RADARS® SYSTEM

Researched Abuse, Diversion and Addiction-Related Surveillance System

QUARTERLY Technical REPORT

Second Quarter, 2019

Multiple Imputation of Missing Data in the Poison Center Program

Key Findings

- The proposed methodology for imputing missing values in this context is validated and reduces bias.
- Population-adjusted rates increase as little as 6.6% (immediate-release (IR) hydrocodone) and as much as 94.5% (IR morphine) after imputation.
- Comparisons of rate estimates within an active pharmaceutical ingredient (API) group are not substantially affected by missing data.
- Rate estimates across API drug groups are substantially affected by missing data and are likely biased when only complete case methods are considered.

Introduction

The RADARS® System Poison Center Program is an active surveillance system capable of identifying geographic and temporal patterns in prescription opioid abuse. However, some data are limited in that information on specific products and/or formulations are not available when exposures are reported. The extent of missing data at the product and formulation level in the Poison Center Program raises concerns that:

- excluding these cases from product and formulation-specific rate calculation estimates will result in an underestimate of the true number of cases within each category.
- if both proportions and patterns of missing data are not consistent across all active pharmaceutical ingredients (APIs) and formulations, across-API comparisons may be biased.
- if the pattern of missing data is not missing completely at random (MCAR)[1,2] with respect to drug group assignment, comparisons of formulation-level rates within and between API groups may be biased.

Multiple imputation is a common option for handling missing values. It uses covariates from a sample to model a posterior predicted distribution from which a set of values are drawn. These values are then used to generate estimates through multiple rounds of imputation, which are then combined to reflect variability across the data set.[1,3] In many instances, it is considered an improvement over complete case and single imputation methods. As such, multiple imputation was investigated to determine if it can be a valid strategy to mitigate bias associated with missing data from the Poison Center Program.

Methods

Data Sources

The study period for this analysis was 1st quarter of 2012 through 1st quarter of 2019. The surveillance population consisted of exposure cases recorded by 51 regional poison centers in 49 states covering people in urban, suburban, and rural regions (96% of total US population). The primary outcomes for this analysis were population-adjusted rates and rate ratios comparing exposures of immediate-release (IR) tablet and capsule

formulations to exposures of extended-release (ER) tablet and capsule formulations of hydrocodone, hydromorphone, morphine, oxycodone, and oxymorphone in the Poison Center Program. Population denominators were extrapolated from the 2000 and 2010 US Censuses and scaled per 100,000 population. Missing data occurs within an API at the formulation and product levels. A summary by API at the formulation level is shown in Table 1.

Table 1: Number of missing cases in the Poison Center Program by formulation and active pharmaceutical ingredient (API)

Formulation	Hydrocodone N (%)	Hydromorphone N (%)	Morphine N (%)	Oxycodone N (%)	Oxymorphone N (%)
IR Tablets/ Capsules	122,363 (88.2%)	6,035 (68.8%)	1,243 (5.8%)	61,722 (60.8%)	197 (7.5%)
ER Tablets/ Capsules	283 (0.2%)	241 (2.7%)	8,865 (41.7%)	11,479 (11.3%)	1,922 (73.0%)
Non-Tablets/ Capsules	7,512 (5.4%)	646 (7.4%)	1,129 (5.3%)	N/A	N/A
Missing	8,577 (6.2%)	1,844 (21.0%)	10,020 (47.1%)	28,314 (27.9%)	515 (19.6%)

Multiple Imputation Analysis

As a first step in evaluating the best approach to handle missing values, the distribution and pattern of missing data were examined. Because the distribution of missing data was not consistent across exposures where the API was known (Table 1), it was determined that API-specific imputation models would be utilized to account for these differences in missingness patterns between APIs. Mutually-exclusive API and formulation-specific drug groups were created and considered as the nominal outcome in the final imputation model.

