Analysis of a Composite Endpoint with Missing Data in Components

Hui Quan, Daowen Zhang, Ji Zhang & Laure Devlamynck
Sanofi-Aventis

February 2007
Rutgers Biostatistics Day
Outline

• The use of composite endpoint
• The case of two components
• The general case
• Simulation Results
• Real data example
• Discussion
Composite Endpoint

Composite endpoints are often used in clinical trials

- GI outcomes study: perforation, ulcer and GI bleeding
- Major CV trials: MI, stroke and all-cause mortality
- VTE trials: CPMP PtC--DVT, PE and all-cause deaths

Classify a patient as having an event if the patient has events in any components
Composite Endpoint (2)

Advantages:

- To increase the overall event rate
- To reduce the size of the trial and achieve desired power
- To shorten the duration and get timely answers
- To avoid multiplicity issue

Conventional approach for analyzing composite endpoint:

- To directly analyze the composite endpoint
- A valid approach when no missing data in any components

As any study endpoints, components could have missing data.
Composite Endpoint (3)

A trial to compare two treatments on prevention of venous thromboembolism in patients with knee surgery

Primary endpoint– a composite endpoint
- DVT through venograph at Week 1 (30% missing)
- Symptomatic DVT and/or PE (pulmonary embolism)
- VTE related death

Two naïve approaches:
• To exclude patients with missing data in components
• To assume patients with missing data in a component to have no event in the component

Both provide inconsistent estimates.
Composite Endpoint (4)

The new approach will:

• Use all available data – consistent with ITT principle
• Derive rates for all potential study outcomes
• Then combine these rates to obtain the rate for the composite endpoint
• Use a likelihood-based approach
• Provide consistent estimate under the MAR assumption for the components
Two Components

Two components: X and Y (=1 for event, =0 otherwise)

When there are no missing data

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(\pi_{00})</td>
<td>(\pi_{01})</td>
</tr>
<tr>
<td>1</td>
<td>(\pi_{10})</td>
<td>(\pi_{11})</td>
</tr>
</tbody>
</table>

Composite endpoint: \(V\) (\(V=1\) if either \(X=1\) or \(Y=1\), \(V=0\) if \(X=0\) and \(Y=0\)).

Focus: \(\Pr(V=1)=1-\pi_{00}\)
Two Components (2)

When there are missing data

<table>
<thead>
<tr>
<th>Scenario</th>
<th>X</th>
<th>Y</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>7</td>
<td>.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>.</td>
<td>0</td>
<td>.</td>
</tr>
<tr>
<td>9</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Scenarios 6 and 8: partial missing data
Scenario 9: complete missing data
Two Components (3)

Naïve 1: directly analyze V after excluding missing data (Scenarios 6, 8 and 9)

- Is valid only when missing data on V are MAR and MLE is used.
- However, even when missing data on all components are MCAR, missing data on V may still not be MAR.
- Since some observed X=0 and Y=0 are ignored in analysis, the derived rate for the composite endpoint will not be consistent with the true rate.

Naïve 2: assume V=0 for Scenarios 6 and 8 => under estimates the true rate.
A new Approach for two Components

Let \((X_i, Y_i), u_i^{jk} = I[X_i = j, Y_i = k], i=1, \ldots, n\) be complete data from \(n\) patients. Then

\[ u_i = (u_i^{00}, u_i^{01}, u_i^{10}, u_i^{11}) \sim \text{Multinomial} \left(1; \pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}\right) \]

The Probability mass of the complete data

\[ P(u, \pi) = \prod \pi_{u_i}^{u_{i00}} \pi_{u_i}^{u_{i01}} \pi_{u_i}^{u_{i10}} \pi_{u_i}^{u_{i11}} \]
A new Approach for two Components (2)

In the case of missing data, define

\[ n_{jk} = \sum I[X_i = j, Y_i = k] \]

\[ n_{j.} = \sum I[X_i = j, Y_i = .] \]

\[ n_{.k} = \sum I[X_i = ., Y_i = k] \]

Likelihood

\[ L(\pi) = \pi_{00}^{n_{00}} \pi_{01}^{n_{01}} \pi_{10}^{n_{10}} \pi_{11}^{n_{11}} \]  \hspace{1cm} \text{(Completely observed data)}

\[ \times (\pi_{00} + \pi_{01})^{n_{0.}} (\pi_{10} + \pi_{11})^{n_{1.}} \]  \hspace{1cm} \text{(X observed and Y missing)}

\[ \times (\pi_{00} + \pi_{10})^{n_{0.}} (\pi_{01} + \pi_{11})^{n_{1.}} \]  \hspace{1cm} \text{(X missing and Y observed)}
A new Approach for two Components (3)

