Handling of Missing Outcome Data in Acute Stroke Trials: Advantages of Multiple Imputation Using Baseline and Post-Baseline Variables

Dashiell F. Young-Saver[1], Jeffrey Gornbein[2], Sidney Starkman[3]
Jeffrey L. Saver[1]
[1]Department of Neurology and Comprehensive Stroke Center, David Geffen School of Medicine at the University of California, Los Angeles,
710 Westwood Plaza, Los Angeles, CA 90095
[2]Department of Biomathematics, University of California, Los Angeles,
621 Charles E Young Dr S, Los Angeles, CA 90024
[3] Departments of Emergency Medicine and Neurology and Comprehensive Stroke Center, David Geffen School of Medicine at the University of California, Los Angeles,
10833 Le Conte Ave, Los Angeles, CA 90095

Abstract
Multiple imputation (MI) is a preferred approach to missing outcome data in longitudinal studies of progressive disease; older techniques, like complete case analysis (CCA) and last observation carried forward (LOCF), can bias toward unduly good outcomes. But the preferred approach for studies of acute illnesses with some recovery, such as acute stroke, is understudied. In these settings, CCA and LOCF may bias toward worse group outcomes than actually occur. Therefore, these methods are viewed as "conservative" and remain often used.

Using data from a well-known acute stroke trial, we simulated data missingness and compared 5 handling methods: 1) CCA, 2) worst outcome assigned, 3) LOCF, 4) MI using baseline covariates (BCVs), and 5) MI using BCVs plus later observed outcomes. Imputation methods that ignored post-baseline data showed poor correlation with actual outcomes and reduced study power. LOCF preserved power but biased outcome estimates to worse than actual. MI with BCVs plus interim outcome observations yielded highest power, accuracy, and lack of directional bias. We describe techniques to assess bias and variance of imputation methods in acute illness trials generally.

Key Words: Missing Data, Imputation, Last Observation Carried Forward, Multiple Imputation, Stroke, Clinical Trial

1. Introduction

Missing outcome data in randomized clinical trials can reduce study power, bias estimates of treatment effect positively or negatively, and diminish the generalizability of findings.[1-3] For randomized clinical trials, simply ignoring missing data violates the intent-to-treat principle and can lead to biased estimates, so imputation of missing values is often instead undertaken.[2, 4] However, several techniques for imputation are available, and their relative strengths and weaknesses when used for acute stroke trial outcomes have not been well-delineated.

In longitudinal studies of progressive or recurrent disease, multiple imputation (MI) has been established as a preferred,[1, 2, 5] though not always implemented,[6] approach to missing outcome data, as older approaches like complete case analysis (CCA) and last observation carried forward (LOCF) bias toward unduly good outcomes. However, in studies of acute illnesses with some recovery, CCA and LOCF may bias toward worse group outcomes than actually occur. Therefore, for acute stroke trials, some investigators consider these methods "conservative," and they remain in use. [7-11]

Using data from a well-known acute stroke trial, we investigated: 1) complete case analysis, 2) worst case imputation, 3) last observation carried forward, 4) multiple imputation using baseline variables only, and 5) multiple imputation using baseline and post-baseline variables.

2. Methods

We analyzed data from NINDS-tPA Study Trial 2, the pivotal trial that led to the approval of tissue plasminogen activator as the first pharmacologic therapy for acute ischemic stroke.[15] We simulated data missingness by creating datasets with some 90 day modified Rankin Scale outcome randomly deleted from both treatment groups, evaluating deletion of 5%, 10%, 15%, 20%, and 25% of outcomes. Missing values were randomly drawn 200 times in each of these missing amounts, creating 200 x 5 = 1,000 total datasets.

In the created datasets, 5 techniques for handling the missing 90d mRS data were evaluated:

1) Complete case analysis (CCA) – analyze only the cases without missing outcome data.

2) Worst case imputation (WCI) – impute the value of the worst possible outcome (6 = dead) for participants with missing outcome data.

3) Last observation carried forward (LOCF): impute the last observed (7-10d) mRS value for the 90d mRS value.

