Missing Data in Orthopaedic Clinical Outcomes Research: A Sensitivity Analysis of Imputation Techniques Utilizing a Large Multicenter Total Shoulder Arthroplasty Database

Kevin A Hao, Terrie Vasilopoulos, Josie A Elwell, Christopher Roche, Keegan M Hones, Jonathan O Wright, Joseph John King, Thomas W Wright, Bradley S Schoch

INTRODUCTION: When missing data is present in clinical outcomes studies, complete-case analysis (CCA) is often performed whereby patients with missing data are excluded. While simple, CCA analysis may impart selection bias and reduce statistical power, leading to erroneous statistical results in some cases. However, there exist more rigorous statistical approaches such as single and multiple imputation, which approximate the associations that would have been present in a full dataset and preserve study power. The purpose of this study is to evaluate how statistical results differ when performed after CCA analysis versus imputation methods.

METHODS: This simulation study analyzed a sample dataset consisting of 2,204 shoulders with complete datapoints from a larger multicenter total shoulder arthroplasty database. From the sampled dataset of demographics, surgical characteristics, and clinical outcomes, we created five test datasets ranging from 100 to 2,000 shoulders and simulated 10-50% missingness in the postoperative American Shoulder and Elbow Surgeons (ASES) score and range of motion in four planes in missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR) patterns (Figure 1). Missingness in outcomes was remedied using CCA, three single imputation techniques, and two multiple imputation techniques. Imputation performance was evaluated relative to the native complete dataset using the root mean squared error (RMSE) and the mean absolute percentage error (MAPE). We also compared the mean and standard deviation (SD) of the postoperative ASES score and results of multivariable linear and logistic regression to understand the effects of imputation on study results.

RESULTS:
The average overall RMSE and MAPE were similar for MCAR (22.6 and 27.2%) and MAR (19.2 and 17.7%) missingness patterns, but were substantially poorer for NMAR (37.5 and 79.2%); sample size and the percentage of data missingness minimally affected RMSE and MAPE (Figure 2). Aggregated mean postoperative ASES scores were within 5% of the true value when missing data was remedied with CCA and all candidate imputation methods for nearly all ranges of sample size and data missingness when data were MCAR or MAR, but not when data were NMAR (Figure 3). When data were MAR, CCA resulted in overestimates of the SD. When data were MCAR or MAR, the accuracy of the regression estimate (β or OR) and its corresponding 95% CI varied substantially based on the sample size and proportion of missing data for multivariable linear regression (Figure 4), but not logistic regression (Figure 5). When data were MAR, the width of the 95% CI was up to 300% larger when CCA was used, whereas most imputation methods maintained the width of the 95% CI within 50% of the true value. Single imputation with k-nearest neighbor (kNN) method and multiple imputation with predictive mean matching (MICE-PMM) best reproduced point estimates and intervariable relationships resembling the native dataset. Availability of correlated outcome scores improved the RMSE, MAPE, accuracy of the mean postoperative ASES score, and multivariable linear regression model estimates.

DISCUSSION AND CONCLUSION: Complete-case analysis can introduce selection bias when data is MAR and results in loss of statistical power, resulting in loss of precision (i.e., expansion of the 95% CI) and predisposition to false-negative findings. Our data demonstrates imputation can reliably reproduce missing clinical data and generate accurate population estimates that closely resemble results derived from native primary shoulder arthroplasty datasets (i.e., prior to simulated data missingness). Further study of the use of imputation in clinical database research is critical, as use of CCA may lead to different conclusions in comparison with more rigorous imputation approaches.