- Direct iterative approach (Newton-Raphson) for MLE
- No closed form solution in general
- Fisher information for asymptotic variance $\sigma^2$ of $1 - \hat{\pi}_{00}$

Alternatively, EM algorithm for MLE is pretty straightforward
A new Approach for two Components (4)

\[ Q(\pi; \pi^{(r)}) = E[\log P(u; \pi) \mid \text{observed data}, \pi^{(r)}] \]

\[ = \hat{n}_{00}^{(r)} \log \pi_{00} + \hat{n}_{01}^{(r)} \log \pi_{01} + \hat{n}_{10}^{(r)} \log \pi_{10} + \hat{n}_{11}^{(r)} \log \pi_{11}, \]

where

\[ \hat{n}_{00}^{(r)} = n_{00} + n_{01}(1 - \delta_{01}^{(r)}) + n_{10}(1 - \delta_{10}^{(r)}), \]

\[ \hat{n}_{01}^{(r)} = n_{01} + n_{01}\delta_{01}^{(r)} + n_{11}(1 - \delta_{11}^{(r)}), \]

\[ \hat{n}_{10}^{(r)} = n_{10} + n_{10}\delta_{10}^{(r)} + n_{00}\delta_{00}^{(r)}, \]

\[ \hat{n}_{11}^{(r)} = n_{11} + n_{11}\delta_{11}^{(r)} + n_{01}\delta_{01}^{(r)} \]

and

\[ \delta_{j1}^{(r)} = \Pr(Y = 1 \mid X = j, \pi^{(r)}) = \pi_{j1}^{(r)} / (\pi_{j0}^{(r)} + \pi_{j1}^{(r)}), \]

\[ \delta_{1k}^{(r)} = \Pr(X = 1 \mid Y = k, \pi^{(r)}) = \pi_{1k}^{(r)} / (\pi_{0k}^{(r)} + \pi_{1k}^{(r)}), \quad j, k = 0, 1. \]

Maximizing \( Q(\pi; \pi^{(r)}) \) subject to \( \sum \pi_{jk} = 1 \) leads to the update:

\[ \pi_{jk}^{(r+1)} = \frac{\hat{n}_{jk}^{(r)}}{n}. \]
Two Components – one with no missing data

Let X be component with no missing data and Y be the other component.

\[ \Pr(X=1 \text{ or } Y=1) = \Pr(X=1) + \Pr(X=0 \text{ and } Y=1) \]
\[ = \Pr(X=1) + \Pr(Y=1|X=0)\Pr(X=0) \]

\[ n_1 = \text{ the \# of patients}, \quad m_1 = \text{ the \# of } (X=1) \]
\[ n_2 = \text{ the \# of } (X=0, \ Y \neq .), \quad m_2 = \text{ the \# of } (X=0, \ Y=1) \]

The MLE for the overall event rate for the composite endpoint:

\[ \hat{r} = \hat{p}_1 + \hat{p}_2(1 - \hat{p}_1) = \frac{m_1}{n_1} + \frac{m_2}{n_2} \frac{n_1 - m_1}{n_1} \]
General Case

For the case of $s$ components

- Calculate the probability $u_h$ of the $h$th outcomes based on $\pi$.
- Count then number $n_h$ of patients with the outcome.
- Use $(u_h)^{n_h}$ as a factor in the likelihood.

When $s=4$, there will be $2^4-1=15$ independent $\pi$’s and $3^4-1=80$ different factors in the likelihood --- too difficult to handle.

