# A Comparison of Approaches for Missing Outcome Data in Randomized Clinical Trials with Censored Time-to-event Endpoints

Andrea Bellavia, Sabina A. Murphy

TIMI Study Group, Brigham and Women's Hospital, Harvard Medical School

abellavia@bwh.harvard.edu

ENAR Spring Meeting, Baltimore, March 13 2024

Acknowledgments:

- Dr. Marc S. Sabatine, TIMI Study Group
- Dr. Stephen D Wiviott, TIMI Study Group
- Min Guo, Department of Biostatistics, Harvard T.H. Chan School of Public Health

Disclaimer:

The authors declare no conflict of interest. AB and SAM are members of the TIMI Study Group which has received institutional research grant support through Brigham and Women's Hospital from: Abbott, Abiomed, Inc., Amgen, Anthos Therapeutics, ARCA Biopharma, Inc., AstraZeneca, Boehringer Ingelheim, Daiichi-Sankyo, Ionis Pharmaceuticals, Inc., Janssen Research and Development, LLC, MedImmune, Merck, Novartis, Pfizer, Regeneron Pharmaceuticals, Inc., Roche, Saghmos Therapeutics, Inc., Siemens Healthcare Diagnostics, Inc., Softcell Medical Limited, The Medicines Company, Verve Therapeutics, Inc., Zora Biosciences

# 1) Background and Introduction

# 1) Background and Introduction

- Randomized Clinical Trials (RCT) are commonly designed in a prospective fashion, following and comparing efficacy outcomes between groups of individuals over time using survival analysis
- The outcome is defined with a pair of random variables $(T, \delta)$
  - $\delta_i = (0, 1)$ is the indicator of whether the event of interest occurred
  - $T_i = min(Y_i, C_i)$ is the time to either the event $(Y)$ or censoring $(C)$, whichever occurred first

- The more common reason for missing outcome data is censoring; individuals might be lost to follow-up or withdraw consent and leave the study
- Even when the censoring reason is known (e.g. safety event leading to discontinuation), the intention-to-treat (ITT) approach remains mostly agnostic to this information
- Limitations of standard approaches are recognized and sensitivity analyses are recommended[1]
- However, methods comparison have generally relied on simulated data alone rather than using actual RCT data

[1] Altman DG. Missing outcomes in randomized trials: addressing the dilemma. Open Medicine. 2009;3(2):e51.; Tan PT, Cro S, Van Vogt E, Szigeti M, Cornelius VR. A review of the use of controlled multiple imputation in randomised controlled trials with missing outcome data. BMC Medical Research Methodology. 2021 Dec;21(1):1-7.

# Goals of the Presentation

- Review the concepts of missing at random/not-at random, and informative/non-informative censoring in RCTs
- Review recent approaches for multiple imputation (MI) that can account for informative censoring in RCTs
- Validate results from a real RCT using safety event information to inform sensitivity analyses for efficacy endpoints
- Simulate more extreme censoring scenarios of this dataset to explore the impact of different imputation approaches on efficacy results

# 2) Multiple imputations for informative censoring

# Missing Outcome Data in RCT

- Primary results of phase 3 RCTs are generally reported with Kaplan-Meier (KM) curves and Cox regression
- Both KM and Cox (as well as several other common techniques in time-to-event analysis) makes the assumption of non-informative (or independent) censoring: participants who drop out of the study should do so due to reasons unrelated to the study
- This corresponds to assuming a missing at random (MAR) distribution for the unobserved event times
- Under MAR, valid inference can be obtained from the likelihood of the observed data only

# Informative Censoring and Missing Not At Random (MNAR)

- In the presence of informative censoring, the assumption of MAR is not met
- Trial results, when obtained with statistical approaches that require non-informative censoring, should be evaluated in sensitivity analysis
- Controlled MI provides a flexible and intuitive tool for this sensitivity analysis[2]

[2] Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide. Statistics in medicine. 2020 Sep 20;39(21):2815-42.

# Controlled MI for Informative Censoring

- Main approaches are risk-imputation and $\gamma$-imputation [focus here on the latter]
- $\gamma$-imputation uses MI to relax the independent (non-informative) assumptions in a Cox model[3]
- Both approaches are implemented in the `informativeCensoring` R package

[3] Jackson D, White IR, Seaman S, Evans H, Baisley K, Carpenter J. Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation. Statistics in medicine. 2014 Nov 30;33(27):4681-94.

