Published in final edited form as:

J Nurs Scholarsh. 2008 ; 40(4): 373–378. doi:10.1111/j.1547-5069.2008.00252.x.

Managing the Common Problem of Missing Data in Trauma Studies

Tessa Rue, MS[Biostatistician],

Department of Biostatistics, University of Washington, Seattle

Hilaire J. Thompson, PhD, APRN, BC, CNRN, Xi & Psi-at-Large[Assistant Professor],

Biobehavioral Nursing and Health Systems, University of Washington, Seattle

Frederick P. Rivara, MD, MPH[Professor],

Pediatrics, University of Washington, Seattle

Ellen J. Mackenzie, PhD[Professor], and

Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD

Gregory J. Jurkovich, MD[Professor]

Surgery, University of Washington, Seattle, WA

Abstract

Purpose—To provide guidance for managing the problem of missing data in clinical studies of trauma in order to decrease bias and increase the validity of findings for subsequent use.

Organizing Construct—A thoughtful approach to missing data is an essential component of analysis to promote the clear interpretation of study findings.

Methods—Integrative review of relevant biostatistics, medical and nursing literature, and case exemplars of missing data analyses using multiple linear regression based upon data from the National Study on the Costs and Outcomes of Trauma (NSCOT) was used as an example.

Findings and Conclusions—In studies of traumatically injured people, multiple imputed values are often superior to complete case analyses that might have significant bias. Multiple imputation can improve accuracy of the assessment and might also improve precision of estimates. Sensitivity analyses which implements repeated analyses using various scenarios may also be useful in providing information supportive of further inquiry. This stepwise approach of missing data could also be valid in studies with similar types or patterns of missing data.

Clinical Relevance—In interpreting and applying findings of studies with missing data, clinicians need to ensure that researchers have used appropriate methods for handling this issue. If suitable methods were not employed, nurse clinicians need to be aware that the findings may be biased.

Keywords

data collection; multiple imputation; sensitivity analyses; bias; precision

Throughout the world, a major and increasing cause of death and disability is traumatic injury (World Health Organization, 2001). The loss of productivity from injury in the United States in 2000 alone was estimated at \$326 billion (Finkelstein, Corso, & Miller, 2006).

Recent efforts, including those by nurse scientists (Richmond et al., 2007; Scheetz, 2005; Sommers et al., 2006; Thompson et al., 2008), have been focused on injury prevention and on improving systems of trauma care in order to reduce morbidity and mortality following injury (Sommers, 2006). Multivariate and longitudinal research approaches are generally taken in order to suitably address questions posed by these issues in injury research. The problem of missing data is frequently encountered as a result by trauma researchers and must be appropriately handled. Additionally, because of the sources and types of data used in injury research, the problem of missing data might be compounded. If missing data issues are addressed improperly, they might lead to inefficient, underpowered analyses or biased estimates (Joseph, Belisle, Tamim, & Sampalis, 2004).

The purpose of this paper is to provide guidance for managing the problem of missing data in clinical studies of trauma in order to minimize bias and increase the validity of findings for subsequent translation. An integrative review and simulated analysis based upon data from the National Study on the Costs and Outcomes of Trauma (NSCOT) was used as an example.

Missing data is an omnipresent problem in clinical research, but this problem is often compounded in trauma research where the prevalence of missing data is often much higher. The first basis for the scope of the problem in trauma research is the data source. The severity of injury and nature of emergent care being provided to patients might interfere with data collection (Joseph et al., 2004). Data might also be missing for a variety of reasons including: missing reports concerning prehospital care (Newgard, 2006) or having an item that is not fully testable in a particular patient (e.g., pupillary response on a patient with trauma to one eye). Traumatic injury, by its very nature, is an unexpected event, thus the injured person might be unable to provide needed information and a proxy informant might not be available (Moore et al., 2005). People at increased risk for trauma include those with comorbidities such as substance abuse which might make follow-up more difficult in longitudinal studies (Gentilello, Donovan, Dunn, & Rivara, 1995; Holavanahalli et al., 2006; Nilsen, Holmqvist, Nordqvist, & Bendtsen, 2007).