Components with non-missing data can be combined to reduce the number of components. More details are given for the case of 3 components.
Three Components

Three binary components: X, Y and Z
• =1 for event and =0 for no event.
• Due to missing data, three outcomes for each component
• A total of 27=3x3x3 possible outcomes \((3^k\) for k comps)

Composite endpoint \(V=1\) if \(X=1\) or \(Y=1\) or \(Z=1\),
\[ \begin{align*}
V &= 0 \text{ if } X=0 \text{ and } Y=0 \text{ and } Z=0, \\
V &= . \text{ otherwise.}
\end{align*} \]

As for the case of two components, there will be a problem if directly analyze \(V\).
Three Components (2)

Define

\[ \pi_{jkl} = \Pr(X = j, Y = k, Z = l) \]
\[ \pi_{jk.} = \Pr(X = j, Y = k, Z = \cdot) = \Pr(X = j, Y = k) = \pi_{jk0} + \pi_{jk1} \]
\[ \pi_{j.l} = \Pr(X = j, Y = \cdot, Z = l) = \Pr(X = j, Z = l) = \pi_{j0l} + \pi_{j1l} \]
\[ \pi_{.kl} = \Pr(X = \cdot, Y = k, Z = l) = \Pr(Y = k, Z = l) = \pi_{0kl} + \pi_{1kl} \]
\[ \pi_{j..} = \Pr(X = j) \quad \pi_{.k.} = \Pr(Y = k) \quad \pi_{..l} = \Pr(Z = l) \]

Similarly, the number of patients with each outcome.
Likelihood for \( \pi = (\pi_{000}, \pi_{001}, \pi_{010}, \ldots, \pi_{111}) \):

\[ G(\pi) = \prod_{jkl}^{n} \pi_{jkl} \prod_{jk.}^{n} \pi_{jk.} \ldots \prod_{.k.}^{n} \pi_{.k.} \prod_{..l}^{n} \pi_{..l} \]
Three Components (3)

In generally, there is no closed form solution for MLE. Newton-Raphson or EM algorithm can be used to obtain

\[ \hat{r} = 1 - \hat{\pi}_{000} \]

and the corresponding asymptotic variance.

When missing data pattern is monotonic

A closed form solution for MLE exits.
Between-Treatment Comparison

\( r_1 \) and \( r_2 \) the true rates for Treat1 & Treat2

\( \hat{r}_1 \) and \( \hat{r}_2 \) the estimates

\( \frac{\sigma_1^2}{n} \) and \( \frac{\sigma_2^2}{n} \) the variances.

Asymptotically,

\[ \log(\frac{\hat{r}_1}{\hat{r}_2}) \sim N(\log(\frac{r_1}{r_2}), \frac{\sigma_1^2}{r_1^2} + \frac{\sigma_2^2}{r_2^2})/n \]

Under null:

\[ T = \sqrt{n} \log(\frac{\hat{r}_1}{\hat{r}_2})/\sqrt{\frac{\hat{\sigma}_1^2}{\hat{r}_1^2} + \frac{\hat{\sigma}_2^2}{\hat{r}_2^2}} \sim N(0,1) \]
Simulation – Two components

Assume:

\[ p_x = \Pr(X = 1) \quad p_{y0} = \Pr(Y = 1 \mid X = 0) \]

The true overall event rate for composite endpoint:

\[ r = p_x + p_{y0} (1 - p_x) \]

In addition, assume a MCAR for X and MAR for Y,

\[ p_{mx} = \Pr(X = . \mid X = j, Y = k) \quad \text{(independent with j & k)} \]
\[ p_{mY0} = \Pr(Y = . \mid X = 0, Y = k) \quad \text{(independent with k)} \]
\[ p_{mY1} = \Pr(Y = . \mid X = 1, Y = k) \quad \text{(independent with k)} \]

\[ \Pr(X = ., Y = .) = 0 \]
## Simulation Results for 2 components

<table>
<thead>
<tr>
<th>Case</th>
<th>$r$</th>
<th>$\hat{r}_{\text{mle}}$</th>
<th>$\hat{r}_{\text{na1}}$</th>
<th>$\hat{r}_{\text{na2}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.79</td>
<td>9.80</td>
<td>11.34</td>
<td>8.31</td>
</tr>
<tr>
<td>2</td>
<td>9.79</td>
<td>9.77</td>
<td>13.08</td>
<td>7.60</td>
</tr>
<tr>
<td>3</td>
<td>14.50</td>
<td>14.49</td>
<td>16.72</td>
<td>12.36</td>
</tr>
<tr>
<td>4</td>
<td>14.50</td>
<td>14.60</td>
<td>19.32</td>
<td>11.44</td>
</tr>
<tr>
<td>5</td>
<td>23.50</td>
<td>23.61</td>
<td>26.76</td>
<td>19.53</td>
</tr>
<tr>
<td>6</td>
<td>23.50</td>
<td>23.31</td>
<td>28.14</td>
<td>18.89</td>
</tr>
<tr>
<td>7</td>
<td>28.00</td>
<td>28.04</td>
<td>31.01</td>
<td>22.60</td>
</tr>
<tr>
<td>8</td>
<td>28.00</td>
<td>27.97</td>
<td>33.01</td>
<td>22.36</td>
</tr>
</tbody>
</table>