As standard practice, basic assumptions about missing data patterns were assessed. Since most of the potential predictor variables were significantly associated with missing drug group, it was assumed that the data were not MCAR.[1,2] For purposes of multiple imputation, it was necessary to assume that the data were missing at random (MAR).[1,2] While the MAR assumption formally cannot be tested, the assumption becomes more plausible as more variables are added to the imputation model, so the goal was to include as many informative predictors as possible.[4] Stepwise model selection for logistic regression (with drug group assignment as the outcome and variable inclusion and deletion thresholds set at $p>0.0001$) was used to select predictors for the imputation model. Standard variables collected on the exposures, including demographic information, location, time, and outcome were considered. The following predictors were used in the final imputation model: ingestion route of administration (ROA), inhalation ROA, injection ROA, dermal ROA, unknown ROA, exposure reason, medical outcome, managed healthcare facility type, region, sex, year, age, and number of substances.

After consideration of several approaches, fully conditional specification (FCS),[5] a semi-parametric iterative Markov chain Monte Carlo (MCMC) procedure, was used to model the imputed outcome. APIs were modeled separately, and each was run with 50 imputations.[3,6,7] Imputed population-adjusted rate and rate ratio calculations with corresponding 95% confidence intervals were generated assuming a Poisson distribution. These estimates were pooled to produce imputed rates and rate ratio estimates with associated measures of variability using Rubin's Rules.[3] Performance of the imputation model was evaluated using 10-fold cross-validation[8-12], with agreement between the imputed count and the true count of each drug group within each fold as the measure of performance. Percent change was calculated for each (k) fold and drug group as: $((\text{mean imputed count}_k - \text{true count}_k)/ \text{true count}_k)*100$.

Complete Case Analysis

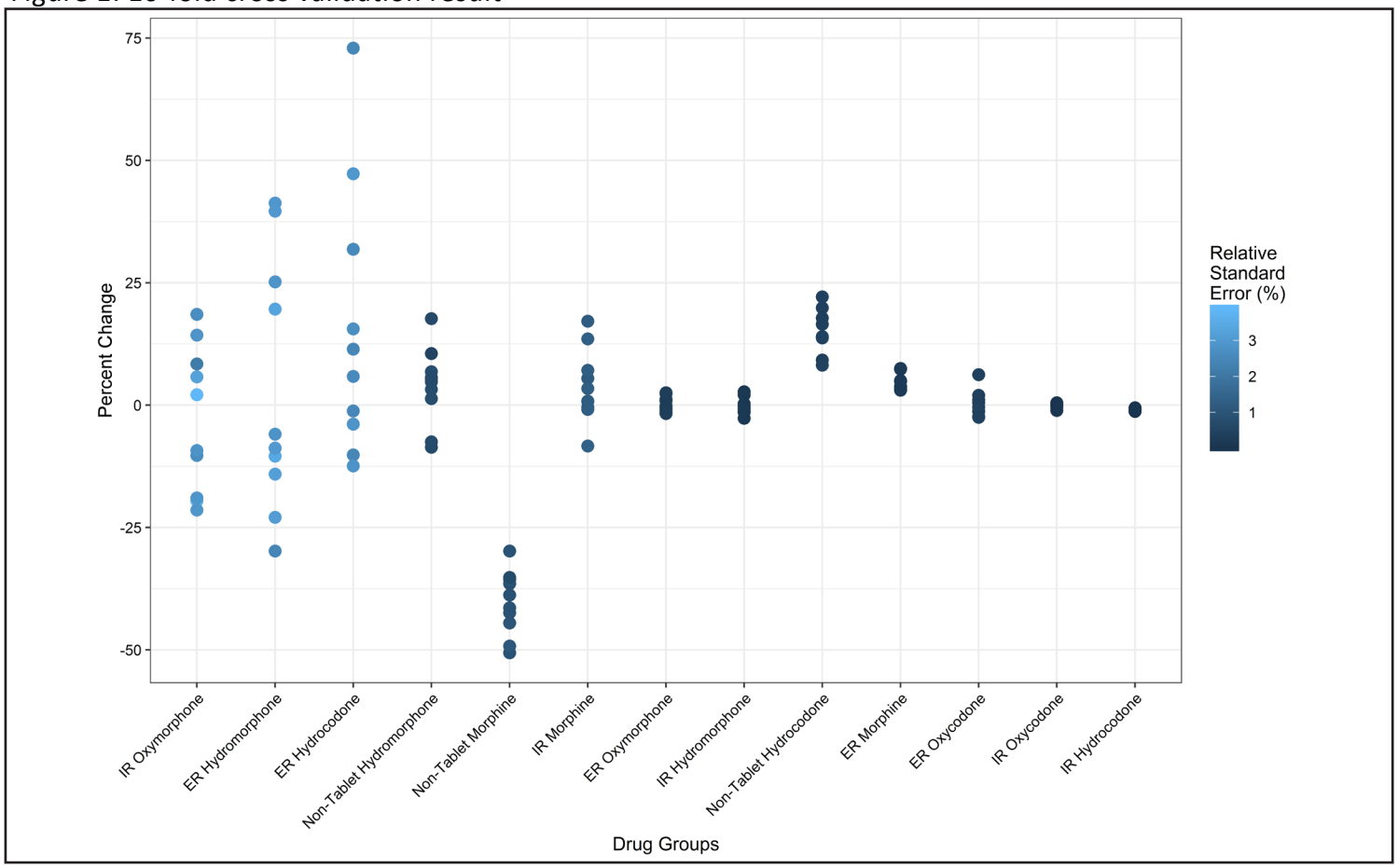
Population-adjusted rates of exposure with 95% exact Poisson confidence intervals were calculated from the complete cases for comparison to the rates calculated from the imputed data. Rate ratios of IR and ER tablet and capsule formulations of oxycodone and morphine were also generated for comparison to their counterparts calculated from the imputed data.