The second area that contributes to the problem of missing data in trauma research is that many studies are conducted as secondary analyses of available data. To facilitate studies of injury, researchers have several sources of secondary clinical data available to them including local and state hospital trauma registries, national hospital data from the Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project (HCUP), and national trauma databases such as the American College of Surgeon's National Trauma Data Bank. When considering the use of secondary sources of data, researchers must balance the knowledge to be gained against the awareness of incomplete cases available for analysis. Thus, regardless if the study is conducted prospectively or retrospectively, the problem of missing data is an issue that injury researchers must be able to aptly deal with on a routine basis in order to ensure that their results are precise, valid, and reliable.

The extent and pattern of missing data may vary greatly from a single item to an entire missing time point of data. In trauma research, items, which are one or more measures on a multi-item survey, may be missing for various reasons. For example, a participant might not want to answer a specific question about alcohol use on a depression questionnaire or be unable to provide an open-ended follow-up answer on a multi-item pain question because of intubation. At the variable level data may be missing, either an entire single-item measure such as age, or an entire multiple-item measure e.g., the Glasgow Coma Scale (GCS). Lastly, an entire time point of data might be missing because factors such as participant death, withdrawal from the study, or inability to be contacted for follow-up. These reasons often result in different patterns of "missingness" (absence) within the data and the

underlying reasons (e.g., withdrawal from the study because of perceived burden of intervention) should be examined separately.

Although one frequently hears that the ideal study is one without missing data, achieving this is generally not feasible. The best study designers, however, attempt to decrease the influence of missing data through various techniques. This would include pilot-testing questionnaires to ensure multiple-item measures are clear and feasible. If members of the healthcare team will be collecting data for the study, it is imperative that the rationale for the study and the need for each item are apparent to all.

Researchers should also allow time for the team to provide feedback regarding feasibility and ask questions during planning and training, to allow the team to have maximal “buy-in” concerning the study. Additionally, employing adequate retention strategies is critical in longitudinal studies to decrease loss to follow-up (Carpenter & Kenward, nd; Fox-Wasylyshyn & El-Masri, 2005). In order to accurately assess the effect of missing data, researchers should report the type, amount, and patterns of missing data on main variables of interest. This reporting also indicates assessment of data quality for readers (for an example of good data reporting, see Mackenzie et al., 2007).

A plethora of current biostatistical research exists indicating the problem of missing data in clinical studies. The issue is complex and one that is inseparable from the general problem of statistics: inference. It does not take sophisticated statistical expertise to understand the two basic concerns caused by missing data in patient care research. They are:

1. Is the available sample a random sample of the patient population of interest?
2. Are all relevant and available data being used in the analysis?

The latter question deals directly with efficiency; that is to say, is the researcher maximizing power to answer the question of interest?

A sample selected at random is what drives statistical inference—the ability to generalize from a sample to a population. A randomly selected sample does not necessarily mean that it will be representative. For instance, if only female participants were randomly selected from a mixed-gender population of patients, the sample is not truly representative of the population of interest. It follows, then, that this characteristic of a sample—whether or not it was selected at random—cannot be assessed by looking at the sample itself. Instead, how the sample was obtained should determine how one assesses the appropriateness of random selection. Similarly, whether missing data are missing randomly, cannot be assessed with the sample. However in practice, in the absence of further information, researchers often check patterns of missing data with the use of a dummy variable 0=present, 1=absent and assess correlation (Fox-Wasylyshyn & El-Masri, 2005). If the correlation coefficient is high, then data are probably not missing completely at random (MCAR). A nomenclature for missing data has been developed to deal with this concern (Table 1). The resulting taxonomy differentiates among three possible situations:

Situation 1

The available sample, although incomplete, is still a random sample. This type of missing data is often referred to as MCAR or ignorable. Because the available sample is still a random sample from the population of interest, the usual statistical inferential methods remain valid when missing data are ignored. Sometimes, it is clear that missing data are MCAR. For example, suppose the admission GCS score was not obtained for patients randomly assigned to one data collector. The reason the GCS score is missing is unrelated to the test itself. Additionally it is unrelated to other variables such as time of admission

because of the random-collection assignment. An available case analysis—an analysis of participants with admission GCS scores available—will be valid. However, the analysis would be less precise than if no data were missing.

