Framing and Answering Scientific Questions About Trajectories Using Complex, Longitudinal Data

Scott L. Zeger
Professor of Biostatistics
The Johns Hopkins University
Bloomberg School of Public Health
Essential Questions about Patient or Population Health

1. What is this person’s health state given current measurements?

2. What is the person’s health “trajectory”?

3. What is the optimal choice of intervention for this person?

4. Is the intervention being used optimally in the population? How much does it improve the population’s health at what cost?
Variability is the law of life, and as no two faces are the same, so... no two individuals react alike and behave alike under the abnormal conditions which we know as disease.
– William Osler
Be a clinician (or parent)

- 5 year old child, no family history of disability X, tests “positive” in a screening test
- What is his prognosis?
- What action do you, his clinician recommend to his parents?

**Data from population of “similar” people**

<table>
<thead>
<tr>
<th>Test result</th>
<th>True disease status</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Yes</td>
<td>15</td>
<td>985</td>
<td>1,000</td>
</tr>
<tr>
<td>Negative</td>
<td>5</td>
<td>8,995</td>
<td>9000</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>9,980</td>
<td>10,000</td>
<td></td>
</tr>
</tbody>
</table>
Johns Hopkins Goals the Next Scientific Revolution in Medicine

- Create the computational analogue of the 2x2 table for more complex measurements
  
  Population ⇔ Patient

- Build capacity to make tables for ever narrower sets of “otherwise similar” patients
  
  Subset, Subset, Subset
Leading the Biomedical Revolution - L.B. Minor, 2015

Preeminent Clinical Care

Transformative Biomedical Platforms

Fundamental Research and Biomedical Data Science
Is this man’s prostate tumor lethal or indolent?
Variability is the law of life, and as no two faces are the same, so... no two individuals react alike and behave alike under the abnormal conditions which we know as disease.

– William Osler
Talk outline

1. OSLER’s big idea on which JHM led the first medical revolution and can decide to lead the second

1. Longitudinal study/data analysis
   – Origins in regression analysis
   – Use each subject as own control -> more precise estimates of change
   – Disentangle direction of causation

2. Models for trajectories using \textit{mixed effects} aka \textit{hierarchical} models

1. Individualized health and longitudinal data analysis to identify subsets
   – Prostate cancer “solution”
Regression: heights of fathers and sons

Growth of Nepali Children
Growth of Nepali Children
Growth of Nepali Children
Growth of Nepali Children
Comparison of cross-sectional and longitudinal models
Comparing cross-sectional (wrong) with longitudinal (right) analyses

|       | Estimate | Std. Error | t value | Pr(>|t|) | CI LB | CI UB |
|-------|----------|------------|---------|----------|-------|-------|
| (Intercept) | 3.2893   | 0.7116     | 4.6225  | 0.0000   | 1.8946 | 4.6841 |
| age    | 0.5489   | 0.1428     | 3.8450  | 0.0001   | 0.2691 | 0.8288 |
| age_sp1| -0.4348  | 0.1830     | 2.3761  | 0.0177   | -0.7935| -0.0761|
| age_sp2| 0.0638   | 0.0709     | 0.8992  | 0.3688   | -0.0752| 0.2028 |
| age_sp3| -0.0375  | 0.0207     | 1.8126  | 0.0703   | -0.0780| 0.0030 |

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>pval</th>
<th>CI LB</th>
<th>CI UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.7675</td>
<td>0.2738</td>
<td>13.7609</td>
<td>0.0000</td>
<td>3.2309</td>
<td>4.3041</td>
</tr>
<tr>
<td>age</td>
<td>0.5021</td>
<td>0.0459</td>
<td>10.9465</td>
<td>0.0000</td>
<td>0.4122</td>
<td>0.5920</td>
</tr>
<tr>
<td>age_sp1</td>
<td>-0.3611</td>
<td>0.0569</td>
<td>6.3484</td>
<td>0.0000</td>
<td>-0.4726</td>
<td>-0.2496</td>
</tr>
<tr>
<td>age_sp2</td>
<td>0.0135</td>
<td>0.0235</td>
<td>0.5740</td>
<td>0.5660</td>
<td>-0.0326</td>
<td>0.0596</td>
</tr>
<tr>
<td>age_sp3</td>
<td>-0.0252</td>
<td>0.0095</td>
<td>2.6443</td>
<td>0.0082</td>
<td>-0.0439</td>
<td>-0.0065</td>
</tr>
</tbody>
</table>
Talk outline

1. OSLER’s big idea on which JHM was built

1. Longitudinal study/data analysis
   – Origins in regression analysis
   – Use each subject as own control-> more precise estimates of change
   – Disentangle direction of causation

2. Models for trajectories using *mixed effects* aka *hierarchical* models

1. Individualized health and longitudinal data analysis to identify subsets
   – Prostate cancer “solution”
Schizophrenia Study

• Phase II double blind randomized trial of risperidone (r) at 2, 6, 10 and 16 mg versus placebo (p) and haloperidol 20 mg (h) for treatment of symptoms of schizophrenia;

