Statistical Considerations
The Use of Mucosal Assays in Microbicide Trials
August 25, 2015

Barbra Richardson, Ph.D.
Department of Biostatistics, University of Washington
Statistical Center for HIV/AIDS Research & Prevention
Seattle, WA, USA
Outline

- **Design issues:**
  - Hypotheses
  - Sampling
  - Sample size/Power

- **Analysis issues:**
  - Statistical Analysis Plan
  - Multiple Comparisons
  - Dimension Reduction
Design Issues - Hypotheses

- Mucosal assay results in microbicide trials
  - Generally secondary or exploratory endpoints
    - Still deserve well defined hypotheses
  - Numerous hypotheses (this is ok)

- A priori: Why do we care about these assay results and what are the hypotheses regarding them?
Timing of sampling and your hypotheses

- Baseline sampling
  - hypotheses re: within participant changes

- Longitudinal sampling
  - Sampling frequency, timing addresses hypotheses
    - Acute versus chronic exposure to microbicide
Mucosal assay results in microbicide trials usually limited by available sample size

Generally 5 relevant variables:
- Sample size
- False positive rate ($\alpha$) – 0.05
- Power (1-false negative rate) – 80% or 90%
- Magnitude of effect size (hypothesized)
- VARIABILITY!
Design Issues – Variability

- Variability
  - Within assay (noise)
  - Within participant
  - Between participant
Design Issues – Variability

- Within assay variability (noise)
  - Consider 3 replicates of one sample

<table>
<thead>
<tr>
<th>Assay</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>100</td>
<td>90</td>
<td>49</td>
</tr>
<tr>
<td>B</td>
<td>40</td>
<td>60</td>
<td>50</td>
<td>10</td>
</tr>
</tbody>
</table>

- Assay A will require much larger sample size than assay B to discern a similar magnitude of difference
Analysis Issues – Statistical Analysis Plan

- Statistical analysis plan includes at minimum
  - Hypotheses
  - Endpoints
  - Analysis population description
  - Statistical methods
    - Transformation of variables – Normality or categorization (lower limit of detection)
    - Statistical tests to be used
    - Potential covariates
    - Methods for accounting for multiple comparisons
## Analysis Issues – Multiple Comparisons 101

<table>
<thead>
<tr>
<th>DECISION</th>
<th>TRUTH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H₀ True</td>
</tr>
<tr>
<td>Do Not Reject H₀</td>
<td>CORRECT 1-α</td>
</tr>
<tr>
<td>Reject H₀</td>
<td>INCORRECT β</td>
</tr>
</tbody>
</table>

- **H₀ True**
  - **Do Not Reject H₀**: CORRECT 1-α
  - **Reject H₀**: INCORRECT β (false positive)

- **H₀ False**
  - **Do Not Reject H₀**: INCORRECT β (false negative)
  - **Reject H₀**: CORRECT 1-β (power)
Want to control probability of a false positive result ($\alpha$)

![Graph showing the relationship between the number of hypothesis tests and the probability of a false positive result.](image)
Analysis Issues – Multiple Testing Methods

I can’t live with ANY false positive results!

- Methods that control the “Family Wise Error Rate” (FWER) = Pr(at least one false positive)
  - Single step
    - Bonferroni: reject any hypothesis with p-value ≤ α/m (m is number of tests)
    - Too conservative – high probability of false negative results
  - Sequential
    - Holm’s Method, Simes’ Method, others
    - Different criteria for magnitude of p-value rejected
    - Choice depends on correlation of hypothesis tests as well as other factors
Analysis Issues – Multiple Testing Methods

I can live with some false positive results…..

- Methods that control the “False Discovery Rate” (FDR) = proportion of false positives among the set of rejected hypotheses
  - Strive to keep the FDR below a threshold “q” – defined as the q-value
  - Benjamini and Hochberg FDR
  - Storey’s positive FDR (pFDR)
Analysis Issues – Multiple Testing Methods

False Discovery Rate (FDR) versus False Positive Rate (FPR)

<table>
<thead>
<tr>
<th>DECISION</th>
<th>TRUTH</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H₀ True</td>
<td>H₀ False</td>
</tr>
<tr>
<td>Call H₀</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>True (do not</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reject)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Call H₀</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>False (reject)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>100</td>
<td>25</td>
</tr>
</tbody>
</table>

FDR = 20% (5/25)

FPR = 5% (5/100)
Analysis Issues – Dimension Reduction

- Numerous mucosal assay outcome variables
  - Are there some variables that cluster together to mark a similar underlying biological mechanism?

- Methods for reducing dimension (combining variables)
  - Principal components analysis
  - Linear discriminant analysis
  - Canonical correlation analysis
  - Others
Example: MTN 004 MTN BSWG Analyses (Pellett Madan, et al, 2015)
- 61 women with 4 visits (baseline, 7 days, 14 days and 21 days)
- IL-1β, IL-6, IL-12p40, MIP-1α, GM-CSF, lactoferrin and SLPI from cervical swabs

Soluble immune mediator score created using factor analysis with principal components extraction
Analysis Issues – Dimension Reduction

- Example: MTN 004 MTN BSWG Analyses (Pellett Madan, et al, 2015)

  - Soluble immune mediator score created using factor analysis with principal components extraction
  - Score used in analyses to see if it was predictive of subsequent endogenous activity against *E. coli*
  - Dimension reduced from 7 hypothesis tests (7 separate assay results) to 1 (score) – probability of at least one false positive reduced from ~30% to 5%
Conclusions

- **Design:**
  - If possible build mucosal assays into study design up front
  - Timing of sampling
  - Sample size/Power
  - DRIVEN BY HYPOTHESES! *A priori*: Why do we care about these assay results and what are the hypotheses regarding them?
Conclusions

- Analysis:
  - Statistical Analysis Plan
    - Multiple testing procedures
    - Possibility of dimension reduction?
  - DRIVEN BY HYPOTHESES! *A priori*: Why do we care about these assay results and what are the hypotheses regarding them?
Acknowledgments

- Fred Hutchinson Cancer Research Center
  - Elizabeth Brown
  - Raphael Gottardo
Design Issues – Sampling Noise

- “Noisy” assays
  - Separate signal from noise
    - Baseline sampling
    - Placebo sampling
Design Issues – Variability

- Within participant variability
  - Consider data on two participants from 3 timepoints for a particular assay

<table>
<thead>
<tr>
<th>Participant</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>10</td>
<td>100</td>
<td>90</td>
<td>49</td>
</tr>
<tr>
<td>Y</td>
<td>40</td>
<td>60</td>
<td>50</td>
<td>10</td>
</tr>
</tbody>
</table>

- Participant X’s assay results are much more variable over time than participant Y’s. Harder to see a smaller signal in participants like X.