Situation 2

The available sample is a biased sample with respect to the population of interest but the bias depends on known variables. Inference will be invalid if systematic bias is not taken into account. This type of missing data is called missing at random (MAR) and, perhaps confusingly, is also ignorable. This is because given available information, participants can be identified to represent other participants with missing data. For example, suppose the GCS score is missing for some trauma patients, but this time, the score is more likely to be missing for older patients. Because age is an available covariate for all participants, selection bias can be corrected. A method that could be used in this instance of MAR data in a regression analysis is inverse probability weighting (Finkelstein et al., 2006; McKnight, McKnight, Sidani, & Figueredo, 2007).

These techniques are indicated to correct for certain under-represented covariate profiles (e.g., those who stay in the study versus those who drop out), assigning weights proportional to the inverse of the probability of being observed (Fitzmaurice, Laird, & Ware, 2004; McKnight et al., 2007). In this approach, only complete cases are included then the data are reweighted so that their distribution more closely approximates that of the full sample based upon the covariate profiles (Little & Rubin, 2002). A drawback to this approach in a multivariate situation is that it does not always allow for making use of all available data. Because of the constraints necessary for its implementation, weighting is generally used only when missing data patterns are few, similar in nature, and for a limited number of analyses (univariate, generalized estimating equations; McKnight et al., 2007).

Multiple imputation (MI) is another appropriate method that can be used for any type of statistical analysis to handle MAR data (Donders, van der Heijden, Stijnen, & Moons, 2006; Rubin, 1987). In this method, multiple (m) versions (typical range 5–20) of the data set are created using available data to predict missing values. These data sets are then used to conduct m analyses, each with its own regression values and confidence intervals, which are then combined into one inferential analysis (Rubin, 1987).

The particular appeal of this method is that once completed data sets have been developed, standard statistical methods can be used. Through the use of MI data sets, researchers are able to reduce the bias associated with single imputation which might lead to narrow confidence intervals (Fitzmaurice et al., 2004; Newgard, 2006). Single imputation techniques (e.g., mean value, last value carried forward) are generally not recommended because they underestimate uncertainty (McKnight et al., 2007; Rubin & Schenker, 1991). The MI approach is equally appropriate for handling missing data in both cross-sectional and longitudinal designs with MAR data (Ali & Siddiqui, 2000). Several statistical packages are available for MI and include R, AMELIA II, SOLAS 3.0, SAS, and S-PLUS. For further detailed information on the steps of MI in nursing research, see the reviews by Patrician (2002), McCleary (2002), and Kneipp and McIntosh (2001).

A third method available for handling MAR data is maximum likelihood (ML). ML is a model-based procedure for parameter estimation that in some situations has results similar to MI (Collins, Schafer, & Kam, 2001). Multiple imputation can be intuitively understood as an extension of simple imputation, a familiar concept to many researchers. In contrast, understanding ML methods requires a high degree of statistical knowledge. Furthermore, implementing ML procedures requires special software (LISREL, EQS, Amos, and others), which might not be familiar to some researchers (Allison, 2002). For these practical reasons

we have focused on MI as the method of choice for MAR data. A detailed discussion of ML methods can be found in Little and Rubin (2002) and Allison (2002).

Situation 3

The available sample is biased and the bias depends on the missing information. This is called nonignorable (NI) missing data or missing not at random (MNAR) because inference ignoring missing data will be biased. Sometimes, it is known that missing data are NI, for example, if admission GCS scores are missing and it is known that GCS scores are more likely to be missing for more mild levels of brain injury. No methods currently exist for making accurate inferences in this situation. However, MI has been shown to indicate estimates that are less biased than the available case analysis in some NI situations (Joseph et al., 2004).

The success of multiple imputation in an NI setting depends on having participants in the sample that can be identified using known covariates to reasonably represent those for whom data are missing—or put another way, having a variable that is highly correlated with missing variables (such as age in the GCS example). Even in such a situation, when data are NI, the question of potential bias lingers. The possible effect of the informative missingness can be assessed using sensitivity analyses (Allison, 2002).

