A Broad Framework for Joint Modeling
And Some Tales From the Unexpected

Geert Molenberghs

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)
Universiteit Hasselt & Katholieke Universiteit Leuven, Belgium
geert.molenberghs@uhasselt.be & geert.molenberghs@med.kuleuven.be
www.ibiostat.be

with: M. Aerts, M. Davidian, M.G. Kenward, D. Rizopoulos, A.A. Tsiatis, G. Verbeke
## Common Ground for Various Settings

<table>
<thead>
<tr>
<th>Setting</th>
<th>Relationship(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential trials</td>
<td>outcome &amp; sample size</td>
</tr>
<tr>
<td>Incomplete data</td>
<td>outcome(s) &amp; missigness process</td>
</tr>
<tr>
<td>Completely random sample size</td>
<td>outcome &amp; sample size</td>
</tr>
<tr>
<td>‘Informative’ cluster size</td>
<td>outcome &amp; cluster size</td>
</tr>
<tr>
<td>Classical survival</td>
<td>time to event &amp; censorship</td>
</tr>
<tr>
<td>(Narrow) joint modeling</td>
<td>time to event(s) &amp; longitudinal process</td>
</tr>
<tr>
<td>Random observation times</td>
<td>outcome &amp; measurement schedule</td>
</tr>
</tbody>
</table>
A Classic: Joint Models with Missing Data

\[ f(y_i, r_i | X_i, \theta, \psi) \]

Selection Models:
\[ f(y_i | X_i, \theta) f(r_i | X_i, y_i^o, y_i^m, \psi) \]

MCAR $\rightarrow$ MAR $\rightarrow$ MNAR

\[ f(r_i | X_i, \psi) f(r_i | X_i, y_i^o, \psi) f(r_i | X_i, y_i^o, y_i^m, \psi) \]

Pattern-mixture Models:
\[ f(y_i | X_i, r_i, \theta) f(r_i | X_i, \psi) \]

Shared-parameter Models:
\[ f(y_i | X_i, b_i, \theta) f(r_i | X_i, b_i, \psi) \]
## Generic Setting

\[ f(y_i, c_i | \theta, \psi) = f(y_i | \theta) \cdot f(c_i | y_i, \psi) = f(y_i | c_i, \theta, \psi) \cdot f(c_i | \theta, \psi) \]

<table>
<thead>
<tr>
<th>Setting</th>
<th>( Y_i )</th>
<th>( C_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequential trials</strong></td>
<td>( Y_i )</td>
<td>( N )</td>
</tr>
<tr>
<td><strong>Incomplete data</strong></td>
<td>( Y_i )</td>
<td>( R_i )</td>
</tr>
<tr>
<td><strong>Completely random sample size</strong></td>
<td>( Y_i )</td>
<td>( N )</td>
</tr>
<tr>
<td><strong>‘Informative’ cluster size</strong></td>
<td>( Y_i )</td>
<td>( T_i )</td>
</tr>
<tr>
<td><strong>Classical survival</strong></td>
<td>( T_i )</td>
<td>( C_i )</td>
</tr>
<tr>
<td><strong>(Narrow) joint modeling</strong></td>
<td>( Y_i )</td>
<td>( T_i, ... )</td>
</tr>
<tr>
<td><strong>Random observation times</strong></td>
<td>( Y_i )</td>
<td>( T_i )</td>
</tr>
</tbody>
</table>
Fundamental Concepts

**Ignorability:**

\[ f(c_i | y_i, \psi) = f(c_i | y_i^0, \psi) \]
combined with **separability**

**Ancillarity:**

- \( S \) minimally sufficient statistic for \( \theta \)
- \( T|S \) does not contain information about \( \theta \)
Completeness:

- $g(\cdot)$ a measurable function of $S$
- $E[g(s)] = 0$ for all $\theta \implies g(s) = 0$ a.e.

Weights:

$$f(y_i|\theta^*, \psi^*) = f(y_i|c_i, \theta^*) \cdot \frac{f(c_i|\psi^*)}{f(c_i|y_i, \theta^*, \psi^*)}$$

$$= f(y_i|c_i, \theta^*) \cdot w_i(y_i, c_i, \theta^*, \psi^*)$$

MAR: $w_i(y_i, c_i) = w_i(y_i^0, c_i)$

Deterministic rule:

$$w_i(y_i^0, c_i) \in \{0, 1\}$$
Simple Setting

- Observe
  \[ Y_i \sim N(\mu, 1) \quad i = 1, \ldots, n \]

- Construct sum statistics
  \[ K = \sum_{i=1}^{n} Y_i \]

- Apply a stopping rule, e.g.,
  \[ F \left( \alpha + \frac{\beta}{n} k \right) \]