# Gamma imputation

$\gamma$-imputation uses the Fleming and Harrington definition of independent censoring (the hazard of failure at $t$ is equal to the hazard of failure at $t$ given that censoring has not occurred): $h(t, Z_i) = h(t|C_i > t, Z_i)$, to extend the Cox model as follows:

$$h(t_i|C_i > t, Z_i) = h_0(t)\exp(\beta Z_i)\exp(\gamma_i)$$

where $\gamma_i$ is the log-HR comparing censored vs uncensored individuals.

- $\gamma_i = 0$ -> uninformative censoring
- $\gamma_i > 0$ -> the $i_{th}$ participant is at elevated risk of the event after censoring
- $\gamma_i < 0$ -> the $i_{th}$ participant is at lower risk of the event after censoring

This approach provides an ideal setting for sensitivity analyses as several values of $\gamma$ can be tested

# Practical procedure

- Define which participants should have the outcome data imputed
- For those participants, define their value of $\gamma$
- Use controlled MI to create imputed dataset where there are no missing values
- Present regression parameters (HRs) over bootstrap samples
- Compare results at different levels of $\gamma$

# 3) Illustrative Example

# 3) Illustrative Example

- Data from DECLARE-TIMI58[4]
- 17,160 patients with a 1:1 randomization followed for a median of 4.2y (IQR: 3.9-4.4). Results indicated non-inferiority with respect to major adverse cardiovascular diseases (MACE, HR=0.93, 95% CI:0.84-1.03) and superiority with respect to hospitalization for heart failure or CVD death (HHF/CVD, HR=0.83, 95% CI:0.73-0.95), the 2 co-primary endpoints of the trial
- Primary analyses were based on intention-to-treat as per FDA guidelines

[4] randomized, double-blind, multinational, placebo-controlled, phase 3 trial of dapagliflozin in patients with type 2 diabetes and established atherosclerotic cardiovascular disease or multiple risk factors for atherosclerotic cardiovascular disease

- Safety data on several adverse events (AE) were collected and compared between treatment arms
- If the AE is associated with treatment, and leads to follow-up withdraw, this censoring would be informative

**Table 2. Safety Events.***

| Event | Dapagliflozin (N=8574) | Placebo (N=8569) | Hazard Ratio (95% CI) | P Value |
|---|---|---|---|---|
| | *no. (%)* | | | |
| Serious adverse event | 2925 (34.1) | 3100 (36.2) | 0.91 (0.87–0.96) | <0.001 |
| Adverse event leading to discontinuation of trial regimen | 693 (8.1) | 592 (6.9) | 1.15 (1.03–1.28) | 0.01 |
| Major hypoglycemic event | 58 (0.7) | 83 (1.0) | 0.68 (0.49–0.95) | 0.02 |
| Diabetic ketoacidosis | 27 (0.3) | 12 (0.1) | 2.18 (1.10–4.30) | 0.02 |
| Amputation | 123 (1.4) | 113 (1.3) | 1.09 (0.84–1.40) | 0.53 |
| Fracture | 457 (5.3) | 440 (5.1) | 1.04 (0.91–1.18) | 0.59 |
| Symptoms of volume depletion | 213 (2.5) | 207 (2.4) | 1.00 (0.83–1.21) | 0.99 |
| Acute kidney injury | 125 (1.5) | 175 (2.0) | 0.69 (0.55–0.87) | 0.002 |
| Genital infection | 76 (0.9) | 9 (0.1) | 8.36 (4.19–16.68) | <0.001 |
| Urinary tract infection | 127 (1.5) | 133 (1.6) | 0.93 (0.73–1.18) | 0.54 |
| Cancer | 481 (5.6) | 486 (5.7) | 0.99 (0.87–1.12) | 0.83 |
| Bladder cancer | 26 (0.3) | 45 (0.5) | 0.57 (0.35–0.93) | 0.02 |
| Breast cancer | 36 (0.4) | 35 (0.4) | 1.02 (0.64–1.63) | 0.92 |
| Hypersensitivity | 32 (0.4) | 36 (0.4) | 0.87 (0.54–1.40) | 0.57 |
| Hepatic event | 82 (1.0) | 87 (1.0) | 0.92 (0.68–1.25) | 0.60 |

* Additional details, data sources, and a complete list of serious adverse events are provided in the Supplementary Appendix. P values and 95% confidence intervals have not been adjusted for multiple comparisons.

