A Multiple Imputation Method for Sensitivity Analysis of Time-to-event Data with possible Informative Censoring

Yue Zhao (Sophie)
Merck Research Lab.
North Wales. PA
• **Important principles for conducting an unbiased study**
  – Randomization
  – Intent-to-treat (ITT) analysis

• **Missing data due to withdrawal from study**
  – Undermine the comparability of randomized groups
  – Potential bias for treatment comparison
  – In many situations, MAR appears to be reasonable; but the possibility of MNAR can never be ruled

• **Follow-up of patients after discontinuation from study mediation**
  – Support ITT analysis
  – Complicate interpretation: receive effective rescue therapies
Methods of handling missing data in time-to-event analysis

• Prematurely discontinue follow-up for the assigned treatment prior to event or the end of study

• **Censor patients at the time of discontinuation**
  – Non-informative Censoring (i.e., MAR-like assumption)
  – Often chosen as the primary analysis
  – Per-protocol (PP) analyses for the ITT population
  – May not be consistent with the ITT principle

• **Manage discontinuation from treatment as clinical failure**
  – A composite endpoint (time to event of interest or withdrawals)
  – Impute missing as failure
  – Worst-case analysis and worst-comparison analysis
  – Not primary analysis, but sensitivity analysis
### Trial example – Maintenance treatment of bipolar disorder

**Primary efficacy endpoint:** time-to-intervention for any mood episode

<table>
<thead>
<tr>
<th>Disposition</th>
<th>Overall N</th>
<th>Overall %</th>
<th>Placebo N</th>
<th>Placebo %</th>
<th>Test treatment N</th>
<th>Test treatment %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Completed study without episode</td>
<td>46</td>
<td>15.3</td>
<td>15</td>
<td>10</td>
<td>31</td>
<td>20.7</td>
</tr>
<tr>
<td>2 Intervention for a mood episode</td>
<td>157</td>
<td>52.3</td>
<td>82</td>
<td>54.7</td>
<td>75</td>
<td>50</td>
</tr>
<tr>
<td>3 Discontinued study prematurely</td>
<td>97</td>
<td>32.3</td>
<td>53</td>
<td>35.3</td>
<td>44</td>
<td>29.3</td>
</tr>
</tbody>
</table>

- **Adverse event:**
  - Overall: 24 (8%), Placebo: 16 (10.7%), Test treatment: 8 (5.3%)
- **Consent withdraw:**
  - Overall: 28 (9.3%), Placebo: 13 (8.7%), Test treatment: 15 (10)
- **Lost to follow-up:**
  - Overall: 20 (6.7%), Placebo: 8 (5.3%), Test treatment: 12 (8)
- **Protocol violation:**
  - Overall: 10 (3.3%), Placebo: 3 (2), Test treatment: 7 (4.7)
- **Other (including missing data):**
  - Overall: 15 (5), Placebo: 13 (8.7), Test treatment: 2 (1.3)

**Administrative Censoring**

1. Event
2. Premature Discontinuation/withdrawal (missing outcome)
Primary analysis (censors patients at the time of dropout) and sensitivity analyses

<table>
<thead>
<tr>
<th>Analysis method</th>
<th>Semi-parametric analysis (Cox PH model)</th>
<th>Non-parametric analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (Std Err)</td>
<td>HR (95% CI)</td>
</tr>
<tr>
<td>(1A) MAR-like</td>
<td>-0.393 (0.161)</td>
<td>0.675 (0.493, 0.925)</td>
</tr>
<tr>
<td>(1B) Worst case</td>
<td>-0.484 (0.127)</td>
<td>0.616 (0.481, 0.790)</td>
</tr>
<tr>
<td>(1C) Worst comparison</td>
<td>0.033 (0.144)</td>
<td>1.033 (0.779, 1.371)</td>
</tr>
</tbody>
</table>

More realistic approaches are worthy of consideration for sensitivity analyses to address robustness of conclusions to possibly informative censoring of time-to-event data.
General ideas for sensitivity analysis
– recommended by the panel on handling missing data in clinical trials NRC (2010)

- Conduct sensitivity analyses based on different assumptions to evaluate the potential impact of missing outcomes on the trial conclusion
- The primary MAR assumption serves as an anchor point for the sensitivity analysis to assess the robustness of the treatment effect inferences
- Supplement the primary analysis with a number of modifications, relating treatment effect inferences to one or more parameters that capture departures from the MAR.
- The degree to which conclusions are stable across such analyses provides an indication of the confidence that can be placed in them
- Modifications to a basic model
  - Untestable assumptions regarding the missing response distribution (pattern-mixture model)
  - How the probability of missingness related to the unobserved outcomes (selection model)
Proposed sensitivity analysis method

- **Multiple imputation (MI):** impute experiences for discontinued patients during their unobserved remaining times as if they continued to be followed until the end of study, based on the conditional survival distribution.

- **Sensitivity parameter ($\theta$):** fixed hazard ratio for a withdrawer having an event after the corresponding censor time relative to the patients still remaining on their treatment, allowing for different post-discontinuation experience. (Pattern-mixture model).

- **Sensitivity assessment:**
  - Anchors on the primary MAR assumption ($\theta = 1$)
  - Investigates the impact of departures from MAR assumption by summarizing the treatment effect inference as a function of $\theta$ over a plausible range.

Kaplan-Meier multiple imputation (KMMI) procedure

- A single treatment group of n patients with the same planned follow-up time t*
- M distinct failure times \((t_1 < t_2 < \cdots < t_M)\)
- The Kaplan-Meier estimates \(\hat{S}(t)\) have support on the observed failure times \((t_1, t_2, \ldots, t_M)\)
- K distinct discontinuation times \((c_1 < c_2 < \cdots < c_K)\) before t*
- With k indexing the censoring times before t*, \(t_{k,j}\) denotes the jth failure time after \(c_k\), \(t_{k,0}\) denotes the latest failure time prior to \(c_k\) if \(c_k \geq t_1\) and \(t_{k,0} = 0\) if \(c_k < t_1\).
Estimate the survival rates for all $K + 1$ censoring times ($t^*$ and $c_k$’s)

Obtain $\hat{S}(c_k)$ by linear interpolation, when $c_k < t_M$

$$\hat{S}(c_k) = \hat{S}(t_{k,0}) - \frac{c_k - t_{k,0}}{t_{k,1} - t_{k,0}} \times (\hat{S}(t_{k,0}) - \hat{S}(t_{k,1}))$$

$t_{k,0}$: Latest failure time prior to $c_k$, when $t_1 \leq c_k$
$t_{k,j}$: The $j$th failure time after $c_k$, when $c_k < t_M$, $j = 1, \ldots, J_k$
Estimate the survival rates for all $K + 1$ censoring times ($t^*$ and $c_k$’s)

Obtain $\hat{S}(t^*)$ and $\hat{S}(c_k)$ when $c_k > t_M$

1. $\hat{S}((t_M - t_{M-f})|t > t_{M-f}) = \frac{\hat{S}(t_M)}{\hat{S}(t_{M-f})} = \exp\{-h \times (t_M - t_{M-f})\}$ to determine $h$

2. $\hat{S}(t^*) = \hat{S}(t_M) \times \hat{S}((t^* - t_M)|t > t_M) = \hat{S}(t_M) \times \exp\{-h \times (t^* - t_M)\}$

Linear interpolation for $c_k$ between $t_M$ and $t^*$:

$$\hat{S}(c_k) = \hat{S}(t_M) - \frac{c_k - t_M}{t^* - t_M} \times (\hat{S}(t_M) - \hat{S}(t^*))$$
Construct the estimated conditional failure time distribution given a fixed hazard ratio $\theta$ (sensitivity parameter)

- **Sensitivity parameter**: the hazard ratio $\theta$ for a discontinued patient having an event after $c_k$ relative to the patients still remaining on their treatment.

- Under the proportional hazard assumption, the survival function for a discontinued patient at time $t$ ($t > c_k$) is $\hat{S}(t)^{\theta}$

- Conditional on not having the event by the time $c_k$, the probability of having an event (PDF) in $[c_k; t_{k;1}]$ and $[t_{kj}; t_{kj+1}]$ is given by
  \[
  \hat{f}_{k,0}(\theta) = \frac{\hat{S}(c_k)^{\theta} - \hat{S}(t_{k;1})^{\theta}}{\hat{S}(c_k)^{\theta}} \\
  \text{and} \\
  \hat{f}_{k,j}(\theta) = \frac{\hat{S}(t_{kj})^{\theta} - \hat{S}(t_{kj+1})^{\theta}}{\hat{S}(c_k)^{\theta}}
  \]

- Conditional cumulative incidence function (CDF) by $t$ in $[t_{kj}; t_{kj+1}]$, for withdrawal at $c_k$
  \[
  \hat{F}_{k,j}(\theta) = \sum_{j' = 0}^{j} \hat{f}_{k,j'}(\theta) = 1 - \frac{\hat{S}(t_{kj+1})^{\theta}}{\hat{S}(c_k)^{\theta}}
  \]

  $t_{kj+1}$: The $(j + 1)$th failure time after $c_k$ and $j = 0, 1, 2, \cdots, J_k$ with $t_{kJ_k+1} = t^*$
Multiple Imputation Scheme: An Illustration of Imputing failure time for $c_1$
Multiple Imputation Scheme: An Illustration of Imputing failure time for $c_1$

A random number $p$ drawn from $U(0,1)$

The imputed failure time is $t$
Multiple Imputation Scheme: An Illustration of Imputing failure time for $c_1$

A random number $p$ drawn from $U(0,1)$

The imputed censor time is $t^*$

No failure event by $t^*$
Multiple Imputation Scheme

- Generate a random number \( p \) from \( U(0,1) \), failure time is imputed by linear interpolation

- For \( c_k < t_M \)
  - If \( 0 \leq p \leq \hat{F}_{k,0}(\theta) \), impute failure time \( t_{k}^{(l)} \) between \( c_k \) and \( t_{k,1} \) as
    \[
    c_k + (t_{k,1} - c_k) \times \frac{p}{\hat{F}_{k,0}(\theta)}
    \]
  - If \( \hat{F}_{k,j}(\theta) \leq p \leq \hat{F}_{k,j+1}(\theta) \) for \( j = 0,1,2,\ldots,(J_k - 1) \), then impute \( t_{k}^{(l)} \) between \( t_{k,j+1} \) and \( t_{k,j+2} \) as
    \[
    t_{k,j+1} + (t_{k,j+2} - t_{k,j+1}) \times \frac{p - \hat{F}_{k,j}(\theta)}{\hat{F}_{k,j+1}(\theta) - \hat{F}_{k,j}(\theta)}
    \]
  - If \( p > \hat{F}_{k,J_k+1}(\theta) \), then manage the patient as having no event by the end of follow-up time \( t^* \)
Multiple Imputation Scheme

- For $t_M < c_k < t^*$
  - If $p \leq \hat{f}_{k,0}(\theta)$, then impute $t^{(l)}_k$ between $c_k$ and $t^*$ as
    \[ c_k + (t^* - c_k) \times \frac{p}{\hat{F}_{k,0}(\theta)} \]
  - Otherwise, manage the patient as having no event by $t^*$

- Repeat $L$ times to generate multiple imputed data sets

- The imputation is performed separately for the patients with premature discontinuation in each treatment group.

- In practice, the appropriate number of imputations should be investigated more closely, especially when the fraction of missing information is large (Horton and Lipsitz, 2001).
Parameter Estimations (rule of Rubin)

- **Point estimates**
  \[ \bar{\beta} = \frac{1}{L} \sum_{l=1}^{L} \hat{\beta}^{(l)} \]
  where \( \hat{\beta}^{(l)} \) denote the point estimate for \( \beta \) from the \( l \)-th data set

- **Variance component**
  \[
  \hat{V}_\beta = V_\beta + (1 + L^{-1})B_\beta \\
  V_\beta = \frac{1}{L} \sum_{l=1}^{L} \hat{V}_\beta^{(l)} \\
  B_\beta = \frac{1}{L - 1} \sum_{l=1}^{L} (\hat{\beta}^{(l)} - \bar{\beta})^2
  \]
  where \( \hat{V}_\beta^{(l)} \) denote the variance estimate from the \( l \)-th data set

- **Logrank (or Wilcoxon) statistic:**
  \[
  \hat{Z}^{(l)} = \frac{\hat{\beta}^{(l)}}{(\hat{V}_\beta^{(l)})^{1/2}} \]
  can serve as \( \hat{\beta}^{(l)} \) with corresponding \( \hat{V}_Z^{(l)} = 1 \)
Performance of KMMI method under $\theta = 1$ (L=50)

Survivor curves obtained from the conventional and the KMMI method

![Graph showing survivor curves](image-url)
Performance of KMMI method under $\theta = 1$ (L=50)

Comparison of analyses from the conventional (MAR-like) and KMMI method

<table>
<thead>
<tr>
<th>Analysis method</th>
<th>Semi-parametric analysis (Cox PH model)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Err</td>
<td>HR (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>1. Original primary analyses</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1A) MAR-like</td>
<td>-0.393</td>
<td>0.161</td>
<td>0.675 (0.493, 0.925)</td>
<td>0.0144</td>
</tr>
<tr>
<td>2. Kaplan-Meier multiple imputation (KMMI) method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2A) $\theta = 1$</td>
<td>-0.322</td>
<td>0.160</td>
<td>0.724 (0.530, 0.991)</td>
<td>0.0436</td>
</tr>
</tbody>
</table>
Performance of KMMI method under $\theta = 1$ (L=50)