4) Multiple imputation using baseline variables only (MI-B) – using cases with a 90d mRS, ordinal logistic regression prediction models were generated for 90d mRS based on patient baseline features through a multivariate imputation by chained equations (MICE) procedure.[16]

5) Multiple imputation using baseline plus post-baseline variables (MI-BP) – using cases with a 90d mRS, ordinal logistic regression prediction models were generated for 90d mRS based on patient baseline features plus post-baseline covariates through a multivariate imputation by chained equations (MICE) procedure.[16]

The imputation models used 18 baseline demographic and clinical characteristics. The models incorporating post-baseline variables included 2 additional covariates: NIHSS at 24 hours (handled as a continuous variable), and mRS at day 7-10 (handled as an ordinal variable).

We evaluated the different missing data-handling methods in two broad ways. The first was by measuring the accuracy of the imputed values versus the gold standard of the actual 90d mRS values. Accuracy was quantified using: 1) exact agreement rate, 2) Spearman correlation coefficient, 3) mean absolute difference, and 4) mean difference.

The second was by analyzing impact on trial-level findings of beneficial study drug effect, for outcomes of: 1) 90d mRS dichotomized at 0-1 (disability-free) vs 2-6 (disabled or dead) (Fisher exact test); and 2) favorable shift to lower disability levels across all 7 disability levels of the mRS (ordinal logistic regression).

Among the 18 baseline variables, most had complete data available in the trial dataset. Several had minimal (<2%), four modest (2-10%), and none substantial missingness. The analyzed data was from the public NINDS-tPA dataset, and the analysis was conducted under the US DHHS exemption for additional consenting or IRB review for secondary analyses of public datasets.

3. Results

In the full NINDS-tPA Trial 2 dataset, 333 patients were enrolled and all had documented 90-day modified Rankin Scale scores (distributions shown in Figure 1). For the dichotomized outcome of freedom from disability (mRS 0-1), trial findings using all actual data were: tPA vs placebo, 38.7% vs 16.1%, odds ratio 1.78 (95%CI, 1.10-2.94), p=0.01). For the outcome of level of disability across all 7 mRS levels, tPA increased the likelihood of a lower disability score, common odds ratio (cOR) 1.50 (95%CI 1.03-2.20), p=0.035.

In the simulated missingness datasets, the concordance between imputed and actual 90-day mRS values is shown in Table 1. Exact agreement rates were highest for MI-BP and LOCF, intermediate for MI-B, and lowest for WCI. Correlations between the imputed mRS and the actual mRS were highest for LOCF, and also high for MI-BP, and lower for MI-B, with no correlation present by definition for WCI. The distance of imputed from actual 90-day mRS scores was highest for MI-BP and LOCF, intermediate for MI-B, and lowest for WCI.

Directional bias for the imputed 90-day mRS values, compared with the actual 90-day mRS scores, had a different pattern (Table 1, histograms displayed in Figure 2). No directional bias was present with MI-BP. Bias toward imputing poorer outcomes than actually observed was present for the other techniques, small for MI-B, moderate for LOCF, and marked for WCI.

The impact of the missing data-handling methods on a formally positive study result (using the conventional $p<0.05$ threshold) is shown in Figure 3. Across simulations, the greater the frequency of missing data, the greater the impact of the different missing data-handling methods. MI-BP and LOCF best retained the actual positive trial result, remaining formally positive, on average, for both dichotomized and shift analysis across all 5 rates of data missingness. MI-B and complete cases analysis performed at an intermediate level, with average p values becoming non-positive at ≥10-20% missingness rates. WCI performed at the lowest level, with average p values becoming non-positive at ≥5-20%. Study power calculations showed a similar pattern (data available on request).

4. Discussion

The most consistently advantageous method was MI-BP. Compared with the actual observed values, MI-BP had high concordance and low directional bias in estimating final outcome, and showed excellent preservation of study power to detect the present treatment benefit. LOCF also performed well, with high concordance, but moderate bias toward estimating outcomes as worse than actual, and excellent preservation of study power. The remaining techniques, that did not draw upon post-baseline variables, did not perform as well.

