Informative’ observation times in longitudinal studies

Benjamin French, PhD
Department of Biostatistics and Epidemiology
University of Pennsylvania
bcfrench@upenn.edu

ENAR 2013
Orlando, Florida
13 March 2013
Collaborators

Kay See Tan, MS  
University of Pennsylvania

Andrea Troxel, ScD  
University of Pennsylvania

Stephen Kimmel, MD, MSCE  
University of Pennsylvania

Kevin Volpp, MD, PhD  
University of Pennsylvania
• Anticoagulant used for prevention and treatment of thromboembolism
  ▶ Formation of a clot that obstructs blood flow in vein or artery
  ▶ Associated outcomes: Deep vein thrombosis, stroke

• Highly efficacious, but has a narrow therapeutic range
  ▶ Over-anticoagulation: Increased risk of bleeding complications
  ▶ Under-anticoagulation: Increased risk of thromboembolic events

• Requires frequent monitoring, which may lead to dose changes
  ▶ INR: International normalized ratio
  ▶ Ratio of patient’s prothrombin time to that of a normal sample

• Poor adherence contributes to poor anticoagulation control
  ▶ Missed doses → under-anticoagulation (Kimmel et al., 2007)
  ▶ Goal: Improve adherence to warfarin therapy (Kimmel et al., 2012)
INR monitoring

Days since enrollment

- Below range
- Within range
- Above range

Informative observation times
Frequency and/or timing of data collection may depend on past outcome values, i.e. ‘informative’ observation times

- Selection bias: Participants with particular outcome values may have more observations; over-represented in the sample
- May result in biased estimates for exposure-outcome associations
Warfarin trial

- Participants randomized to one of four interventions
- Followed for a maximum of six months
- Primary outcome: Out-of-range INR (binary)

<table>
<thead>
<tr>
<th>Group</th>
<th>$P_0$</th>
<th>$P_{25}$</th>
<th>$P_{50}$</th>
<th>$P_{75}$</th>
<th>$P_{100}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>14</td>
</tr>
</tbody>
</table>
Warfarin trial: Analysis plan

- Logistic regression model for odds of out-of-range INR ($Y_{ij} = 1$)
  \[
  \text{logit } P[Y_{ij} = 1] = \beta_0 + \beta_1[B] + \beta_2[C] + \beta_3[D]
  \]

- GEE with working independence (Liang and Zeger, 1986)
  - Adjusted for stratification factors
  - Included a linear term for study time

- How to account for different number of INRs for each participant, which may be ‘informative’ of outcome values?
  - Use only ‘monthly’ INRs, closest to scheduled follow-up visits
  - Use all available INRs in an ‘unweighted’ model
  - Use all available INRs in a ‘weighted’ model, in which participant-level weights are equal to the inverse of the number of INRs, i.e. $1/N_i$; cluster-weighted GEE (Williamson et al., 2003)
Warfarin trial: Analysis plan

- Logistic regression model for odds of out-of-range INR ($Y_{ij} = 1$)

$$\text{logit } P[Y_{ij} = 1] = \beta_0 + \beta_1[B] + \beta_2[C] + \beta_3[D]$$

- GEE with working independence (Liang and Zeger, 1986)
  - Adjusted for stratification factors
  - Included a linear term for study time

- How to account for different number of INRs for each participant, which may be ‘informative’ of outcome values?
  - Use only ‘monthly’ INRs, closest to scheduled follow-up visits
  - Use all available INRs in an ‘unweighted’ model
  - Use all available INRs in a ‘weighted’ model, in which participant-level weights are equal to the inverse of the number of INRs, i.e. $1/N_i$; cluster-weighted GEE (Williamson et al., 2003)
### Warfarin trial: Results

<table>
<thead>
<tr>
<th>Group</th>
<th>Monthly</th>
<th>Unweighted</th>
<th>Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1.14</td>
<td>0.95</td>
<td>0.98</td>
</tr>
<tr>
<td>C</td>
<td>0.76</td>
<td><strong>0.66</strong></td>
<td><strong>0.64</strong></td>
</tr>
<tr>
<td>D</td>
<td>0.90</td>
<td>0.78</td>
<td>0.77</td>
</tr>
</tbody>
</table>

- Clearly there are benefits to using all available data
  - Reduced bias
  - Increased efficiency
- Can we do better?
Joint models

- Semi-parametric outcome model

\[ \text{logit } P[Y_{ij} = 1 \mid X_i(t)] = \mu(t) + \beta'X_i(t) \] (1)

- Observation-time model

\[ E[dN_i(t) \mid Z_i(t)] = \exp\{\gamma'Z_i(t)\}d\Lambda(t) \] (2)

