Statistical Issues in Assessing Treatment Efficacy and Correlates of Protection in mAb Efficacy Trials

Peter Gilbert, Michal Juraska, Yunda Huang, Allan deCamp

Vaccine and Infectious Disease Division
Fred Hutchinson Cancer Research Center

Passive Immunization Trials to Inform Vaccine Design
August 8, 2014
Outline

1. Correlates trial objectives and possible designs
2. mAb markers and mAb sieve analysis
3. Sample size for overall treatment efficacy
4. Sample size for correlates of protection
5. Discussion
Correlates Trial Objectives

1. To develop a marker of the mAb treatment that correlates with protection against HIV infection
2. To provide insight into the mechanistic correlates of protection

Application: Help define immunogenicity study endpoints in Phase I/II trials for evaluating candidate neutralizing antibody-based HIV vaccines
Competing Trial Objectives

- **Efficacy trial**: Maximize the overall treatment efficacy
- **Correlates trial**: Maximize the variability of treatment efficacy while achieving reasonably high overall efficacy
Correlates Trial Design Requirements for Success

1. Beneficial overall treatment efficacy (TE)
   - TE between about 30% and 80% (50%–70% ideal)

2. Wide variability in neutralization characteristics of the mAb over time and/or individuals, and the ability to measure the characteristics
   - Wide potentially protection-relevant variability and low assay noise

3. Trial implementation enabling statistical analysis to characterize a correlate
   - Adequate sample size (number of breakthrough cases in mAb recipients)
   - Adequate HIV testing schedule and diagnostics
   - Adequate mAb marker sampling design and modeling over time
Will Discuss Two Potential Trial Designs

1. **[Two-arm Design]** mAb vs. placebo on a regular infusion schedule such as bi-monthly

<table>
<thead>
<tr>
<th>Infusions (Months)</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>mAb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. **[Multi-Arm Design]** Multiple mAb interventions + placebo

<table>
<thead>
<tr>
<th>Infusions (Months)</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>mAb trt1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mAb trt2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mAb trt3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sampling of mAb Markers for Assessing Immune Correlates

- Measure markers in the mAb group: all HIV infected cases and a random sample of HIV uninfected controls:
  - All infusion/trough visits prior to HIV infection
  - ~3 additional time-points for estimating individual mAb concentration-time curves for the entire follow-up period
  - More intensive PK sub-study in a random sample of subjects

Samples for Measurement in the mAb group

HIV infected (cases)   XX X X X X X X X
HIV uninfected (controls) XX X X X X X X X X X X X
mAb Marker Variability

• Two types of variability
  • Over time within mAb recipients
  • Among mAb recipients
mAb Marker Variability Over Time

- **Approach:** Identify a correlate by comparing HIV incidence during low mAb periods vs. high mAb periods.

- Must include HIV tests between peaks and troughs to learn about correlates.
Extending the Dosing Schedule May Improve Correlates Assessment

- Ideally will have large segments of person-years at risk in putative high protection zones and in putative lower protection zones
mAb Marker Variability Among mAb Recipients

- Variability exists naturally and may be created by design

- How much natural inter-individual variability?
  - VRC 601, 602, HVTN 104 are providing information
  - May create variability by randomizing to different mAb doses or schedules
Outline

1. Correlates trial objective and possible designs
2. mAb markers and mAb sieve analysis
3. Sample size for overall treatment efficacy
4. Sample size for correlates of protection
5. Discussion
Which mAb Markers to Assess as Correlates?

• Serum concentration of mAb (BAMA or ELISA)
  • **Advantage:** Marker defined without reference to a particular HIV-1(s)

• Serum mAb neutralization magnitude and breadth to HIV-1 Envs reflecting variability of the mAb footprint
  • **Advantage:** May link to an immunogenicity endpoint that would be used in future vaccine trials

• mAb binding and other effector functional activities such as ADCC
Mucosal mAb Markers for Correlates

• Of interest for assessing mechanistic correlates of protection and for interpreting the serum results

• Questions:
  • Practicable for predictive correlates assessment? E.g., can samples be collected from almost all subjects using swabs?
  • Do the mucosal assays have low enough noise?
Which HIV-1 Env Panel(s)?