It is noteworthy that the two best performing missing data handling methods both drew upon the last observed functional outcome to predict the final outcome, either alone (LOCF) or in combination with other variables (MI-BP). Several acute stroke studies have shown that, among all baseline and post-baseline variables, the last observed functional state is the single most powerful variable for predicting the final functional state.[13, 14]

A limitation of the current study is that the mechanism of data missingness explored in the simulations was solely MCAR. But, for MAR, evaluating the MCAR condition is a conservative approach. The better performance of imputation methods that incorporate post-baseline covariates observed in the current study with MCAR missingness would be expected to be even greater with MAR missingness, as post-baseline variables capture important information not only about the missing outcome, but also the reason for its missingness. In addition, with MAR mechanisms of missingness, complete case analysis would show directional bias not present in a pure MCAR simulation.[17]

In addition, the findings of the current study also suggest likely benefit of the MI-BP imputation technique in MNAR conditions of missingness. Compared with complete case analysis, the MI-BP technique's incorporation of post-baseline variables captures important information not only about the missing outcome, but also the reason for its missingness. In addition, in MNAR settings, the directional bias of the LOCF technique would often amplify loss to follow-up biases.

In conclusion, this analysis of empirical trial data and simulated missingness indicates that, for acute stroke, multiple imputation using baseline variables plus post-baseline variables (especially the last observation of the functional outcome) is an advantageous approach to handling missing final functional outcome data, with high concordance with actual outcomes, absence of directional bias, and greater preservation of study power to confirm an actually-present treatment benefit.

5. Tables and Figures

Table 1: Correlation, Exact Agreement, Distance, and Directional Bias of Estimated versus Actual mRS Outcomes with Different Imputation Methods^

	Correlation (Spearman rho)	Exact Agreement percent (95%CI)	Distance Root mean sq error, mean (95% CI)	Directional Bias mean difference (95% CI)
Worst Case Imputation	N/A^^	13.3 (12.7, 13.9)	3.8 (3.7, 3.8)	3.22 (3.19, 3.26)
Multiple Imputation – Baseline Vars Only	0.46	21.4 (20.6, 22.2)	1.9 (1.9, 1.9)	0.16 (0.12, 0.19)
Last Observation Carried Forward	0.85	38.4 (37.5, 39.3)	1.2 (1.1, 1.2)	0.37 (0.35, 0.39)
Multiple Imputation – Baseline+Post-Baseline Vars	0.78	37.6 (36.6, 38.5)	1.3 (1.3, 1.3)	-0.01 (-0.04, 0.02)

Vars = Variables; Sq = square
^Table shows values for simulations of 15% missingness, the midpoint of explored missingness rates. Values were very similar with the other rates of missingness.
^^Correlation coefficient cannot be calculated for worst case imputation as there is no variance among imputed values (value of 6 is always imputed in WCI).

Figure 1: Actual 3-Month mRS Disability Outcomes in NINDS-tPA Trial 2

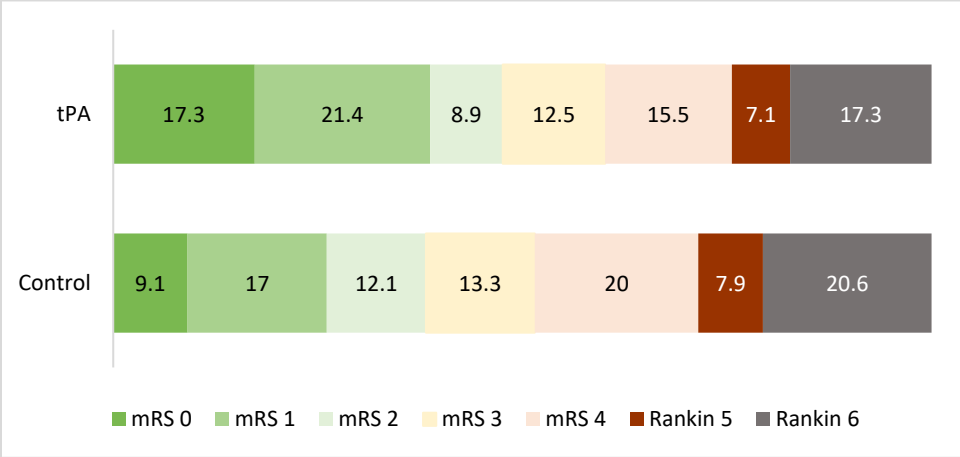

Figure 2: Frequency of differences in imputed versus actual values for 90-Day mRS, with simulated 15% missingness rates, over all 200 simulations.

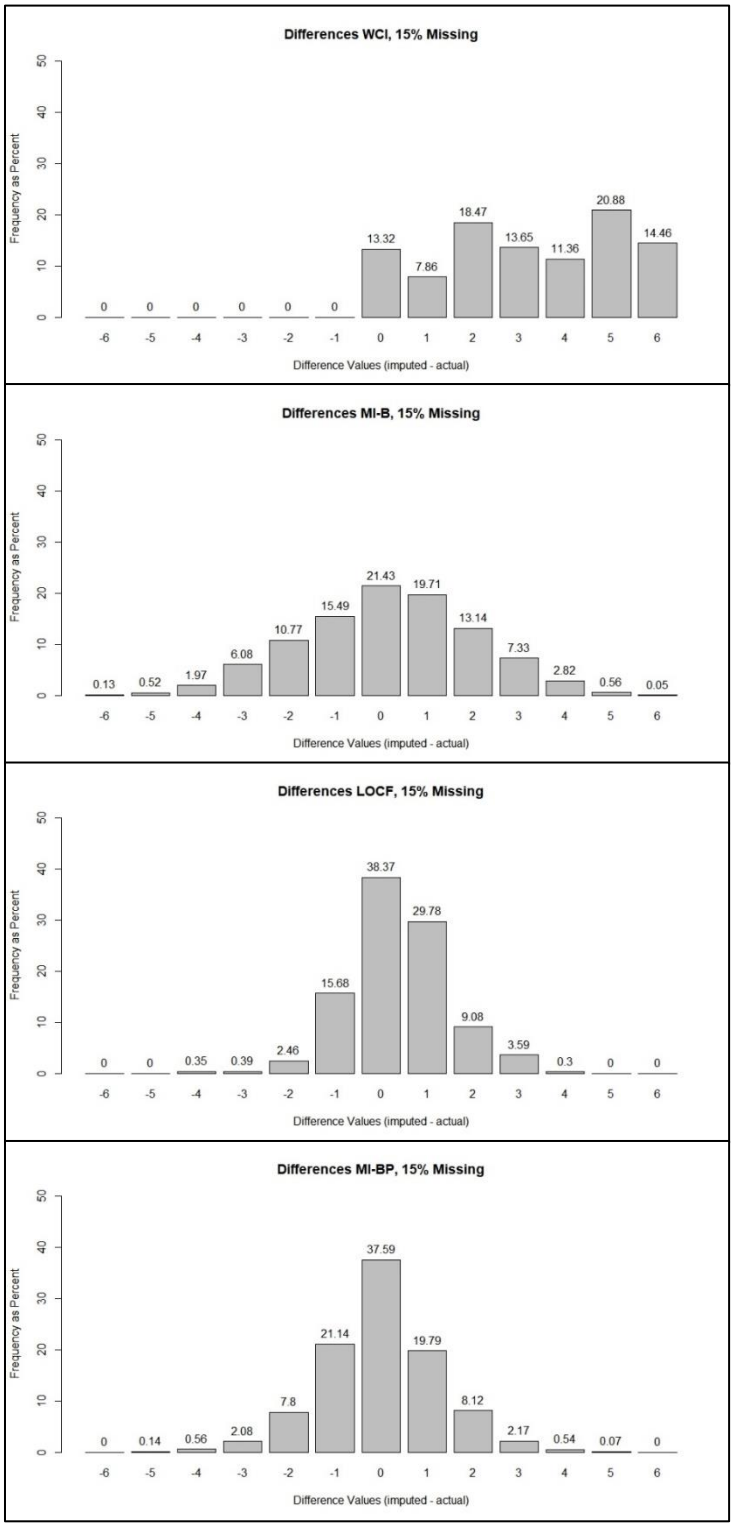

Figure 3: Heat Map of Rates of Trial Positivity with Different Missing Outcome Data Handling Methods. Average p values across the 200 simulations are shown. Dark green indicates strongly positive ($p < 0.03$), light green positive ($p = 0.03$-0.049), yellow trend positive ($p = 0.05$-0.099), red neutral ($p \geq 0.10$) findings.