In general, a sensitivity analysis is any repetition of analysis under an alternative assumption. Such analyses, then, are used to test the robustness of the statistical analysis to underlying assumptions (Allison, 2002; Saltelli, Chan, & Scott, 2000). In a study with missing data, sensitivity analyses allow researchers to model data across a reasonable range of expected values and assess the effects of the type of missing data assumed to be present. For example, if previous studies indicate an expected median GCS score of 13 in a sample, the researcher could impute higher values of scores for those missing. Specifically for multiple imputation analysis, which as typically implemented relies on the assumption that missing data are ignorable, it is possible to specify NI imputation models (Rubin, 1987). If results of a sensitivity analysis are not substantially different under alternative assumptions about the type of missing data present (models), a researcher can have confidence that even if the data are NI, for practical purposes the potential effect of the missing data is ignorable.

Besides the important concern of having a representative sample for inferential statistics, there is also a concern about efficient use of data. If the incomplete variable is a covariate, but not interesting itself, imputation of a covariate(s) can be used to improve precision of the analysis. It allows patients' records to be used that would otherwise not be included, thereby increasing sample size and power. Changes in regression equations will be shown by increases in the amount of variance accounted for in statistical models. However, if a predictor or outcome of interest in a clinical study is incomplete, imputation, while potentially correcting for bias, does not actually improve precision of an association estimate.

Example

To show the effects of missing data and an approach to missing data in a multivariate analysis, a data set was created that could be used to address the following research question: Is elevated intracranial pressure (ICP) in the first 72 hours following traumatic brain injury (TBI) associated with lower subsequent functional status? From ideal data, three data sets with missing data were simulated.

The ideal data set that was constructed contains the outcome of interest, Glasgow Outcome Scale-Extended (GOSE; range: 1=died to 8=upper good recovery) 3 months after TBI and

the exposure of interest: an indicator of elevated average ICP (>10 mm Hg) in the first 72 hours following TBI. Also included were three important adjustment variables (age, initial GCS score and type of insurance), as well as two additional variables highly correlated with the other measures (pupillary response on admission and injury severity score). The data set was drawn from the NSCOT study; detailed description of the enrollment and data collection procedures has been previously published (MacKenzie et al., 2006). Participants were selected from the NSCOT database for the current analysis if they met the following criteria: adult patient with a TBI defined using ICD-9-CM (International Classification of Diseases) codes (Coronado, Thomas, Sattin, & Johnson, 2005) who received ICP monitoring. A total of 373 participants met inclusion criteria. In order to create an ideal data set, the data matrix was completed using data generated from similar cases.

From the ideal data, three incomplete data sets were generated, each missing approximately 25% of each variable, a percentage typical of trauma databases (Joseph et al., 2004). In other types of clinical research, typical percentages of missing data are substantively lower than 25%; however the techniques presented in this paper remain both useful and efficient. Missing items in the first dataset were generated randomly; in the second data set, the probability of missingness was based on admission pupillary response. For the third data set missingness was based on the probability of having an elevated ICP.

For each data set, linear regression was used to estimate the association of elevated ICP in the first 72 hours following TBI and functional status 3 months after TBI adjusting for initial GCS score, age, and type of insurance. To handle missing data the following were conducted (a) available case analysis, (b) analyses with multiple imputation, and (c) sensitivity analysis. The multiple imputation was conducted with the R software package for multivariate imputation by chained equations (MICE). For the sensitivity analysis, the investigator assumed informative missingness, specifically that missing values of ICP were elevated, and then performed an imputation analysis based on this a priori assumption to handle missing outcomes and covariates.

Results

Using the ideal data, adjusting for age, initial GCS score and type of insurance, the average GOSE at 3 months postinjury between patients with elevated mean ICP (>10 mm Hg) and those with normal mean ICP (<10 mm Hg) was -1.2 ± 0.2 (Table 2). This indicates that elevated ICP adversely affects functional status. This number is an estimate of the “true” association between ICP and GOSE. Table 2 shows difference estimates obtained in three hypothetical studies with missing data. The purpose of these studies is to show the effect of missing data and to indicate appropriate methods for handling missing data and interpreting results.