Results

Validation of the Model

Figure 1 summarizes agreement between the mean imputed count and the true mean count within each fold of the cross validation, with each dot providing information on a specific drug group within a fold. The x-axis is sorted by ascending mean true count per fold, and the figure is colored by relative standard error of the mean count across imputations per fold. Relative standard errors and percent change variability are highest for the following low volume drug groups: IR oxymorphone, ER hydromorphone, and ER hydrocodone. Within-fold true counts of these groups are between 14 and 35. Since estimates are combined across many imputations, these characteristics of the low volume drug groups are not of high concern. The model performs best when the true counts per fold are above 100, which includes all of the groups to the right of IR morphine in Figure 1. Non-tablet/capsule hydrocodone and ER morphine groups were consistently overestimated while the non-tablet/capsule morphine group was consistently underestimated, suggesting there is room for improvement in the final imputation model chosen.

Figure 1: 10-fold cross validation result

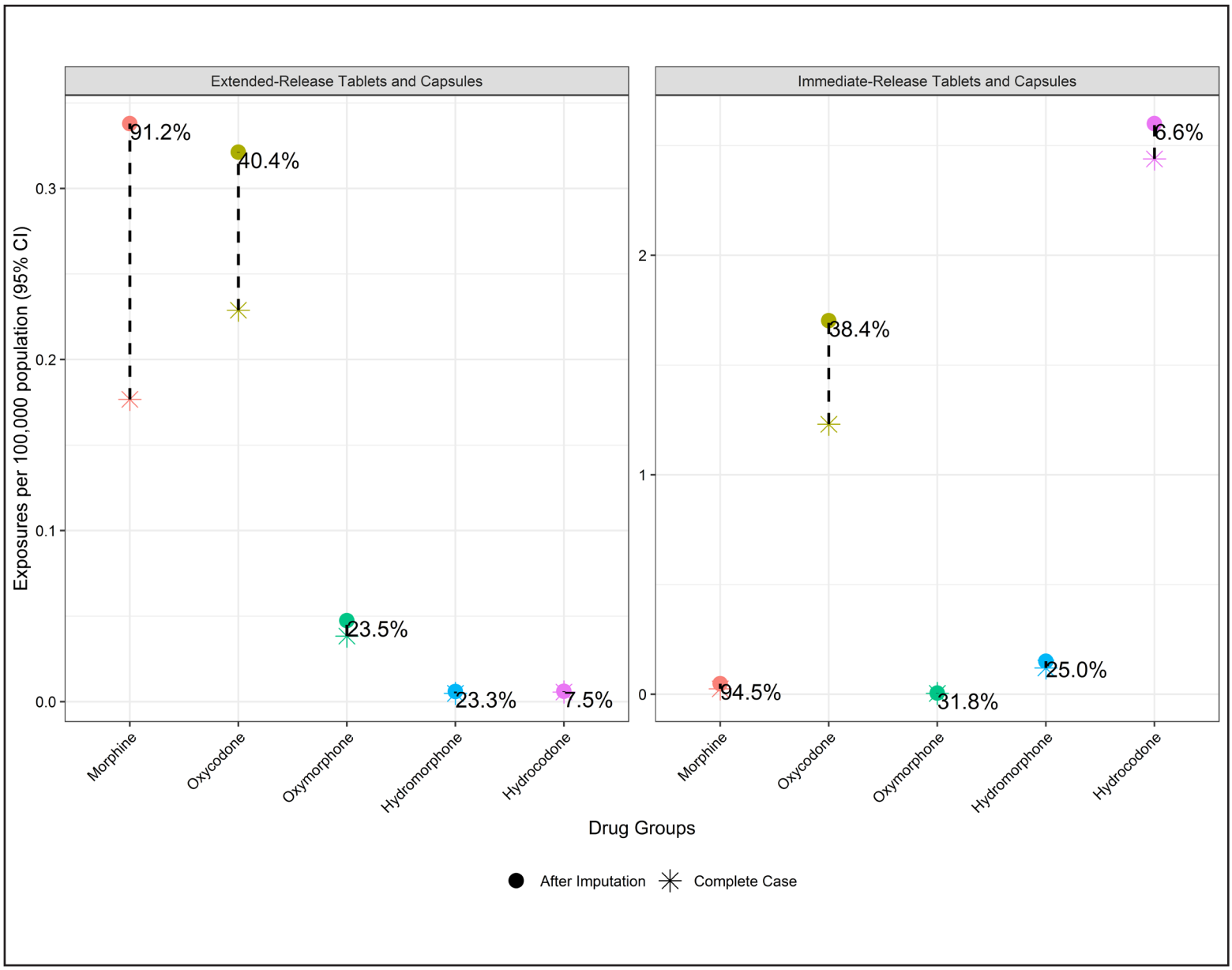


*Relative standard error calculated as ($SE_{\text{imputed count within k}}$ / $\text{Mean}_{\text{imputed count within k}}$)*100

Multiple Imputation Impact on Rate Estimates

As expected, the mean number of exposure cases and pooled rates across the final model's 50 imputations are greater than the number of cases present in the original data and direct rate estimation without imputation for all of the drug groups (Figure 2). The high percent changes in IR (94.5%) and ER (91.2%) morphine and IR (38.4%) and ER (40.4%) oxycodone groups raise concerns that population-adjusted exposure rates of these groups may be underestimated using complete case methods.

Figure 2: Population-adjusted rates of exposures with percent change of multiple imputation relative to complete case analysis



In order to evaluate the impact of imputation on comparisons of rate estimates, rate ratios generated from complete case data were compared to rate ratios generated from the imputed data. Rate ratios comparing IR tablet and capsule exposures to ER tablet and capsule exposures within and across select API drug groups are displayed in Table 2. The percent changes are small for within-API comparisons and substantially larger for across-API comparisons. While comparisons of rate estimates within an API group are not greatly affected by missing data, comparisons of rate estimates across API drug groups are affected considerably. For example,

there is over a 28% decrease in the estimated relationship of IR oxycodone exposure rates to IR morphine exposure rates when calculated with the imputed data compared to when it was calculated with complete case data.

