smdi: An R package to perform structural missing data investigations

**Authors**: Janick Weberpals1, Sudha R. Raman2, Pamela A. Shaw3, Hana Lee4, Massimiliano Russo1, Bradley G. Hammill2, Sengwee Toh5, John G. Connolly5, Kimberly J. Dandreo5, Fang Tian4, Wei Liu4, Jie Li4, José J. Hernández-Muñoz4, Robert J. Glynn1, Rishi J. Desai1

Author affiliations:

1 Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

2 Department of Population Health Sciences, Duke University School of Medicine, Durham, NC

3 Biostatistics Division, Kaiser Permanente Washington Health Research Institute, Seattle, WA

4 Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD

5 Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA

**Correspondence:**

Janick Weberpals, RPh, PhD

Division of Pharmacoepidemiology and Pharmacoeconomics,

Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School,

1620 Tremont Street, Suite 3030-R, Boston, MA 02120, USA

Phone: 617-278-0932

Fax: 617-232-8602

Email: [jweberpals@bwh.harvard.edu](mailto:jweberpals@bwh.harvard.edu)

**Word count:** xxxx words

**Tables:**

**Figures:**

**Supplementary material:**

**Short running title**: An R package to perform structural missing data investigations

**Keywords:** Missing data, Confounder, EHR, R, Software,

**Funding:** This project was supported by Master Agreement 75F40119D10037 from the US Food and Drug Administration (FDA).

**Disclosures/COI:res/COI:** The FDA approved the study protocol, statistical analysis plan and reviewed and approved this manuscript. Coauthors from the FDA participated in the results interpretation and in the preparation and decision to submit the manuscript for publication. The FDA had no role in data collection, management, or analysis. The views expressed are those of the authors and not necessarily those of the US FDA. Janick Weberpals reports prior employment by Hoffmann-La Roche and previously held shares in Hoffmann-La Roche. Pamela Shaw is a named inventor on a patent licensed to Novartis by the University of Pennsylvania for an unrelated project. Sengwee Toh serves as a consultant for Pfizer, Inc. and TriNetX, LLC.. Robert J Glynn has received research funding through his employer from Amarin, Kowa, Novartis, and Pfizer. Dr. Desai reports serving as Principal Investigator on investigator-initiated grants to the Brigham and Women’s Hospital from Novartis, Vertex, and Bristol-Myers-Squibb on unrelated projects. All remaining authors report no disclosures or conflicts of interest.

Analytical code and data sharing statement: This manuscript was written using Quarto version 1.3.433 (<https://quarto.org/>) and R version 4.1.2. All R code and materials can be found at <https://gitlab-scm.partners.org/drugepi/smdi-manuscript>. Detailed information on packages and versions can be found in the renv.lock file in the linked repository. The R package and data used in this study can be downloaded from <https://janickweberpals.gitlab-pages.partners.org/smdi/>.

**Acknowledgments:** None.

# Abstract

|  |
| --- |
| Highlights |
| * … |

# Introduction

# Methods

# Results

# Discussion

# References

# Figures