Missing Data Handling Method	Disability-Free (mRS 0-1) Actual p value with no missing data: 0.014					Reduced Disability (all 7 mRS ranks) Actual p value with no missing data: 0.035				
	5% Missing	10% Missing	15% Missing	20% Missing	25% Missing	5% Missing	10% Missing	15% Missing	20% Missing	25% Missing
No data used from cases with missing final outcome										
Complete Case Analysis	0.022	0.028	0.042	0.051	0.059	0.045	0.049	0.067	0.074	0.094
Worst Case Imputation	0.024	0.032	0.049	0.061	0.073	0.069	0.088	0.154	0.160	0.245
Baseline data only used from cases with missing final outcome										
Multiple Imputation – Baseline Variables Only	0.022	0.032	0.049	0.058	0.079	0.046	0.054	0.072	0.079	0.105
Post-baseline data used from cases with missing final outcome										
Last Observation Carried Forward	0.015	0.015	0.015	0.015	0.013	0.036	0.033	0.036	0.036	0.037
Multiple Imputation – Baseline + Post-Baseline Variables	0.017	0.018	0.020	0.021	0.022	0.036	0.035	0.037	0.039	0.043

Disclosures

SS and JLS served as unpaid site investigators in a multicenter trial sponsored by Genentech, and JLS served as an unpaid site investigator in a multicenter prevention trial sponsored by Boehringer Ingelheim, for which the UC Regents received payments on the basis of clinical trial contracts for the number of subjects enrolled. JLS received funding for services as a scientific consultant regarding rigorous trial design and conduct to Boehringer Ingelheim (prevention only). JLS serves as an unpaid consultant to Genentech advising on the design and conduct of the PRISMS trial; neither the University of California nor Dr. Saver received any payments for this voluntary service.

References

1. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012;367:1355-1360
2. Council NR. *The prevention and treatment of missing data in clinical trials*. Washington, DC: The National Academies Press; 2010.
3. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ*. 2009;338:b2393
4. Yeatts SD, Martin RH. What is missing from my missing data plan? *Stroke*. 2015;46:e130-132
5. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91:473-489
6. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in rcts; a review of the top medical journals. *BMC Med Res Methodol*. 2014;14:118
7. Hill MD, Martin RH, Palesch YY, Tamariz D, Waldman BD, Ryckborst KJ, et al. The albumin in acute stroke part 1 trial: An exploratory efficacy analysis. *Stroke*. 2011;42:1621-1625
8. Albers GW, von Kummer R, Truelsen T, Jensen JK, Ravn GM, Gronning BA, et al. Safety and efficacy of desmoteplase given 3-9 h after ischaemic stroke in patients with occlusion or high-grade stenosis in major cerebral arteries (dias-3): A double-blind, randomised, placebo-controlled phase 3 trial. *Lancet Neurol*. 2015;14:575-584
9. Davalos A, Alvarez-Sabin J, Castillo J, Diez-Tejedor E, Ferro J, Martinez-Vila E, et al. Citicoline in the treatment of acute ischaemic stroke: An international, randomised, multicentre, placebo-controlled study (ictus trial). *Lancet*. 2012;380:349-357
10. Mocco J, Zaidat OO, von Kummer R, Yoo AJ, Gupta R, Lopes D, et al. Aspiration thrombectomy after intravenous alteplase versus intravenous alteplase alone. *Stroke*. 2016;47:2331-2338
11. Saver JL, Goyal M, Bonafe A, Diener HC, Levy EI, Pereira VM, et al. Stent-retriever thrombectomy after intravenous t-pa vs. T-pa alone in stroke. *N Engl J Med*. 2015;372:2285-2295
12. Qureshi AI, Palesch YY, Barsan WG, Hanley DF, Hsu CY, Martin RL, et al. Intensive blood-pressure lowering in patients with acute cerebral hemorrhage. *N Engl J Med*. 2016;375:1033-1043
13. Ovbiagele B, Lyden PD, Saver JL. Disability status at 1 month is a reliable proxy for final ischemic stroke outcome. *Neurology*. 2010;75:688-692
14. Ovbiagele B, Saver JL. Day-90 acute ischemic stroke outcomes can be derived from early functional activity level. *Cerebrovasc Dis*. 2010;29:50-56
15. NINDS rt-PA Stroke Group. Tissue plasminogen activator for acute ischemic stroke. *New England Journal of Medicine*. 1995;333:1581-1587
16. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20:40-49
17. Kontopantelis E, White IR, Sperrin M, Buchan I. Outcome-sensitive multiple imputation: A simulation study. *BMC Med Res Methodol*. 2017;17:2