Improved GSimp - a Flexible Missing Value Imputation Method to Support Generic Drug Development and Regulatory Assessment

Jing Wang, Xiajing Gong, Liang Zhao, and Meng Hu
Division of Quantitative Methods and Modeling, Office of Research and Standards, Office of Generic Drugs
Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland, U.S.A.

INTRODUCTION

- Missing values are commonly seen in pharmacokinetics (PK) or pharmacodynamics (PD) bioequivalence (BE) studies.
- Missing data typically appear as a mixture of different types, such as Missing Not at Random (MNAR) and Missing Completely at Random (MCAR).
- The study objective is to develop a data imputation method suitable for handling missing values in BE studies (e.g., PK or PD) to support generic drug development and regulatory assessment.

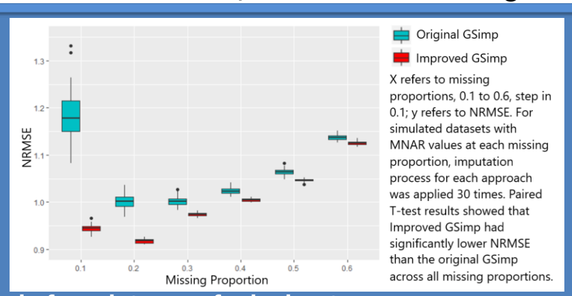
METHODS

- An iterative Gibbs sampler-based missing value imputation approach (GSimp) showed superior performance over other mainstream data imputation methods for MNAR data [1]. However, this approach determines predicted values by randomly stopping at a preset number of iterations that may lead to varying predictions and cannot handle missing data of mixed types (e.g., MNAR and MCAR mixture).
- We introduced an improved GSimp method ("Improved GSimp" thereafter) to address forementioned drawbacks by
  1. Determining predicted values based on the distribution of the whole iteration process.
  2. Handling different types of missing values by adaptively operating the imputation limit parameters.
- Simulations were conducted to compare the performance of Improved GSimp with other mainstream methods.
- As case examples, real PK and PD BE studies were used to demonstrate the usefulness of the Improved GSimp method in facilitating BE assessment.

RESULTS

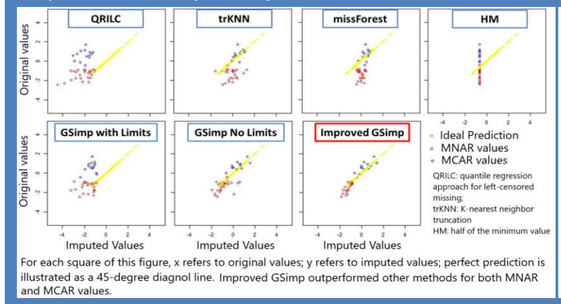
➢ Simulation 1: Compare Improved GSimp to the original GSimp for MNAR data

Normalized root mean square error (NRMSE) results showed that Improved GSimp always showed significant lower imputation error than the original GSimp with MNAR data.

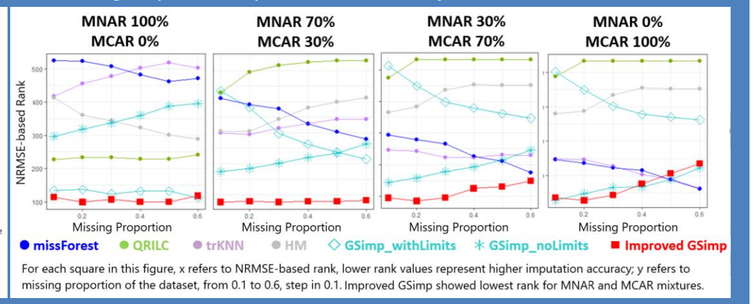


➢ Simulation 2: Compare Improved GSimp vs. Other methods for mixtures of missing types

For datasets with mixtures of MNAR and MCAR values, compared to the original GSimp and other methods, Improved GSimp always showed lowest NRMSE indicating superior imputation accuracy.



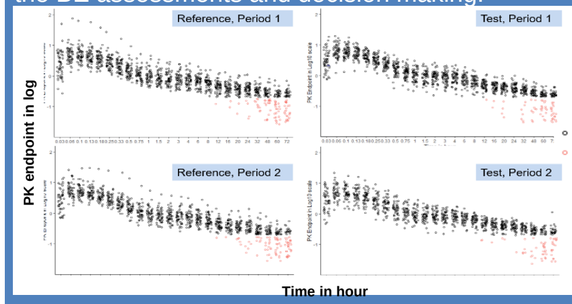
For each square of this figure, x refers to original values; y refers to imputed values; perfect prediction is illustrated as a 45-degree diagnol line. Improved GSimp outperformed other methods for both MNAR and MCAR values.



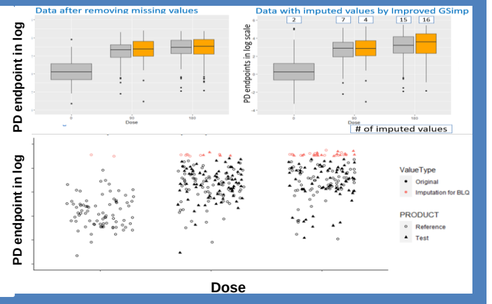
For each square in this figure, x refers to NRMSE-based rank, lower rank values represent higher imputation accuracy; y refers to missing proportion of the dataset, from 0.1 to 0.6, step in 0.1. Improved GSimp showed lowest rank for MNAR and MCAR mixtures.

➢ Case Examples: Improved GSimp applied to real PK and PD BE studies

Missing values pose challenges for BE demonstration in PK or PD studies. Conventional data imputation methods often impose unrealistic assumptions on the data (e.g., Half of the Minimum value method). The developed Improved GSimp demonstrates superior performance over other methods, thus it may be used to conduct more accurate risk analysis for missing data. As case examples, the improved GSimp was applied to the real PK (left) and PD (right) study data for handling missing data points. With imputed values, downstream regulatory analysis for the two studies can be successfully conducted, which improves the credibility of the BE analysis and supports the BE assessments and decision making.



This approach was applied to PK (left) and PD (right) data with Beyond the Limit of Quantifications (BLQ) missing values. In both figures, black dots refer to the observed data; red dots refer to the imputed data using Improved GSimp approach.



CONCLUSION(S)

In this study, we developed an improved GSimp data imputation method for handling mixed types of missing data.

The superior imputation accuracy and reliable performance of the improved GSimp showed its great potential to address the missing value imputation challenges in PK or PD BE assessment for generic drug development and regulatory assessment.

REFERENCES

[1] Lazar, Cosmin, et al. Journal of proteome research 15.4 (2016): 1116-1125.
[2] Wei, Runmin, et al. PLoS computational biology 14.1 (2018): e1005973.
[3] Lenz, Michael, et al. Plos one 15.12 (2020): e0243487.

ACKNOWLEDGEMENTS

Dr. Jing Wang was supported by an appointment to the Research Participation Program at the Center for Drug Evaluation and Research, FDA, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. FDA. This poster reflects the views of the authors and should not be construed to represent FDA's views or policies.

Special thanks for the support and guidance from Dr. Lei Zhang and Dr. Robert Lionberger, Office of Research and Standards, Office of Generic Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration.