Combining Survival Trials Using Aggregate Data
Based on Misspecified Models

Tinghui Yu *, Yabing Mai †, Sherry Liu *, Xiaofei Hu †

* FDA, Center for Devices and Radiological Health
† Merck Research Laboratories

2015 Midwest Biopharmaceutical Statistics Workshop
The contents of this presentation are our own, and do not necessarily represent the views and/or policies of the Food and Drug Administration or Merck Research Laboratories or their staff.
Motivating example
Research literature
New method
A real life example
Motivating Example
Targeted treatments are validated with patients selected by specific biomarkers.

Diagram:

- Biomarker
  - +: Enroll -> Randomization
  - -: Exclude
Targeted treatments are validated with patients selected by specific biomarkers.
Targeted treatments are validated with patients selected by specific biomarkers.
Targeted treatments are validated with patients selected by specific biomarkers

\[ h_0(t)e^{\tilde{\theta}'Z} \text{ vs } h_0(t) \]
Clinical trials producing \((X_i, \delta_i, Z_i), \; i = 1, \ldots, n\).

- Right censored survival time \(X_i = \min(T_i, C_i)\)
- Survival time \(T_i\) follows a Cox proportional hazard model

\[
h(t|Z) = h_0(t) e^{\beta'Z}, \quad H_0(t) = \int_0^t h_0(u)du.
\]

- Independent censoring time \(C_i\)
- \(\delta_i = 1\) if \(T_i \leq C_i\), \(= 0\) otherwise
- Log hazard ratios \(\tilde{\beta} = (\beta_1, \ldots, \beta_k)\), hazard ratios \(\tilde{b} = e^\tilde{\beta}\)
- Covariates \(Z_i = (Z_{i1}, \ldots, Z_{ik})\)

\[
Z_i \text{ iid } \sim f(z_1, \ldots, z_k)
\]

- Treatment arm indicator \(Z_1\) (1=treatment, 0=control) is independent from \((Z_2, \ldots, Z_k)\) with the randomization ratio defined by \(q = P(Z_1 = 1)\)
**Trial 1**: \((X_i, \delta_i, Z_i), \ i = 1, \ldots, n\) follows a Cox model

\[
h(t|Z) = h_0(t)e^{\tilde{\alpha}'Z},
\]
with MPLE \(\hat{\alpha} = \{\hat{\alpha}_1, \ldots, \hat{\alpha}_k\}\):

\[
\hat{\alpha} \approx N_k(\alpha, I(\hat{\alpha})^{-1}).
\]

**Trial 2**: \((Y_j, \delta_j, Z_j), \ j = 1, \ldots, m\) follows an other Cox model

\[
h(t|Z) = h_0(t)e^{\tilde{\beta}'Z},
\]
with an independent MPLE \(\hat{\beta} = \{\hat{\beta}_1, \ldots, \hat{\beta}_k\}\).

Assumptions:

- Trial 1 \(\perp\) Trial 2
- \(\tilde{\alpha} \neq \tilde{\beta}, \ \alpha_1 < \beta_1\)
- Same baseline hazard function \(h_0(t)\)
- Fixed ratio between sample sizes: \(\frac{n}{n+m} = p\) as \(m, n \to \infty\)
**Question:** How to estimate the overall treatment effect $\tilde{\theta}$?

Combined trial

$$(W_i, \delta_i, Z_i) = Trial1 + Trial2$$

$$= \{(X_i, \delta_i, Z_i), i = 1, \ldots, n\} \cup \{(Y_j, \delta_j, Z_j), j = 1, \ldots, m\}$$

Assuming a misspecified Cox model (working model) for $W_i$

$$h(t|Z) = h_0(t)e^{\tilde{\theta}'Z}.$$ 

- $\tilde{\theta} = (\theta_1, \ldots, \theta_k) =$?
- Same baseline hazard function $h_0(t)$
- Individual patient data $X_i, Y_j$ may not be available
- $\hat{\alpha}, I(\hat{\alpha})^{-1}, \hat{\beta}$ and $I(\hat{\beta})^{-1}$ are reported
Literature Review
Pooled patient line data \((N = n + m)\)

\[
\{(W_i, \delta_i, Z_i), i = 1, \ldots, n + m\} = \{(X_i, \delta_i, Z_i), i = 1, \ldots, n\} \cup \{(Y_j, \delta_j, Z_j), j = 1, \ldots, m\}.
\]

MPLE for the overall treatment effect

\[
\hat{\theta}_{PL} = \arg \max_{\tilde{\theta}} \prod_{i=1}^{n+m} \left[ \frac{\exp(\tilde{\theta}' Z_i)}{\sum_{j \in \mathcal{R}_i} \exp(\tilde{\theta}' Z_j)} \right]^{\delta_i}.
\]

**Question:** is this the best estimate for the overall treatment effect?
Methods used for meta-analysis

- Inverse variance fixed-effect method

\[ \hat{\theta}_L = \left( \sum_{r=1}^{K} V_r^{-1} \right)^{-1} \sum_{r=1}^{K} V_r^{-1} \hat{\theta}_r. \]

+ Can be well approximated by \( \hat{\theta}_L \approx \sum_{r=1}^{K} p_r \hat{\theta}_r \)

**Note:** any linear combination of \( \hat{\theta}_r \) would lead to a biased estimate for \( \tilde{\theta} \) if \( \tilde{\theta}_r \)'s are essentially different.

- DerSimonian and Laird (1986) random effect model

\[ \theta_r = \theta + a_r + \epsilon_r, \quad r = 1, \ldots, K. \]

+ \( \theta \) true overall treatment effect
+ \( a_r \sim N(0, \tau^2) \) random effect
+ \( \epsilon_r \sim N(0, \sigma_r^2) \) random error

**Note:** need a large \( K \) to estimate the unknown parameters.
Some idea

The (true) pdf for \((W_i, \delta_i, \bm{Z}_i)\) is

\[
ph_0(t)e^{\tilde{\alpha}'\bm{Z}} - \exp(\tilde{\alpha}'\bm{Z})H_0(t) + (1 - p)h_0(t)e^{\tilde{\beta}'\bm{Z}} - \exp(\tilde{\beta}'\bm{Z})H_0(t)
\]

True hazard

\[
h(t|\bm{Z}) = h_0(t) \frac{pe^{\tilde{\alpha}'\bm{Z}} - \exp(\tilde{\alpha}'\bm{Z})H_0(t) + (1 - p)e^{\tilde{\beta}'\bm{Z}} - \exp(\tilde{\beta}'\bm{Z})H_0(t)}{pe^{-\exp(\tilde{\alpha}'\bm{Z})H_0(t)} + (1 - p)e^{-\exp(\tilde{\beta}'\bm{Z})H_0(t)}}
\]

Estimate \(\tilde{\theta}\) under a misspecified model

\[
h(t|\bm{Z}) = h_0(t)e^{\tilde{\theta}'\bm{Z}}.
\]

**Note:**

- \(\tilde{\theta}\) is not a uniquely defined parameter
- \(\tilde{\theta}\) is the limit of a justifiable estimator
- There can be various versions of the “overall treatment effect” \(\tilde{\theta}\)
- Efficient estimates follows the valid definition of \(\tilde{\theta}\), not vice versa
New Method (1)

- Combined hazard ratio as a limit of the MPLE from pooled data
Proposition 1. ($\theta_{PL}^*$ based on patient line data and MPLE)

Struthers and Kalbfleisch (1986) and Lin and Wei (1989)
- $(W_i, \delta_i, Z_i), i = 1, \ldots, n + m = N \sim$ misspecified $h(tZ) = h_0(t)e^{\tilde{\theta}'Z}$
- $h_i(t)$ true hazard function of the $i$th patient
- $Y_i(t) = 1_{W_i \geq t}$ at-risk process

Define

$$s^{(0)}(t) = E \left[ \sum_{i=1}^{N} Y_i(t)h_i(t) \right], \quad s^{(1)}(t) = E \left[ \sum_{i=1}^{N} Y_i(t)h_i(t)Z_i \right],$$

$$s^{(0)}(\tilde{\theta}, t) = E \left[ \sum_{i=1}^{N} Y_i(t)e^{\tilde{\theta}'Z_i} \right], \quad s^{(1)}(\tilde{\theta}, t) = E \left[ \sum_{i=1}^{N} Y_i(t)e^{\tilde{\theta}'Z_i}Z_i \right].$$

The MPLE $\hat{\theta}_{PL} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ calculated from $N$ patient records converges in probability to the solution of

$$\int_0^\infty s^{(1)}(t)dt - \int_0^\infty \frac{s^{(1)}(\tilde{\theta}, t)}{s^{(0)}(\tilde{\theta}, t)} s^{(0)}(t)dt = 0.$$

The asymptotic variance of $\hat{\theta}_{PL}$ can be consistently estimated by a “robust sandwich matrix estimator”.

Combining Survival Trials Using Aggregate Data, Based on Misspecified Models
Corollary 1

Assume no censoring in the data $X$ and $Y$. The true treatment effects (log hazard ratios) from either trial are known and denoted by $\tilde{\alpha}$ and $\tilde{\beta}$ respectively. The MPLE for the overall log hazard ratio $\hat{\theta}_{PL}$ converges in probability to a constant $\theta^*_{PL}$ when $n, m \to \infty$. Here $\theta^*_{PL}$ is the unique solution to the following equation about $\tilde{\theta}$:

$$
E(Z) = \int_0^\infty \frac{pE\left(e^{\tilde{\theta}'Z - \exp(\tilde{\alpha}'Z)u}Z\right) + (1 - p)E\left(e^{\tilde{\theta}'Z - \exp(\tilde{\beta}'Z)u}Z\right)}{pE\left(e^{\tilde{\theta}'Z - \exp(\tilde{\alpha}'Z)u}\right) + (1 - p)E\left(e^{\tilde{\theta}'Z - \exp(\tilde{\beta}'Z)u}\right)}
\left[ pE\left(e^{\tilde{\alpha}'Z - \exp(\tilde{\alpha}'Z)u}\right) + (1 - p)E\left(e^{\tilde{\beta}'Z - \exp(\tilde{\beta}'Z)u}\right) \right] du.
$$

$\hat{\theta}_{PL} \xrightarrow{p} \theta^*_{PL}$ is a good definition for the “overall treatment effect”

- Notations: $\hat{c}_{PL} = e^{\hat{\theta}_{PL}} \xrightarrow{p} e^{\theta^*_{PL}} = c^*_{PL}$
- Substitute $\tilde{\alpha}$ by $\hat{\alpha}$, $\tilde{\beta}$ by $\hat{\beta}$. Denote the solution by $\hat{\theta}_M$
- $\hat{\theta}_M$ is asymptotically efficient, semi-parametric, only requires aggregate statistics
- The variance of $\hat{\theta}_M$ can be simulated using parametric bootstrap
Example 1 ($\theta^*_M$)

- Univariate covariate - arm indicator $z \sim \text{Bernoulli}(q)$
- Notations: Trial 1 $a = e^\alpha$, Trial 2 $b = e^\beta$, Combined trial $c = e^\theta$

The MPLE $\hat{c}_{PL}$ converges to the solution of

$$1 = \int_0^\infty \frac{(1 - q)e^{-u} + pqae^{-au} + (1 - p)qbe^{-bu}}{(1 - q)e^{-u} + pqce^{-au} + (1 - p)qce^{-bu}} \cdot e^{-u} du.$$

**Remark:**

- In most cases, can prove that $c^*_{PL} < e^{p\tilde{\alpha}+(1-p)\tilde{\beta}}$
- $\hat{c}_M$ is the solution of

$$1 = \int_0^\infty \frac{(1 - q)e^{-u} + pq\hat{a}e^{-\hat{a}u} + (1 - p)q\hat{b}e^{-\hat{b}u}}{(1 - q)e^{-u} + pqce^{-\hat{a}u} + (1 - p)qce^{-\hat{b}u}} \cdot e^{-u} du.$$

$\hat{c}_M$ is an asymptotically efficient, semi-parametric estimate for $c^*_{PL}$ requiring only aggregate statistics.
Example 2 (Bias in the inverse-variance-weighted effects)

**Rule of thumb:** For most decent $a$, $b$ values (e.g., $a > 0.2$), $p\alpha + (1 - p)\beta$ overestimates the overall log hazard ratio. $\hat{\theta}_L = p\hat{\alpha} + (1 - p)\hat{\beta}$ is positively biased. Let $p = 0.5$, $q = 0.5$.

**Figure:** Contour plot for the percentage difference: $\frac{\exp(\theta^*_L) - c^*_PL}{c^*_PL} \times 100\%$. 
Corollary 1: Assume no censoring in the data $X$ and $Y$ ...

Q: Why "no censoring"?