• Outcome: PANSS scale (higher -> more symptoms);

• Patients observed at 0, 1, 2, 4, 6 and 8 weeks;

• Substantial drop-out
# Drop Outs

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>h</th>
<th>r</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drop-outs</td>
<td>26</td>
<td>52</td>
<td>172</td>
<td>270</td>
</tr>
<tr>
<td>Completers</td>
<td>53</td>
<td>43</td>
<td>174</td>
<td>252</td>
</tr>
<tr>
<td>Total</td>
<td>79</td>
<td>95</td>
<td>346</td>
<td>522</td>
</tr>
</tbody>
</table>
Modeling trajectories for individuals; “Mixed effects” or “hierarchical” models

Basic ideas

1. There is natural heterogeneity across individuals in their true levels and changes over time (regression coefficients) and this heterogeneity can be described by a probability distribution

1. Conditional on true pattern for each individual, responses are independent
Mixed model fits

Model 1 – Random intercept (solid)

Model 2 – Random intercept and random linear slope (dashed)
Model-based estimates of treatment effect: R–vs-P with valid confidence intervals

Comparison of Risperidone to Control

Total PANSS Score Difference

Weeks
Correlations among PANSS repeated scores for each person

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00</td>
<td>0.64</td>
<td>0.53</td>
<td>0.48</td>
<td>0.43</td>
<td>0.37</td>
</tr>
<tr>
<td>1</td>
<td>0.64</td>
<td>1.00</td>
<td>0.80</td>
<td>0.68</td>
<td>0.60</td>
<td>0.58</td>
</tr>
<tr>
<td>2</td>
<td>0.53</td>
<td>0.80</td>
<td>1.00</td>
<td>0.80</td>
<td>0.72</td>
<td>0.64</td>
</tr>
<tr>
<td>4</td>
<td>0.48</td>
<td>0.68</td>
<td>0.80</td>
<td>1.00</td>
<td>0.83</td>
<td>0.77</td>
</tr>
<tr>
<td>6</td>
<td>0.43</td>
<td>0.60</td>
<td>0.72</td>
<td>0.83</td>
<td>1.00</td>
<td>0.86</td>
</tr>
<tr>
<td>8</td>
<td>0.37</td>
<td>0.58</td>
<td>0.64</td>
<td>0.77</td>
<td>0.86</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Covariances by treatment group

9/3/2015 Bayview Trajectories
Estimates of Individual Curves
Talk outline

1. OSLER’s big idea on which JHM was built

1. Longitudinal study/data analysis
   – Origins in regression analysis
   – Use each subject as own control-> more precise estimates of change
   – Disentangle direction of causation

2. Models for trajectories using *mixed effects* aka *hierarchical* models

1. Individualized health and longitudinal data analysis to identify subsets
   – Prostate cancer “solution”
True Prostate Cancer State (Latent) → Observed PSA → Biopsy Performed → Biopsy Results → Surgical Removal (Observe True State)
Diagnosis

P(Gleason 7+)

40%

5 Years Follow-up

2%

10 Years Follow-up

2%

All Available

PSA (ng/mL)

Age (years)

PSA (ng/mL)

Age (years)

PSA (ng/mL)

Age (years)

Probability Reclassification

Age (years)

Probability Reclassification

Age (years)

Probability Reclassification

Age (years)
Main Points Again

1. OSLER’s big idea on which JHM was built

1. Longitudinal study/data analysis
   – Origins in regression analysis
   – Use each subject as own control-> more precise estimates of change
   – Disentangle direction of causation

2. Models for trajectories using *mixed effects* aka *hierarchical* models

1. Individualized health and longitudinal data analysis to identify subsets
   – Prostate cancer “solution”
Health State/Trajectory ($\eta_{it}$) with Person-specific Indicator ($\delta_i$)
Effects of Exogenous (X) and Endogenous (Rx) Covariates on Health State/Trajectory with Person-specific Regression Coefficients ($\beta_i$)
Observations ($Y$) that Inform about Health State through Coefficients ($\varphi_i$)
\[
\beta_i, \Sigma
\]

\[
\theta
\]

\[
\phi_i
\]

\[
\delta_i
\]

\[
\xi
\]

\[
\eta_i
\]

\[
X_{it}, R_{xit}
\]

\[
Y_{it-1}, Y_{it+1}
\]

\[
9/3/2015 Bayview Trajectories
\]

\[
52
\]
Why are cross-sectional and longitudinal data model estimates of average trend so different?

Least Squares Fits of Natural Splines

Random Effects (Int, Time) Fits of Natural Splines

Failure to model trajectories leads to incorrect inferences in the presence of substantial drop-out (drop-out from death is another issue)
Boole -a- Bayes

George Boole 1815-1864

Thomas Bayes 1701-1761
Individualized Health
Active Surveillance of Prostate Cancer

Keys to Success
1. Identify potentially lethal cancer when present
2. Correctly diagnose indolent cancer
3. Maintain patient engagement

Patient Priorities
1. Avoid negative side effects of curative intervention
2. Limit the number, frequency of painful biopsies
3. Reduce anxiety