Table 2: Population-adjusted rate ratios for select tablets and capsules groups

	Complete Case Analysis	After Imputation Analysis	Absolute Change	Percent Change[1]
	Rate Ratio (95% CI)			
Within-API[2]				
IR Morphine/ ER Morphine	0.14 (0.13, 0.15)	0 0.14 (0.10, 0.19)	0.00	1.7%
IR Oxycodone/ ER Oxycodone	5.38 (5.27, 5.49)	5.30 (5.24, 5.36)	-0.08	-1.4%
Across-API[3]				
ER Oxycodone/ ER Morphine	1.29 (1.26, 1.33)	0.95 (0.93, 0.97)	-0.34	-26.6%
IR Oxycodone/ IR Morphine	49.66 (46.94, 52.52)	35.34 (33.99, 36.68)	- 14.32	-28.8%

[1]Calculated as $((\text{rate ratio}_{MI} - \text{rate ratio}_{\text{complete case}})/ \text{rate ratio}_{\text{complete case}})*100$

[2]Drug groups being compared are subgroups of the same API

[3]Drug groups being compared are subgroups of different APIs

Conclusions

The multiple imputation methodology evaluated for formulation-level missing data in the Poison Center Program appears to effectively address missing formulation values within an API. The example model displayed good predictive performance, as shown through consistency in imputed vs. true counts within each fold of the cross validation. The high percent changes in rates for IR and ER morphine and oxycodone groups demonstrate that population-adjusted exposure rates of these groups may be underestimated using only complete case methods. Rate ratios generated with complete cases compared to rate ratios generated from imputed data show small changes in within-API drug group comparisons and large changes in across-API drug group comparisons. This suggests that comparisons of rate estimates across API drug groups are substantially affected by missing data and are likely biased when only complete case methods are considered. Accounting for missing data should be considered when comparisons across API drug groups are of interest, and multiple imputation has proven to be a valid strategy in this context.

Suggested Citation

Amioka E, Burnham RI, Kreider SED, Black JC, Severtson SG, Dart RC (2019). Multiple Imputation of Missing Data in the Poison Center Program. RADARS® System Technical Report, 2019-Q2.

References

[1]Rubin, DB. Inference and missing data. Biometrika. 1976; 63(3): 581–592.

[2]Little RJA, Rubin DB. Statistical analysis with missing data (2nd ed). New York, NY: John Wiley & Sons. 2002.

[3]Rubin, DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons. 1987.

[4]van Buuren S, Boshuizen HC, Knook, DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in Medicine. 1999; 18: 681-694.

[5]van Buuren, S. Multiple Imputation of Discrete and Continuous Data by Fully Conditional Specification. Statistical Methods in Medical Research. 2007; 16(3): 219-42.

[6]Graham WJ, Allison OE, Tamika GD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science.2007; 8: 206–213.

[7]Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. The American Statistician. 2001; 55(3).

[8]Stone, M. Cross–validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological). 1971; 36(2): 111-147.

[9]Stone, M. Cross–validation: A review. Statistics.1978; 9(1).

[10]Mosteller F, Tukey JW. Data analysis, including statistics. Revised Handbook of Social Psychology. Addison–Wesley. 1968; 2.

[11]Su YS, Gelman A, Hill J, Yajima M. Multiple imputation with diagnostics (mi) in R: opening windows into the black box. Journal of Statistical Software.2011; 45(2):1–31.

[12]Eddings W, Marchenko Y. Diagnostics for multiple imputation in Stata. Stata Journal. 2012; 12(3): 353.