- (Independent) censoring contains no information about the survival times
- $\hat{\theta}_{PL}$ is not robust against heterogeneity in the study population when there is censoring!
Corollary 2

Assume independent right censoring for both $X$ and $Y$ such that the survival functions of the censoring times are denoted $C_X(t|Z)$ and $C_Y(t|Z)$ respectively. The limit of $\hat{\theta}_{PL}$ with $n, m \to \infty$ is the unique solution to the following equation with respect to $\tilde{\theta}$ with known values of $\tilde{\alpha}$ and $\tilde{\beta}$:

$$
\int_0^\infty \left[ pE\left( f_X(t|Z)C_X(t|Z) \right) + (1 - p)f_Y(t|Z)C_Y(t|Z) \right] dt =
$$

$$
\int_0^\infty \frac{pE\left( e^{\tilde{\theta}'Z-\exp(\tilde{\alpha}'Z)H_0(t)} C_X(t|Z) \right) + (1 - p)E\left( e^{\tilde{\theta}'Z-\exp(\tilde{\beta}'Z)H_0(t)} C_Y(t|Z) \right)}{pE\left( e^{\tilde{\theta}'Z-\exp(\tilde{\alpha}'Z)H_0(t)} C_X(t|Z) \right) + (1 - p)E\left( e^{\tilde{\theta}'Z-\exp(\tilde{\beta}'Z)H_0(t)} C_Y(t|Z) \right)} \cdot 
$$

$$
[pE\left( f_X(t|Z)C_X(t|Z) \right) + (1 - p)E\left( f_Y(t|Z)C_Y(t|Z) \right)] dt.
$$

$\square$
Example 3 (Censoring at fixed time $T_{\text{max}}$)

- Univariate covariate - arm indicator $z \sim \text{Bernoulli}(q)$
- Both clinical trials will be terminated at given time $T_{\text{max}} > 0$, i.e. the censoring times follow

$$C_X(t|Z) = C_Y(t|Z) = 1 \text{ if } t < T_{\text{max}}, \ 0 \text{ otherwise}.$$ 

The MPLE $\hat{c}_{PL}(W_1, \ldots, W_N)$ converges to the solution (denoted $c^*$) of

$$1 - e^{-H_0(T_{\text{max}})} = \int_0^{H_0(T_{\text{max}})} \frac{(1 - q)e^{-u} + pqae^{-au} + (1 - p)qbe^{-bu}}{(1 - q)e^{-u} + pace^{-au} + (1 - p)qce^{-bu}} \cdot e^{-u} du.$$ 

Can prove $c^* > c_{PL}$ for arbitrary $T_{\text{max}} < +\infty$. That is, $\hat{\theta}_{PL}$ is always positively biased if the data are censored at $T_{\text{max}}$. 
The overall treatment effect $\theta^*_\text{PL}$ (or $c^*_\text{PL}$) can be defined using an integral equation

$$1 = \int_0^\infty \frac{(1 - q)e^{-u} + pqae^{-au} + (1 - p)qbe^{-bu}}{(1 - q)e^{-u} + pqce^{-au} + (1 - p)qce^{-bu}} \cdot e^{-u} du.$$ 

The MPLE based on patient line data $\hat{\theta}_\text{PL}$ is good only if there is no censoring.

The inverse-variance-weighted estimate $\hat{\theta}_L \approx p\hat{\alpha} + (1 - p)\hat{\beta}$ is almost surely biased when $\alpha \neq \beta$.

$\hat{\theta}_M$ is recommended instead:

- Asymptotically efficient
- Semi-parametric
- Only requires aggregate statistics $\hat{\alpha}$ and $\hat{\beta}$
- Robust with respect to censoring, because $\hat{\alpha}$ and $\hat{\beta}$ are unbiased despite of censoring
Example 4 (simulation for bias due to censoring)

1000 rounds of simulation
- Let the trial sizes $p : (1 - p) = 0.7 : 0.3$. Equal size randomization $q = 0.5$
- 200 treated patients $X_1, \ldots, X_{200} \sim \text{Exp}(0.3)$
- 200 control group patients $X_{201}, \ldots, X_{400} \sim \text{Exp}(1)$
- 85 treated patients $Y_1, \ldots, Y_{85} \sim \text{Exp}(0.8)$
- 85 control group patient $Y_{86}, \ldots, Y_{170} \sim \text{Exp}(1)$

When $T_{\text{max}} = \infty$

$$E(\hat{\theta}_{PL}) = -0.926, (-1.088, -0.756)$$
$$E(\hat{\theta}_M) = -0.928, (-1.092, -0.758)$$
Example 4 (simulation for bias due to censoring)

Figure: (i) Mean and the 95% CI for $\hat{\theta}_M$. (ii) Mean and the 95% CI for $\hat{\theta}_{PL}$.