Consider researchers with Data set 1, the set with approximately 25% of data MCAR. If they use only available cases (drastically reducing the sample size to only 94 patients), the difference estimate is inefficient: The standard error for the difference (0.5) is greater than that of the standard error obtained using MI (0.3). Multiple imputation provides estimates that most closely match the estimates obtained from the complete data because it increases the amount of data being used (and thus the statistical power). Results of the sensitivity analysis (difference -0.9 ± 0.3), are not qualitatively different from the MI results even though based on the incorrect assumption that the missing data are informative.

The researchers analyzing these data could have some confidence that an association exists between ICP and GOSE scores. Even if the type of missing data is unknown, researchers can state that the results are robust to assumptions about the type of missing data.

For researchers with Data set 2, with 25% of the data MAR (missingness related to pupillary response), the available case analysis is again inefficient. Results of both the imputation and sensitivity analyses are similar, surprisingly, both indicating estimates similar to the complete data estimates. The missing data depend on admission pupillary response, not on ICP as assumed in the sensitivity analysis. However, because ICP is correlated with admission pupillary response, the sensitivity analysis does well at “correcting the bias” in the sample. Again, the relative agreement of the estimates would allow researchers to be assured that the association between elevated ICP and GOSE is real despite the large amount of missing data in the study because the results are robust to assumptions about the type of missing data.

For researchers with Data set 3, missing data that are NI (related to probability of high ICP), using only available cases is extremely inefficient. The standard error (1.5) is greater than the expected effect size (~1.2) indicating the analysis shows very little power to detect this effect. In comparison, using MI slightly increases efficiency and accuracy of the estimates, yet still fails to help find an association between ICP and functional status. In contrast, the sensitivity analysis which imputes high values for ICP, indicates that elevated ICP is associated with worse functioning (difference of -0.6 ± 0.3).

With these conflicting results, researchers cannot make definitive conclusions. Their conclusions, therefore, depend on the type of missing data assumed. If researchers assume the data are MCAR or MAR, neither the available case nor the MI has power to allow investigators to detect the expected effect.

If the researchers suspect nonignorable missing data, the sensitivity analysis is the one they would be most likely to believe. However, because sensitivity analyses are based on extra assumptions about true composition of the data, the evidence they can provide for an association is not considered as strong as an association discovered with an MI analysis. This is because an MI analysis imposes no structure on the data other than what is observed in the available data. Nevertheless, the evidence for an association observed by researchers with the third data set would at least indicate support for further inquiry.

In summary, if given the available information (insurance status, injury severity and so on), participants with complete data are representative of participants with missing data (Data sets 1 and 2), MI indicates unbiased estimates of the association between ICP and functional outcome. If the data are truly NI (Data set 3), MI might still be a better estimate than is the case analysis. In all situations, sensitivity analysis is useful in establishing whether the end result is robust to assumptions about the type of data missing.

Recommendations

Most often, one does not know why data are missing nor in which category the missing data would be. For example, suppose researchers are missing preinjury functional status for some participants because they could not be assessed. In this situation, researchers cannot say if the data are MCAR (e.g., if the participant was simply not assessed), MAR (e.g., if the participant was too economically disadvantaged to have a reliable means of contact) or NI (e.g., if the participant died). In these situations, researchers frequently use MI. As discussed previously, this might be an advantage when the data are MCAR of improving precision, and if the data are NI, of providing a better approximation than the available case analysis. A thorough analyst will also perform sensitivity analyses to more fully address the potential of NI missing data. In summary, an analysis of a typical clinical study with missing data might include the following steps: (a) perform available case analysis, (b) apply multiple imputation, and (c) run sensitivity analysis.

Conclusions

In clinical studies of trauma, researchers will invariably encounter missing data. The most important thing with missing data is to assess the possibility of having a biased sample. If possible to address the sample bias, this should be done. If potential bias exists that cannot be corrected with available information, this should be clearly stated as a limitation of the study. The approach to missing data presented in this paper would also be valid in other types of clinical studies (nontrauma) with similar types or patterns of missing data. Despite the challenges inherent when data are missing, information can be gained when a thoughtful and systematic analytical approach is used.