Cumulative hazard curves

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Coefficient (SE)</th>
<th>HR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole period</td>
<td>-0.393 (0.161)</td>
<td>0.675 (0.493, 0.925)</td>
<td>0.0144</td>
</tr>
<tr>
<td>0-3 weeks</td>
<td>-0.728 (0.284)</td>
<td>0.483 (0.277, 0.842)</td>
<td>0.0103</td>
</tr>
<tr>
<td>4-5 weeks</td>
<td>-0.700 (0.356)</td>
<td>0.497 (0.247, 0.998)</td>
<td>0.0495</td>
</tr>
<tr>
<td>6-20 weeks</td>
<td>-0.096 (0.342)</td>
<td>0.908 (0.465, 1.774)</td>
<td>0.7782</td>
</tr>
<tr>
<td>21-76 weeks</td>
<td>0.138 (0.367)</td>
<td>1.148 (0.559, 2.357)</td>
<td>0.7065</td>
</tr>
</tbody>
</table>
Comparison of analyses from the conventional (MAR-like) KMMI, and PHMI method

<table>
<thead>
<tr>
<th>Analysis method</th>
<th>Semi-parametric analysis (Cox PH model)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
</tr>
<tr>
<td>1. Original primary analyses</td>
<td></td>
</tr>
<tr>
<td>(1A) MAR-like</td>
<td>-0.393</td>
</tr>
<tr>
<td>2. Kaplan-Meier multiple imputation (KMMI) method</td>
<td></td>
</tr>
<tr>
<td>(2A) $\theta = 1$</td>
<td>-0.322</td>
</tr>
<tr>
<td>3. Proportional hazard multiple imputation (PHMI) method</td>
<td></td>
</tr>
<tr>
<td>(3A) $\theta = 1$</td>
<td>-0.388</td>
</tr>
</tbody>
</table>
Specify the sensitivity parameter ($\theta$)

- Usually $\theta_T > \theta_P$ is specified
- With $\theta_P = 1$, $\theta = \theta_T / \theta_P = \theta_T$ becomes a single parameter for calibrating sensitivity analyses.

The choice of $\theta$:
- Arbitrary (such as 1:05; 1:10; 1:15, etc.)
  $\theta = \infty$ corresponds to the worst comparison analysis
- A range ($L; U$), where $(1/U; 1/L)$ is a range of hazard ratios from previous related studies or clinical judgment for the comparison of effective medicines with placebo.
  - $\theta = \theta_T / \theta_P = \theta_T \subset (1; 2.5)$ where $(1/U; 1/L) = (1/2.5 = 0.4; 1)$ is a range of hazard ratios.
  - 0.4 might represent a reasonably large effect size for a clearly effective treatment versus placebo
Sensitivity analyses (KMMI)

HR with 95% CI with Cox model

P-values of Wald test (Cox model), Logrank test, and Wilcoxon test
Sensitivity analyses (PHMI)

HR with 95% CI with Cox model

P-values of Wald test (Cox model), Logrank test, and Wilcoxon test
Summary

• Does not attempt to get more accurate estimation for treatment effects, but to understand robustness of the treatment effect inference.

• Address the question for what the long-term benefit of initial assignment would be if patients with premature discontinuation were followed to the end of the study without other treatment.

• If it is unlikely to reach the values that would alter the study conclusion, then the results of the primary analysis is considered to be robust from a clinical perspective.

• When the inference about treatment effects could be overturned for plausible values of sensitivity parameter, the primary analysis results should be viewed as equivocal, therefore should be interpreted with more caution.
Summary

• **KMMI method**: evaluate the implication of missing outcomes and the non-proportional hazards related to withdrawal simultaneously.

• **PHMI method**: the missing data issue is addressed separately in its own right

• Anchors on the primary MAR assumption, directly exploring the effect of departures from the non-informative censoring assumption made in the primary analysis.

• Calibration toward the worst comparison analysis through how it penalizes premature discontinuation for the test treatment.

• The interpretation of the sensitivity parameter is transparent in the sense that the parameter is based on a standard measurement for analyzing time-to-event data, and consequently may be more understandable to non-statisticians.
Acknowledgement

Many thanks to Dr. Gary Koch and other co-authors for their helps and comments during the development of the methods and the paper.
Thank You Very Much