When $T_{\text{max}} = 1$, observed 51% of censoring. $E(\hat{\theta}_{PL}) = -0.854$, (-1.074, -0.601), accounts for 7.8% of positive bias when compared to $E(\hat{\theta}_{PL}) = -0.926$, (-1.088, -0.756).
A real life example
Example 5 (A drug for cardiovascular disease)

- Primary endpoint: composite of all cause death, myocardial infarction, stroke or severe recurrent ischemia requiring revascularization
- Secondary endpoint: composite of all cause death, myocardial infarction or stroke

**Table:** Overall Treatment effect of primary and key secondary efficacy endpoints †

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Parameter</th>
<th>Pooled Placebo Event / N (%)</th>
<th>New treatment Event / N (%)</th>
<th>HR (95% CI)</th>
<th>Two-sided Wald p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Strata</td>
<td>Primary</td>
<td>83/1160 (7.2)</td>
<td>141/2331 (6.0)</td>
<td>0.84 (0.64, 1.1)</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>66/1160 (5.7)</td>
<td>101/2331 (4.3)</td>
<td>0.76 (0.55, 1.03)</td>
<td>0.077</td>
</tr>
<tr>
<td>Stratum 1</td>
<td>Primary</td>
<td>34/253 (13.4)</td>
<td>40/508 (7.9)</td>
<td>0.57 (0.36, 0.9)</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>29/253 (11.5)</td>
<td>35/508 (6.9)</td>
<td>0.59 (0.39, 0.96)</td>
<td>0.034</td>
</tr>
<tr>
<td>Stratum 2</td>
<td>Primary</td>
<td>49/907 (5.4)</td>
<td>101/1823 (5.5)</td>
<td>1.03 (0.73, 1.45)</td>
<td>0.853</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>37/907 (4.1)</td>
<td>66/1823 (3.6)</td>
<td>0.89 (0.59, 1.33)</td>
<td>0.568</td>
</tr>
</tbody>
</table>

† The data displayed here are quoted from FDA.gov http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/ucm381461.htm

**Note:**
- Heterogeneous strata (Primary HR, 0.75 vs 1.03. Secondary HR 0.59 vs 0.89)
- High censoring rate (86.5% ~ 96.4%)
### Table: New analyses on the overall treatment effect using various models †

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimator</th>
<th>Combined HR (95% CI)</th>
<th>Combined log HR (95% CI)</th>
<th>(%) difference vs $\hat{\theta}_M$ §</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>$\hat{c}_{PL}$ ‡</td>
<td>0.84 (0.64, 1.1)</td>
<td>-0.17 (-0.45, 0.10)</td>
<td>-42</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>$\hat{c}_M$</td>
<td>0.88 (0.68, 1.15)</td>
<td>-0.12 (-0.39, 0.14)</td>
<td></td>
<td>0.368</td>
</tr>
<tr>
<td></td>
<td>$\exp(\hat{\theta}_L)$</td>
<td>0.91 (0.68, 1.21)</td>
<td>-0.10 (-0.39, 0.19)</td>
<td>17</td>
<td>0.498</td>
</tr>
<tr>
<td>Secondary</td>
<td>$\hat{c}_{PL}$ ‡</td>
<td>0.76 (0.55, 1.03)</td>
<td>-0.27 (-0.60, 0.03)</td>
<td>-23</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>$\hat{c}_M$</td>
<td>0.81 (0.59, 1.10)</td>
<td>-0.22 (-0.53, 0.10)</td>
<td></td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td>$\exp(\hat{\theta}_L)$</td>
<td>0.81 (0.58, 1.14)</td>
<td>-0.21 (-0.55, 0.13)</td>
<td>5</td>
<td>0.233</td>
</tr>
</tbody>
</table>

† $p = 761/3491 = 0.218$, $q = 2/3$.
‡ Quoted from row 1, 2 of Table 1. Calculated from the pooled patient line data.
§ $(\hat{\theta} - \hat{\theta}_M)/\hat{\theta}_M \times 100$, where $\hat{\theta}$ varies by the fitting model.
Example 5 (power of the Wald tests)

Figure: Power of the two-sided Wald tests for the ATLAS ACS TIMI 46 trial