Acknowledgments

We thank Drs. Jin Wang and Patrick Heagerty for helpful discussions regarding this work. Supported, in part, by grant number R49/CCR316840 from the National Center for Injury Prevention and Control, a Claire M. Fagin Building Geriatric Nursing Capacity Fellowship from the John A. Hartford Foundation, and the Roadmap for Medical Research.

References

- Ali MW, Siddiqui O. Multiple imputation compared with some informative dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. *Journal of Biopharm Statistics* 2000;10(2):165–181.
- Allison, PD. Missing data. Thousand Oaks, CA: Sage; 2002.
- Carpenter, J.; Kenward, M. Guidelines for handling missing data in social science research. n.d. Retrieved January 19: 2008, from <http://www.lshtm.ac.uk/msu/missingdata/index.html>
- Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods* 2001;6(4):330–351. [PubMed: 11778676]
- Coronado VG, Thomas KE, Sattin RW, Johnson RL. The CDC traumatic brain injury surveillance system: Characteristics of persons aged 65 years and older hospitalized with a TBI. *Journal of Head Trauma Rehabilitation* 2005;20(3):215–228. [PubMed: 15908822]
- Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology* 2006;59(10):1087–1091. [PubMed: 16980149]
- Finkelstein, EA.; Corso, PS.; Miller, TR. Incidence and economic burden of injuries in the United States. New York: Oxford University Press; 2006.
- Fitzmaurice, GM.; Laird, NM.; Ware, JH. Applied longitudinal analysis. New York: Wiley; 2004. Missing data and dropout; p. 375–400.
- Fox-Wasylyshyn SM, El-Masri MM. Handling missing data in self-report measures. *Research in Nursing & Health* 2005;28(6):488–495. [PubMed: 16287052]
- Gentilello LM, Donovan DM, Dunn CW, Rivara FP. Alcohol interventions in trauma centers. Current practice and future directions. *JAMA* 1995;274(13):1043–1048. [PubMed: 7563455]
- Holavanahalli RK, Lezotte DC, Hayes MP, Minhajuddin A, Fauerbach JA, Engrav LH, et al. Profile of patients lost to follow-up in the burn injury rehabilitation model systems' longitudinal database. *Journal of Burn Care Research* 2006;27(5):703–712. [PubMed: 16998404]
- Joseph L, Belisle P, Tamim H, Sampalis JS. Selection bias found in interpreting analyses with missing data for the prehospital index for trauma. *Journal of Clinical Epidemiology* 2004;57(2):147–153. [PubMed: 15125624]
- Kneipp SM, McIntosh M. Handling missing data in nursing research with multiple imputation. *Nursing Research* 2001;50(6):384–389. [PubMed: 11725942]
- Little, RJA.; Rubin, DB. Statistical analysis with missing data. 2. Hoboken, NJ: Wiley; 2002.
- MacKenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, et al. A national evaluation of the effect of trauma-center care on mortality. *New England Journal of Medicine* 2006;354(4):366–378. [PubMed: 16436768]