Same $r$ different missing data rates
Simulation: Between-Treatment Comparison

<table>
<thead>
<tr>
<th>Case</th>
<th>$\lambda$</th>
<th>$\hat{\lambda}_{mle}$</th>
<th>$\hat{\lambda}_{na1}$</th>
<th>$\hat{\lambda}_{na2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>1.01</td>
<td>1.00</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>1.01</td>
<td>0.84</td>
<td>1.03</td>
</tr>
<tr>
<td>3</td>
<td>0.42</td>
<td>0.42</td>
<td>0.41</td>
<td>0.38</td>
</tr>
<tr>
<td>4</td>
<td>0.42</td>
<td>0.42</td>
<td>0.36</td>
<td>0.42</td>
</tr>
<tr>
<td>5</td>
<td>1.19</td>
<td>1.19</td>
<td>1.15</td>
<td>1.11</td>
</tr>
<tr>
<td>6</td>
<td>1.19</td>
<td>1.20</td>
<td>1.03</td>
<td>1.22</td>
</tr>
</tbody>
</table>

Same $\lambda$ different missing data rates
Example


- Treatment Duration: 5 – 9 postoperative days
- Venograph: Day 5 to Day 11
- X: fatal or non-fatal PE – non-missing
- Y: DVT through venograph – 30% missing
- Primary endpoint: composite endpoint of X and Y.
<table>
<thead>
<tr>
<th></th>
<th>Fond</th>
<th>Enox</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>517</td>
<td>517</td>
</tr>
<tr>
<td>$n_{00}$</td>
<td>316</td>
<td>262</td>
</tr>
<tr>
<td>$n_{01}$</td>
<td>44</td>
<td>97</td>
</tr>
<tr>
<td>$n_{10}$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$n_{11}$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$n_0.$</td>
<td>156</td>
<td>154</td>
</tr>
<tr>
<td>$n_1.$</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>$\hat{r}_{\text{MLE}}$ (%)</td>
<td>12.39</td>
<td>27.58</td>
</tr>
<tr>
<td>$\hat{r}_{\text{NA1}}$ (%)</td>
<td>12.47</td>
<td>27.82</td>
</tr>
<tr>
<td>$\hat{r}_{\text{NA2}}$ (%)</td>
<td>8.70</td>
<td>19.54</td>
</tr>
<tr>
<td>MLE $\hat{\lambda}_{\text{MLE}}$ (95% CI)</td>
<td>0.45 (0.33, 0.62)</td>
<td></td>
</tr>
<tr>
<td>NA1 $\hat{\lambda}_{\text{NA1}}$ (95% CI)</td>
<td>0.45 (0.30, 0.64)</td>
<td></td>
</tr>
</tbody>
</table>
Example (3)

\[ \hat{r}_{na1} = \frac{(n_{01} + n_{10} + n_{1.} + n_{11})}{(n - n_{0.})} \]

\[ \hat{r}_{na2} = \frac{(n_{01} + n_{10} + n_{1.} + n_{11})}{n} \]

\[ \hat{r}_{MLE} = \frac{(n_{10} + n_{1.} + n_{11})}{n} + \frac{n_{01}}{(n_{00} + n_{01})(n_{00} + n_{01} + n_{0.})} \]

\[ \hat{r}_{na1} \geq \hat{r}_{MLE} \geq \hat{r}_{na2} \]
Individual Components

After the demonstration of treatment effect on the composite endpoint, there may be the need to assess treatment effects on individual components. For the case of two components:

- $\pi_{10} + \pi_{11}$ for X component
- $\pi_{01} + \pi_{11}$ for Y component
- Joint model for all components not individual model for individual components even under MAR
- No need for multiplicity
Individual Components (2)

For the case of three components:

\[
\{X, Y, Z\} \\
\downarrow \\
\{X, Y\} \text{ and } \{X, Z\} \\
\downarrow \\
\{X\}
\]
Discussion

• Theoretically and based on simulation, the conventional approaches of directly analyzing the composite endpoint provide inconsistent estimates.

• The new method provides consistent and more efficient estimate under MAR for components.

• The new method uses all available data – consistent with the ITT principle.

• Incorporation of covariates will be in future research.
References