- Most relevant HIV-1 Env panel would represent the mAb footprint diversity of viruses exposing trial participants

Logo plot of VRC01 mAb contact residues*

- Logo plot based on 275 subtype B U.S./Peru sequences deposited in the LANL sequence database 2006 or later

*Zhou et al. Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01 (2010, Science)
mAb Epitope-Specific Markers

- Marker = Measurement of the ability of a mAb recipient’s sera to neutralize the epitope targeted by the mAb
  - Account for mutations in the epitope that do not cause neutralization resistance

- For each mAb group infected subject:
  Marker = Ability of his/her pre-infection sera to neutralize his/her breakthrough virus

  - Sera at peak and at the trough time point prior to infection
  - Model the mAb concentration at the time of infection
mAb Immunoprophylaxis Sieve Analysis

1. **Genotypic**: Identify “signatures” that differentiate mAb footprint breakthrough sequences in the mAb vs. placebo group. E.g., comparing:
   i. Number of known mAb escape mutations
   ii. Number of PNGs within a given radius of the mAb footprint
   iii. Length of variable loops in proximity to the mAb footprint

2. **Phenotypic [Neutralization Sieve Analysis]***: Compare breakthrough viruses between groups using an immunological assay. E.g., assess with the TZM-bl assay:
   i. Sensitivity of breakthrough viruses to the mAb
   ii. Sensitivity of breakthrough viruses to mAb recipient sera

*Neutralization sieve analysis in the context of HIV vaccine efficacy trials: Gilbert, deCamp, Montefiori (2010, JID); Montefiori (2014, Hum Vacc & Immunotherapy)
Simulated Genotypic & Phenotypic Comparisons

**Genotypic**

- Number of Breakthrough Viruses
  - Placebo: 50
  - mAb: 40

- Number of Escape Mutations
  - Placebo: 3
  - mAb: 2

**Phenotypic**

- mAb Concentration (μg/ml) for 50% neutralization (IC50)
- Placebo (n=50)
- mAb (n=22)

Percentage of Breakthrough Strains

- Placebo
- mAb
Average Magnitude-Breadth of Neutralization Curves for Simulated ID50 Values

Average MB-curves

- Peak sera, Placebo strains
- Trough sera, Placebo strains
- Peak sera, mAb group strains
- Trough sera, mAb group strains

Percentage of Breakthrough Strains

ID50 for mAb recipient sera
Simulated Genotypic and Autologous Neutralization Response Comparisons

Result: Autologous neutralization is low and typically undetectable (≤ 10) for escape counts ≥ 1
Outline

1. Correlates trial objective and possible designs
2. mAb markers and mAb sieve analysis
3. **Sample size for overall treatment efficacy**
4. Sample size for correlates of protection
5. Discussion
Some Questions for Trial Design

• What study population and HIV incidence projected in the placebo group?
• How to balance the tradeoff of the false positive rate and power for testing $H_0: \text{TE} \leq 0\%$ vs. $H_1: \text{TE} > 0\%$?
• What allocation ratio to mAb treatment vs. placebo?
What Population and Incidence in the Placebo Group? Consider MSMs in the Americas Including Peru

Annual Incidence in Placebo Group, with 4% Baseline Incidence, Under Different Levels of PrEP Use and Levels of PrEP Efficacy*

*Similar analysis to that conducted in Janes et al. (2013, AIDS Research and Human Retroviruses)
What Population and Incidence in the Placebo Group? Consider MSMs in the Americas Including Peru

Annual Incidence in Placebo Group, with 4% Baseline Incidence, Under Different Levels of PrEP Use and Levels of PrEP Efficacy*

*Similar analysis to that conducted in Janes et al. (2013, AIDS Research and Human Retroviruses)
Total Sample Size [mAb Group + Placebo Group] for 90% Power to Detect Overall TE with 2 Years of Follow-up for Infection*

*1-sided $\alpha=0.025$- or 0.10-level log-rank test of $H_0$: TE $\leq 0\%$ in the presence of sequential monitoring

3% annual placebo incidence
5% annual drop-out rate
HIV testing monthly for up to 24 months
Results for Overall TE [Two-Arm Trial]

• Small sample size increase moving from 1:1 to 2:1
  • Prefer 2:1 to improve power for correlates

• Large sample size decrease moving from $\alpha=0.025$ to $\alpha=0.10$
  • While $\alpha=0.10$ can be fitting for a proof-of-concept efficacy trial*, for a correlates trial it is likely too small
  • We focus on the 2:1, $\alpha=0.025$, N=4000 design, with more justification later with correlates calculations

*Fleming and Richardson (2004, JID); Mehrotra et al. (2006, Biometrics); Gilbert (2010, Statistics in Medicine)
Total Sample Size N=4000 [2:1 mAb:Placebo]: Probabilities of Reaching Each Possible Conclusion

Calculations Include group sequential monitoring for:
- Efficacy futility
- Potential harm
- High efficacy
Total Sample Size **N=4000 [2:1 mAb:Placebo]**: Cumulative Numbers of HIV Infection Endpoints by Treatment for TE=50%