- If continuation applies, observe \( Y_i \), \( i = n + 1, \ldots, 2n \)

- Data:
  \( (Y_1, \ldots, Y_N, N) \)
Stopping Rule

\[ F \left( \alpha + \frac{\beta}{n} \right) = \Phi \left( \alpha + \frac{\beta}{n} \right) \]

- Three important cases:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Purely random sample size</td>
<td>( \beta = 0 )</td>
</tr>
<tr>
<td>Probabilistic stopping</td>
<td>( \beta \neq 0 )</td>
</tr>
<tr>
<td>Purely random sample size</td>
<td>( \beta \to +\infty )</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Y</strong></td>
<td>( \prod_{i=1}^{N} \phi(y_i; \mu) )</td>
</tr>
<tr>
<td>( N</td>
<td>Y )</td>
</tr>
<tr>
<td>**Y</td>
<td>N**</td>
</tr>
<tr>
<td>( N )</td>
<td>( \Phi \left( \frac{\alpha + \beta \mu}{\sqrt{1 + \beta^2/n}} \right) )</td>
</tr>
</tbody>
</table>
Joint Likelihood

\[ L(\mu) = \prod_{i=1}^{N} \phi(y_i; \mu) \cdot \Phi \left( \alpha + \frac{\beta}{n} \right)^{z} \cdot \left\{ 1 - \Phi \left( \alpha + \frac{\beta}{n} \right) \right\}^{1-z} \]

\[ \ell(\mu) = -\frac{1}{2} \sum_{i=1}^{N} (y_i - \mu)^2 \]

\[ S(\mu) = \sum_{i=1}^{N} (y_i - \mu) \]

\[ H(\mu) = -N \]
Conditional Likelihood

\[ L_c(\mu) = \prod_{i=1}^{N} \phi(y_i; \mu) \Phi\left( \alpha + \frac{\beta k}{n} \right)^z \left[ 1 - \Phi\left( \alpha + \frac{\beta k}{n} \right) \right]^{1-z} \]

\[ \Phi\left( \frac{\alpha + \beta \mu}{\sqrt{1+\beta^2/n}} \right)^z \left[ 1 - \Phi\left( \frac{\alpha + \beta \mu}{\sqrt{1+\beta^2/n}} \right) \right]^{1-z} \]

\[ N = n : \]

\[ \ell(\mu) = -\frac{1}{2} \sum_{i=1}^{N} (y_i - \mu)^2 - \ln \Phi(\nu) \]

\[ S(\mu) = \sum_{i=1}^{N} (y_i - \mu) - \bar{\beta} \cdot \frac{\phi(\nu)}{\Phi(\nu)} \]

\[ H(\mu) = -N + \bar{\beta}^2 \cdot \left[ \nu \cdot \Phi(\nu) + \phi(\nu) \right] \cdot \frac{\phi(\nu)}{\Phi(\nu)^2} \]

\[ N = 2n : \]

\[ \ell(\mu) = -\frac{1}{2} \sum_{i=1}^{N} (y_i - \mu)^2 - \ln \left[ 1 - \Phi(\nu) \right] \]

\[ S(\mu) = \sum_{i=1}^{N} (y_i - \mu) + \bar{\beta} \cdot \frac{\phi(\nu)}{1 - \Phi(\nu)} \]

\[ H(\mu) = -N - \bar{\beta}^2 \cdot \left\{ \nu \cdot [1 - \Phi(\nu)] - \phi(\nu) \right\} \cdot \frac{\phi(\nu)}{[1 - \Phi(\nu)]^2} \]
Information, Bias, MSE

- Notation:

\[
\tilde{\alpha} = \frac{\alpha}{\sqrt{1 + \beta^2/n}} \\
\tilde{\beta} = \frac{\beta}{\sqrt{1 + \beta^2/n}} \\
\nu = \frac{(\alpha + \beta\mu)}{\sqrt{1 + \beta^2/n}}
\]

- Information:

\[
I(\mu) = n[2 - \Phi(\tilde{\alpha} + \tilde{\beta}\mu)]
\]

\[
I_c(\mu) = n[2 - \Phi(\tilde{\alpha} + \tilde{\beta}\mu)] - \frac{\tilde{\beta}^2\phi(\tilde{\alpha} + \tilde{\beta}\mu)^2}{\Phi(\tilde{\alpha} + \tilde{\beta}\mu)[1 - \Phi(\tilde{\alpha} + \tilde{\beta}\mu)]}
\]
Counterintuitive? $N$ is not ancillary

