
[3] Gormican SP. CRAMS scale: field triage of trauma victims. Ann Emerg Med 1992;11:132–5.
[4] Champion HR, Sacco WJ, Carnazzo AJ, Copes W, Fouty WJ. Trauma score. Crit Care Med 1981;9:672–6.
[5] Baxt WG, Jones G, Fortlage D. The trauma triage rule: a new, resource-based approach to the prehospital identification of major trauma victims. Ann Emerg Med 1990;19:1401–6.
[6] Koehler JJ, Baer LJ, Malafa SA, Meindertsma MS, Navitskas NR, Huizenga JE. Prehospital Index: a scoring system for field triage of trauma victims. Ann Emerg Med 1986;15:178–82.
[7] Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the Trauma Score. J Trauma 1989;29:623–9.
[8] Rubin D. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
[9] Schafer JL. Analysis of incomplete multivariate data. London: Chapman and Hall; 1997.
[10] Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002;7:147–77.
[11] Rubin D. Multiple imputation after 18+ years. J Am Stat Assoc 1996; 91:473–89.
[12] Little R, Rubin D. Statistical analysis with missing data. New York: Wiley; 1987.
[13] Ohsfeldt RL, Morrisey MA, Johnson V, Treat R. Simplifying the assessment of rural emergency medical service trauma transport. Med Care 1996;34:1085–92.
[14] Kuhls D, Malone D, McCarter R, Napolitano L. Predictors of mortality in adult trauma patients: the physiologic trauma score is equivalent to the trauma and injury severity score. J Am Coll Surg 2002;194: 695–704.

[15] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B 1997;39: 1–38.
[16] Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. Biometrics 1988;44:1049–60.
[17] Troxel AB, Lipsitz SR, Brennan TA. Weighted estimating equations with nonignorably missing response data. Biometrics 1997;53:857–69.
[18] Tanner MA, Wong WH. The calculation of posterior densities by data augmentation. (with discussion). J Am Stat Assoc 1987;82:528–50.
[19] Glynn RJ, Laird NM, Rubin DB. Multiple imputation in mixture models for nonignorable nonresponse with followups. J Am Stat Assoc 1993;88:984–93.
[20] Sampalis JS, Denis R, Frechette P, Brown R, Fleiszer D, Mulder D. Direct transport to tertiary trauma centers versus transfer from lower level facilities: impact on mortality and morbidity among patients with major trauma. J Trauma 1997;43:288–95.
[21] Greenspan L, McLellan B, Greig H. Abreviated Injury Scale and Injury Severity Score: a scoring chart. J Trauma 1985;25:60–4.
[22] Baker SP, O'Neill B. The injury severity score: an update. J Trauma 1976;16:882–5.
[23] Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974;2:81–4.
[24] Koehler JJ, Malafa SA, Hillesland J, Baer LJ, Rogers RN, Navitskas NR, Briggs D, Simpson D, Roller B, Lilleboe P. A multicenter validation of the prehospital index. Ann Emerg Med 1987;16:380–5.
[25] SAS/STAT software. Changes and enhancements release 8.2. Cary, NC: SAS Institute Inc.; 2001.
[26] Base SPSS. 10 Users guide. Chicago: SPSS Inc.; 2001.
[27] Gelman A, Carlin B, Stern H, Rubin D. Bayesian data analysis. New York: Chapman Hall; 1995.
[28] Albert JA, Chib S. Bayesian analysis of binary and polychotomous response data. J Am Stat Assoc 1993;88:669–79.
[29] Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 1990;85:398–409.
[30] Raftery A, Lewis S. How many iteration in the Gibbs sampler? In: Bernardo J, Berger J, Dawid A, et al., editors. Bayesian statistics 4. Oxford: University Press; 1992. p. 763–73.
[31] Spiegelhalter D, Thomas A, Best N, Gilks W. BUGS 0.6—Bayesian inference using Gibbs sampling. Cambridge: MRC Biostatistics Unit; 1997.
[32] Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc 1995;90:773–95.
[33] Collins L, Schafer J, Kam C. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods 2001;6:330–51.
[34] Kmetic A, Joseph L, Berger C, Tenenhouse A. Multiple imputation to account for missing data in a survey: estimating the prevalence of osteoporosis. Epidemiology 2002;13:437–44.
[35] Clark T, Altman D. Developing a prognostic model in the presence of missing data. An ovarian cancer case study. J Clin Epidemiol 2003;56:28–37.