- \( \mu(t) \): arbitrary function of time
- \( \beta \): parameters of interest
- \( d\Lambda(t) \): baseline intensity function
- \( \gamma \): intensity parameters

★ Dependence induced by shared covariates in \( X_i(t) \) and \( Z_i(t) \)
★ Requires that \( Z_i(t) \) is correctly specified
Joint models

- Semi-parametric outcome model

\[
\text{logit } P[Y_{ij} = 1 \mid X_i(t)] = \mu(t) + \beta' X_i(t)
\] (1)

▷ \( \mu(t) \): arbitrary function of time
▷ \( \beta \): parameters of interest

- Observation-time model

\[
E[dN_i(t) \mid Z_i(t)] = \exp\{\gamma' Z_i(t)\} d\Lambda(t)
\] (2)

▷ \( d\Lambda(t) \): baseline intensity function
▷ \( \gamma \): intensity parameters

★ Dependence induced by shared covariates in \( X_i(t) \) and \( Z_i(t) \)
★ Requires that \( Z_i(t) \) is correctly specified
Joint models: Estimation

Observation-time model (Lin et al., 2000)

\[ U_\gamma(\gamma) = \sum_{i=1}^{n} \int_0^\tau \{ X_i(t) - \bar{X}(t; \gamma) \} dN_i(t) \]  

(3)

with

\[ \bar{X}(t; \gamma) = \frac{\sum_i 1[C_i > t] \exp\{\gamma'X_i(t)\}X_i(t)}{\sum_i 1[C_i > t] \exp\{\gamma'X_i(t)\}} \]
Joint models: Estimation

For continuous outcomes it is straightforward to estimate $\mu(t)$

$$M_i(t; \beta, \gamma) = \int_0^t Y_i(u)1[C_i > u]dN_i(u)$$

$$- \int_0^t \{\mu(u) + \beta'X_i(u)\}1[C_i > t] \exp\{\gamma'X_i(u)\}d\Lambda(u)$$

- $M_i(t; \beta, \gamma)$ is a mean-zero process
- $dM_i(t)$ and $\int_0^t X_i(u)dM_i(u)$ are also mean-zero processes
- Estimation is based a set of estimating equations of the form

$$\sum_{i=1}^n dM_i(t; \beta, \Lambda) = 0 \quad (4)$$

$$\sum_{i=1}^n \int_0^T X_i(t)dM_i(t; \beta, \Lambda) = 0 \quad (5)$$

- Solve for $\mu(t)$ in (4) and plug into (5) to solve for $\beta$
Joint models: Estimation

Continuous outcome model (Bůžková and Lumley, 2009)

\[ U_\beta(\beta, \hat{\gamma}, \delta) = \sum_{i=1}^{n} \int_{0}^{T} \frac{W(t)}{\omega_i(t; \hat{\gamma}, \delta)} [X_i(t) - \bar{X}(t; \delta)] \]
\[ \times \{ Y_i(t) - \bar{Y}^*(t; \delta) - \beta'[X_i(t) - \bar{X}(t; \delta)] \} \ dN_i(t) \]  

(6)

with

\[ \omega_i(t; \gamma, \delta) = \frac{\exp\{\gamma' Z_i(t)\}}{\exp\{\delta' X_i(t)\}} \]
\[ \bar{X}(t; \delta) = \frac{\sum_i 1[C_i > t] \exp\{\delta' X_i(t)\} X_i(t)}{\sum_i 1[C_i > t] \exp\{\delta' X_i(t)\}} \]
\[ \bar{Y}^*(t; \delta) = \frac{\sum_i 1[C_i > t] \exp\{\delta' X_i(t)\} Y_i^*(t)}{\sum_i 1[C_i > t] \exp\{\delta' X_i(t)\}} \]

* Observations are weighted by inverse intensity rate ratio (IIRR)
Joint models: Estimation

For binary outcomes it is not straightforward to estimate $\mu(t)$

$$M_i(t; \beta, \gamma) = \int_0^t Y_i(u)1[C_i > u]dN_i(u)$$

$$- \int_0^t \left( \frac{\exp\{\mu(t) + \beta'X_i(t)\}}{1 + \exp\{\mu(t) + \beta'X_i(t)\}} \right) 1[C_i > u] \exp\{\gamma'X_i(u)\}d\Lambda(u)$$

- For continuous outcomes used a set of estimating equations

$$\sum_{i=1}^n dM_i(t; \beta, A) = 0 \quad (4)$$

$$\sum_{i=1}^n \int_0^\tau X_i(t)dM_i(t; \beta, A) = 0 \quad (5)$$

- For binary outcomes cannot directly solve for $\mu(t)$ in (4) and plug into (5)
Alternatives for binary outcomes

1. Solve for $\mu(t)$ and $\beta$ simultaneously using an iterative procedure

\[
\sum_{i=1}^{n} dM_i(t; \beta, A) = 0
\]  

\[
\sum_{i=1}^{n} \int_{0}^{\tau} X_i(t)dM_i(t; \beta, A) = 0
\]

- $\mu(t)$ is infinite dimensional so computational issues may arise
- Sparse data resulting from few subjects at unique observation times

2. Impose some (still flexible) structure on $\mu(t)$ using b-splines
   - Iterative procedure with b-splines
   - GEE with b-splines
Warfarin trial: Joint models

- Semi-parametric outcome model
  - Logistic regression model for odds of out-of-range INR
    \[
    \text{logit } P[Y_{ij} = 1] = \mu(t) + \beta' X_i(t)
    \]

- Observation-time models
  1. Outcome-model covariates, including treatment group
    \[
    \lambda_i(t) = \lambda(t) \exp\{ \gamma' X_i(t) \}
    \]
  2. Outcome-model covariates and out-of-range INR at previous visit
    \[
    \lambda_i(t) = \lambda(t) \exp\{ \gamma' X_i(t) + \gamma' Y_i(t - 1) \}
    \]

used to calculate observation-level IIRR weights for outcome model
Warfarin trial: Joint models

- Semi-parametric outcome model
  - Logistic regression model for odds of out-of-range INR
    \[
    \text{logit } P[Y_{ij} = 1] = \mu(t) + \beta'X_i(t)
    \]

- Observation-time models
  1. Outcome-model covariates, including treatment group
    \[
    \lambda_i(t) = \lambda(t) \exp\{\gamma'_X X_i(t)\}
    \]
  2. Outcome-model covariates and out-of-range INR at previous visit
    \[
    \lambda_i(t) = \lambda(t) \exp\{\gamma'_X X_i(t) + \gamma'_Y Y_i(t - 1)\}
    \]

used to calculate observation-level IIWR weights for outcome model
<table>
<thead>
<tr>
<th>Group</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.032</td>
<td>0.026</td>
</tr>
<tr>
<td>C</td>
<td>−0.036</td>
<td>−0.014</td>
</tr>
<tr>
<td>D</td>
<td>0.004</td>
<td>0.022</td>
</tr>
<tr>
<td>Out-of-range</td>
<td>—</td>
<td>0.367</td>
</tr>
</tbody>
</table>

- No clear differences in visit intensity between randomized groups, but agrees with differences in number of INRs per participant
- Out-of-range INR associated with an increase in visit intensity
### Warfarin trial: Results

IIRR weights for outcome model

<table>
<thead>
<tr>
<th></th>
<th>$P_0$</th>
<th>$P_{25}$</th>
<th>$P_{50}$</th>
<th>$P_{75}$</th>
<th>$P_{100}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.86</td>
<td>1.05</td>
<td>1.11</td>
<td>1.16</td>
<td>1.36</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.55</td>
<td>0.70</td>
<td>0.95</td>
<td>1.00</td>
<td>1.16</td>
</tr>
</tbody>
</table>

- Modest variability in weights within models
- Spread of weights similar between models
<table>
<thead>
<tr>
<th>Group</th>
<th>GEE models</th>
<th>Joint models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monthly</td>
<td>Unweighted</td>
</tr>
<tr>
<td>B</td>
<td>1.14</td>
<td>0.95</td>
</tr>
<tr>
<td>C</td>
<td>0.76</td>
<td>0.66</td>
</tr>
<tr>
<td>D</td>
<td>0.90</td>
<td>0.78</td>
</tr>
</tbody>
</table>

- Similar results between models that use all available data
Summary

- Longitudinal data with potentially ‘informative’ observation times motivate consideration of joint models that specify a dependence structure between outcome and observation-time processes.
- Outcome-observation dependence may be induced by shared covariates in outcome and observation-time processes.
- Considered a semi-parametric outcome model that did not specify the functional form for temporal trends:
  - Outcome model weighted by inverse intensity rate ratio
  - Required correctly specified observation-time model
  - Model verification and/or sensitivity analyses may be required
- Extended methods for continuous outcomes to binary outcomes, which required some flexible specification for temporal trends.
Future directions

- Consider latent variables, in addition to covariates, to induce dependence between outcomes and visit process
- Exploit joint models to estimate prediction rules
  - Patient perspective
    - Risk of future poor outcomes given outcome history
    - Optimal time to next visit given visit and outcome history
  - Provider perspective
    - Evaluate patient prognosis, inform personalized treatment strategies
    - Quantify demand for care infrastructure, e.g., beds, staff
References

dbe.med.upenn.edu/biostat-research/bcfrench