Some limits:

$$\beta = 0 \implies I(\mu) = I_c(\mu)$$

$$n \to \infty \implies I(\mu) = I_c(\mu)$$

Both estimators asymptotically unbiased

The conditional score $E[S_c(\mu)] = 0$:

conditional likelihood estimator unbiased in finite samples
• Mean squared error

\[ \beta < +\infty: \]

\[
\text{MSE}(\hat{\mu}) = \frac{1}{n[2 - \Phi(\nu)]} + \frac{1}{4n^2} \beta^2 \phi(\nu)^2
\]

\[
\text{MSE}(\hat{\mu}_c) \approx \frac{1}{n[2 - \Phi(\nu)]} + \frac{1}{[2 - \Phi(\nu)]^2 \Phi(\nu)[1 - \Phi(\nu)]n^2} \beta^2 \phi(\nu)^2
\]

\[ \beta \rightarrow +\infty: \]

\[
\text{MSE}(\hat{\mu}) = \frac{1}{n[2 - \Phi(\sqrt{n}\mu)]} + \frac{1}{4n} \phi(\sqrt{n}\mu)^2
\]

\[
\text{MSE}(\hat{\mu}_c) \approx \frac{1}{n[2 - \Phi(\sqrt{n}\mu)]} + \frac{1}{[2 - \Phi(\sqrt{n}\mu)]^2 \Phi(\sqrt{n}\mu)[1 - \Phi(\sqrt{n}\mu)]n} \phi(\sqrt{n}\mu)^2
\]
Likelihood Estimators: Conditional Bias

- Conditional likelihood estimator: conditionally unbiased

- Joint likelihood estimator ($\beta < +\infty$):

$$E(\mu|N = n) = \mu + \frac{\beta}{\sqrt{1 + \beta^2/n}} \cdot \frac{\phi(\nu)}{n\Phi(\nu)}$$

$$E(\mu|N = 2n) = \mu - \frac{\beta}{\sqrt{1 + \beta^2/n}} \cdot \frac{\phi(\nu)}{2n[1 - \Phi(\nu)]}$$

- Joint likelihood estimator ($\beta \to +\infty$):

$$E(\mu|N = n) = \mu + \frac{\phi(\sqrt{n}\mu)}{\sqrt{n}\Phi(\sqrt{n}\mu)}$$

$$E(\mu|N = 2n) = \mu - \frac{\phi(\sqrt{n}\mu)}{2\sqrt{n}[1 - \Phi(\sqrt{n}\mu)]}$$
• $\beta \to +\infty$ & $n \to +\infty$:

▷ $\mu < 0$

\[ E(\mu | N = n) \to 0 \]
\[ E(\mu | N = 2n) \to \mu \]

▷ $\mu > 0$

\[ E(\mu | N = n) \to \mu \]
\[ E(\mu | N = 2n) \to \frac{\mu}{2} \]

▷ **Yet: asymptotically unbiased**
Lack of Incompleteness $\rightarrow$ Counterintuitive Results

- $N$ fixed $\implies K$ complete

- Nice results apply

- **Lehmann-Scheffé theorem:**
  
  \[
  \begin{align*}
  \text{unbiased} & \quad \implies \quad \text{best mean-unbiased} \\
  \text{complete} & \quad \text{and} \\
  \text{sufficient} & 
  \end{align*}
  \]
• **Basu’s theorem:**

  \[
  \text{complete } \quad \text{sufficient } \quad \implies \quad \text{independent of any ancillary statistic}
  \]

• **Not** so in our case!

• Sufficient statistic is \((N, K)\)
• Joint distribution:

\[
p_\mu(N, k) = p_0(N, k) \cdot \exp \left( k\mu - \frac{1}{2}n\mu^2 \right)
\]

\[
p_0(n, k) = \phi_n(k) \cdot \Phi \left( \alpha + \frac{\beta}{n}k \right)
\]

\[
p_0(2n, k) = \phi_{2n}(k) \cdot \left[ 1 - \Phi \left( \frac{\alpha + \frac{\beta k}{2n}}{\sqrt{\frac{2n+\beta^2}{2n}}} \right) \right]
\]

• To establish incompleteness, we must find a function \( g(k, N) \) satisfying:

\[
g(k, 2n) \cdot p_0(2n, k) = -\int \phi_n(k - z) \cdot g(z, n) \cdot \phi_n(z) \cdot \Phi \left( \alpha + \frac{\beta}{n}z \right) \, dz
\]
Example:

\[ g(k, n) = \frac{\ell}{\Phi \left( \alpha + \frac{\beta}{n}k \right)} \]

\[ g(k, 2n) = -\frac{\ell}{1 - \Phi \left( \frac{\alpha + \frac{\beta k}{2n}}{\sqrt{\frac{2n + \beta^2}{2n}}} \right)} \]
Ordinary Sample Average is Biased and Not Optimal