Table 2 from Wiviott et al. NEJM. 2019

# Analyses

1. Controlled MI to validate trial results by relaxing the non-informative censoring assumption implicitly made for discontinued observations.
2. Controlled MI to impute AE regardless of whether they resulted in censoring.

- These results allow assessing the potential effect of severe informative censoring, had the AE led to discontinuation of trial follow-up for efficacy analysis, thus informing general use in settings where AE might not be available or reported

We evaluated several scenarios based on AEs reported in the previous table:

A) Controlled MI for censored individuals ($\sim 8\%$) due to *adverse events leading to discontinuation*: range of $\gamma < 0$ with moderate to null effect
B) Controlled MI had those with *genital infection* ($\sim 1\%$) been censored: range of $\gamma < 0$ with strong effect
C) Controlled MI had those with *major hypoglycemic events* ($\sim 1\%$) been censored: range of $\gamma > 0$ with moderate effect
D) Controlled MI had those with any adverse event ($\sim 35\%$) been censored: a range of individual-specific $\gamma$

3. Finally, we simulated more extreme scenarios based on the real data distributions observed in the dataset. Specifically:

- Censoring event as in A but affecting $\sim 50\%$ participants rather than $\sim 8\%$
- Censoring event as in B but affecting $\sim 25\%$ participants rather than $\sim 1\%$
- Censoring event as in C but affecting $\sim 25\%$ participants rather than $\sim 1\%$

For comparison, we then repeated all settings using standard MI (MAR assumption)

# Results: sensitivity analysis on observed trial data

| Setting | MACE (HR) | | HHF/CVD | (HR) |
|---|---|---|---|---|
| ITT | 0.93 | | 0.83 | |
| On trt | 0.93 | | 0.81 | |
| | MI | c-MI | MI | c-MI |
| A | 0.94 | 0.93 | 0.84 | 0.82 |
| B | 0.94 | 0.93 | 0.84 | 0.82 |
| C | 0.94 | 0.93 | 0.84 | 0.82 |
| D | 0.94 | 0.93 | 0.83 | 0.82 |

Only showing results for one set of $\gamma$. Results consistent over evaluated ranges:

A) *AE leading to discontinuation* ($\sim 8\%$): $\gamma = -0.5$
B) *Genital infection* ($\sim 1\%$) : $\gamma = -2$
C) *Major hypoglycemic events* ($\sim 1\%$) : $\gamma = 0.5$
D) Any AE ($\sim 35\%$); varying $\gamma$ based on the specific event that led to censoring

c-MI=controlled-MI;both ITT and on-trt make MAR assumption

# Results: simulated scenarios

| Setting | MACE (HR) | | HHF/CVD | (HR) |
|---|---|---|---|---|
| ITT | 0.93 | | 0.83 | |
| On trt | 0.93 | | 0.81 | |
| | MI | c-MI | MI | c-MI |
| A | 0.90 | 0.93 | 0.80 | 0.84 |
| B | 0.89 | 0.94 | 0.81 | 0.85 |
| C | 0.91 | 0.93 | 0.80 | 0.86 |

A) Censoring due to *adverse events leading to discontinuation* occurring in 40% of participants (previously $\sim$ 8%)
B) Censoring due to *genital infection* occurring in 25% of participants (previously $\sim$ 1%)
C) Censoring due to *major hypoglycemic events* occurring in 25% of participants (previously $\sim$ 1%)

c-MI=controlled-MI

# 5) Summary and Discussion

# 5) Summary and Discussion

- Results from DECLARE-TIMI58 were not affected by the assumption of non-informative censoring made for discontinued participants and ITT assumptions
- Based on simulations, failure to account for informative censoring would have affected the results only at substantially high censoring proportions, which were not observed in the trial
- Controlled MI might be of critical importance in studies with long follow-up and subject to strong informative censoring

# Practical recommendations

- Detailed account and analysis of safety events that led to potentially censored observations (as in the DECLARE trial) can be used to inform the sensitivity analysis parameters to validate results in the presence of potential informative censoring
- Censored observations with no explanation (e.g. lost to follow-up) can be investigated with controlled MI to assess the hypothetical level of $\gamma$ that would affect ITT results.
- Tools to address missing outcome data are available and should be included routinely as sensitivity analysis in protocols and analysis plans

Thanks for your attention!

Contact:

abellavia@bwh.harvard.edu

timi.org/biostatistics/

@TimiStudyGroup @andreabellavia