- Mackenzie EJ, Rivara FP, Jurkovich GJ, Nathens AB, Frey KP, Egleston BL, et al. The National Study on Costs and Outcomes of Trauma. *Journal of Trauma* 2007;63(6 Suppl):S54–S67. Discussion S81–S86. [PubMed: 18091213]
- McCleary L. Using multiple imputation for analysis of incomplete data in clinical research. *Nursing Research* 2002;51(5):339–343. [PubMed: 12352784]
- McKnight, PE.; McKnight, KM.; Sidani, S.; Figueredo, AJ. *Missing data: A gentle introduction*. New York: Guilford; 2007.
- Moore L, Lavoie A, LeSage N, Liberman M, Sampalis JS, Bergeron E, et al. Multiple imputation of the Glasgow Coma Score. *Journal of Trauma* 2005;59(3):698–704. [PubMed: 16361915]
- Newgard CD. The validity of using multiple imputation for missing out-of-hospital data in a state trauma registry. *Academic Emergency Medicine* 2006;13(3):314–324. [PubMed: 16495420]
- Newgard CD, Haukoos JS. Advanced statistics: Missing data in clinical research—Part 2: Multiple imputation. *Academic Emergency Medicine* 2007;14(7):669–678. [PubMed: 17595237]
- Nilsen P, Holmqvist M, Nordqvist C, Bendtsen P. Linking drinking to injury—causal attribution of injury to alcohol intake among patients in a Swedish emergency room. *International Journal of Injury Control and Safety Promotion* 2007;14(2):93–102. [PubMed: 17510845]
- Patrician PA. Multiple imputation for missing data. *Research in Nursing & Health* 2002;25(1):76–84. [PubMed: 11807922]
- Richmond TS, Hollander JE, Ackerson TH, Robinson K, Gracias V, Shults J, et al. Psychiatric disorders in patients presenting to the Emergency Department for minor injury. *Nursing Research* 2007;56(4):275–282. [PubMed: 17625467]
- Rubin, DB. *Multiple imputation for nonresponse in surveys*. New York: Wiley; 1987.
- Rubin DB, Schenker N. Multiple imputation in health-care databases: An overview and some applications. *Statistics in Medicine* 1991;10 (4):585–598. [PubMed: 2057657]
- Saltelli, A.; Chan, K.; Scott, EM. *Sensitivity analysis: Gauging the worth of scientific models*. New York: Wiley; 2000.
- Scheetz LJ. Relationship of age, injury severity, injury type, comorbid conditions, level of care, and survival among older motor vehicle trauma patients. *Research in Nursing & Health* 2005;28(3):198–209. [PubMed: 15884027]
- Sommers MS. Injury as a global phenomenon of concern in nursing science. *Journal of Nursing Scholarship* 2006;38:314–320. [PubMed: 17181078]
- Sommers MS, Dyehouse JM, Howe SR, Fleming M, Fargo JD, Schafer JC. Effectiveness of brief interventions after alcohol-related vehicular injury: A randomized controlled trial. *Journal of Trauma* 2006;61(3):523–531. Discussion 532–533. [PubMed: 16966982]
- Thompson HJ, Rivara FP, Jurkovich GJ, Wang J, Nathens AB, MacKenzie EJ. Evaluation of the effect of intensity of care on mortality after traumatic brain injury. *Critical Care Medicine* 2008;36(1):282–290. [PubMed: 18007264]
- World Health Organization. *Facts about injuries: Preventing global injuries [Electronic Version]*. 2001. Retrieved August 24, 2007, from http://www.who.int/violence_injury_prevention/publications/factsheets/en/index.html

Table 1

Taxonomy of Missing Data

Nomenclature in common use	Reason data missing	Approach
Ignorable		
•Missing Completely at Random (MCAR)	Unrelated to any variables in analysis	Ignore
•Missing at Random (MAR)	Unrelated to missing data when available data taken into account	Multiple Imputation
Nonignorable (NI)		
•Missing Not at Random (MNAR)	Dependent on missing data	Sensitivity analysis

Note. Adapted from Donders, van der Heijden, Stijnen, and Moons, 2006; Joseph, Belisle, Tamim, and Sampalis, 2004; Newgard and Haukoos, 2007.

Table 2

Average Glasgow Outcome Scale—Extended Scores for Exposure Groups and Differences With Standard Errors; Controlling for Age, Type of Insurance, and Initial Glasgow Coma score

	<i>N</i>	Low ICP (1–10 mm Hg)	High ICP (>10 mm Hg)	GOSE score difference (standard error)
Complete Data	373	3.8	2.5	−1.2 (0.2)
Data set 1 (MCAR)				
available case	94	3.4	3.2	−0.2 (0.5)
multiple imputation		3.3	2.2	−1.1 (0.3)
sensitivity		2.9	2.0	−0.9 (0.3)
Data set 2 (MAR)				
available case	180	4.5	3.7	−0.8 (0.4)
multiple imputation		3.3	2.4	−0.9 (0.3)
sensitivity		3.9	2.8	−1.1 (0.3)
Data set 3 (NI)				
available case	192	3.7	3.8	+0.1 (1.5)
multiple imputation		3.5	3.4	−0.1 (0.8)
sensitivity		3.7	3.0	−0.6 (0.3)