- **Generalized sample average:**

\[
\bar{\mu} = \frac{K}{N} \cdot \left[ c \cdot I(N = n) + d \cdot I(N = 2n) \right]
\]

- **Expectation:**

\[
E(\bar{\mu}) = d\mu + (c - d)\mu \Phi(\nu) + \frac{2c - d}{2n} \frac{\beta}{\sqrt{1 + \beta^2/n}} \cdot \phi(\nu)
\]
Generalized Sample Size

\[ \beta = 0 \quad \text{— completely random sample size} \]

- A family of unbiased estimators:
  \[ d = \frac{1 - c\Phi}{1 - \Phi} \quad \implies E(\mu) = \mu \]

- Minimum variance:
  \[ c_{\text{opt}} = 1 - \frac{1}{2n} \cdot \frac{1 - \Phi}{\mu^2 + \frac{2 - \Phi}{2n}}, \quad d_{\text{opt}} = 1 + \frac{1}{2n} \cdot \frac{\Phi}{\mu^2 + \frac{2 - \Phi}{2n}}. \]

- No uniform optimum

- Ordinary sample average is not optimal
Generalized Sample Size

$\beta \neq 0$ — stopping rule

- All generalized sample averages biased

- This includes ordinary sample average!

- Asymptotically unbiased (as we known from likelihood)
Generalization 1: Arbitrary Number of Looks

• Our stopping rule

\[ \Phi \left( \alpha + \frac{\beta}{n} \right) \]

for \( \beta \to +\infty \) can be generalized to arbitrary numbers of looks

\[ n_1 < n_2 < \ldots < n_L \]

by applying

\[ \Phi \left( \alpha_j + \frac{\beta_j}{n_j} \right) \]

• Results carry over
Generalization 2: Completely Random Sample Size

- Allow for sample sizes with probabilities

<table>
<thead>
<tr>
<th>$n$</th>
<th>0</th>
<th>1</th>
<th>$\cdots$</th>
<th>$m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_n$</td>
<td>$\pi_0$</td>
<td>$\pi_1$</td>
<td>$\cdots$</td>
<td>$\pi_m$</td>
</tr>
</tbody>
</table>

- Incomplete sufficient statistic:

$$g(z, n) = b_n$$

$$\sum_{n=0}^{m} \pi_n \cdot b_n = 0$$
• Generalized sample average:

\[ \bar{\mu} = \sum_{n=0}^{m} \frac{K}{n} \cdot a_n \cdot I(N = n) \]

• **Unbiased** generalized sample average:

\[ a_n = 1 + b_n \]

• Optimal estimator:

\[ a_n = \frac{1}{\sum_{k=0}^{m} \frac{\mu^2 + \sigma^2/k}{\pi_k}} \]
Generalization 3: Missing Data

- Our stopping rule

\[ \Phi \left( \alpha + \frac{\beta}{n} \right) \]

for \( 0 < \beta < +\infty \) corresponds to MAR missingness in independent data

- Can be generalized to longitudinal sequences

Joint likelihood \( \equiv \) selection model representation

Conditional likelihood \( \equiv \) pattern-mixture model
• Assume, for example

\[ Y_i = (Y_{i1}, Y_{i2}) \]

▷ first block always observed
▷ second one possibly missing

• Sample average is **not** an option:

\[
\hat{\mu}_1 = \frac{1}{N} \sum_{i=1}^{N} y_{i1} \\
\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} y_{i2} - \Sigma_{21} \Sigma_{11}^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} y_{i1} - \frac{1}{N} \sum_{i=1}^{N} y_{i1} \right)
\]
Generalization 4: Informative Cluster Sizes

- **Either:** Model for two random variables:
  - Outcomes $Y_i$
  - Cluster size $t_i$

- **Or:** model for three random variables:
  - Outcomes $Y_i$
  - Observed cluster size $t_i$
  - Theoretical cluster size $n_i \geq t_i$

- Entirely similar to missing-data setting
Generalization 5: Joint Model for Longitudinal and Survival Data

- Model for two, three, or four random variables:
  - Longitudinal process $Y_i$
  - Time-to-event outcome $T_i$
  - Censoring time $C_i$
  - Missing data process $R_i$

- Formulate an appropriate “shared-parameter model”

- Results carry over
Conclusions

- Commonality across seemingly unrelated settings

- Ordinary sample average loses its ‘perfect status’ ← lack of completeness
  - Already with completely random sample size
  - More so with deterministic or probabilistic stopping rule

- It nevertheless is a valid estimator:
  - Asymptotically unbiased (unbiased for CRSS)
  - Almost optimal
  - Asymptotically optimal
  - Consistent with the likelihood