<table>
<thead>
<tr>
<th>Trial Time Since First Person In (Month)</th>
<th>Median Number of Placebo Endpoints</th>
<th>Median Number of mAb Group Endpoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td>24</td>
<td>55</td>
<td>51</td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>66</td>
</tr>
<tr>
<td>36</td>
<td>77</td>
<td>73</td>
</tr>
</tbody>
</table>

- Expect a total of ~150 HIV infection endpoints
Outline

1. Correlates trial objective and possible designs
2. mAb markers and mAb sieve analysis
3. Sample size for overall treatment efficacy
4. Sample size for correlates of protection
5. Discussion
Correlates Analysis

- **Goal:** Estimate how the TE varies over subgroups defined by the mAb-marker level
Power Calculations for a Binary mAb Marker

- We focus on power to detect a binary correlate*
- Divide the mAb group by a threshold of marker response
  - Lower protected vs. Higher protected

<table>
<thead>
<tr>
<th>Different Correlate Effect Sizes</th>
<th>Overall TE</th>
<th>TE Lower Protected</th>
<th>TE Higher Protected</th>
<th>Strength of CoP</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>None/Null</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>40%</td>
<td>60%</td>
<td>Weak</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>30%</td>
<td>70%</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>20%</td>
<td>80%</td>
<td>Strong</td>
<td></td>
</tr>
</tbody>
</table>

*The same analysis applies for designs with one or multiple doses of the mAb treatment
Power Calculations for a Binary mAb Marker

- Divide the mAb group into subgroups defined by $S(t) = 1$ vs. $S(t) = 0$
  - $S(t) = 1$ if the mAb marker at time $t$ is above a fixed threshold
  - Need plausibility of a large difference in TE above and below the threshold
Power to Discriminate 25% vs. 75% TE [Overall TE = 50%, 2-Arm Trial]

Total Sample Size [N=4000 2:1 mAb:Placebo]

- 2-phase logistic regression; 2-sided alpha = 0.05; controls:cases = 5:1, 50% in higher protection group

Number of Infections in mAb Group vs. Power

- Measurement error level:
  - Power rho=1
  - Power rho=0.9
  - Power rho=0.7
  - Power rho=0.5

2000 4000 6000 8000
Power
0.0 0.2 0.4 0.6 0.8 1.0

08/08/2014 • 33
Power to Discriminate Different TEs [Overall TE = 50%, 2-Arm Trial]: rho = 0.9

2-phase logistic regression; 2-sided alpha = 0.05; controls: cases = 5:1, 50% in higher protection group
Conclusions: Correlates Power Results for N = 4000 Two-Arm Trial

- Well-powered for correlates under the scenarios:

<table>
<thead>
<tr>
<th>TE Lower Protected</th>
<th>TE Higher Protected</th>
<th>Overall TE</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>75%</td>
<td>50%</td>
</tr>
<tr>
<td>0%</td>
<td>60%</td>
<td>30%</td>
</tr>
<tr>
<td>55%</td>
<td>85%</td>
<td>70%</td>
</tr>
</tbody>
</table>

- Under-powered for smaller gradients of TE
Assigning Marker Variability to Assess a CoP

- Assigning marker variability is very helpful for CoPs assessment, because it allows a direct assessment of the causal effect of different marker levels on protection (i.e., a CoR = CoP by design)

- E.g., dose-response design to directly assess a CoP

<table>
<thead>
<tr>
<th>Infusions (Months)</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>22</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>mAb 10 µg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mAb 20 µg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mAb 50 µg/kg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Power Calculations for the Dose-Response Design**

**Test H$_0$: Same Infection Rate in the 4 Groups**

<table>
<thead>
<tr>
<th></th>
<th>$T_{E_{\text{low}}}$</th>
<th>$T_{E_{\text{med}}}$</th>
<th>$T_{E_{\text{high}}}$</th>
<th>Power to Detect any Departure from H$_0$</th>
<th>Power to Detect Linear Trend in log HR$^*$</th>
<th>Power to Detect $T_{E_{\text{low}}}$ Different from $T_{E_{\text{high}}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>50%</td>
<td>75%</td>
<td>&gt;0.99</td>
<td>&gt;0.99</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>30%</td>
<td>60%</td>
<td>0.98</td>
<td>0.99</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>55%</td>
<td>70%</td>
<td>85%</td>
<td>&gt;0.99</td>
<td>&gt;0.99</td>
<td>0.84</td>
<td></td>
</tr>
</tbody>
</table>

N=4000 with 1:1:1:1 allocation
3% annual placebo incidence
5% annual drop-out rate
HIV testing monthly for up to 24 months

*4-sample log-rank test (2-sided $\alpha=0.05$)
$^*$Score test for linear trend in log HR across placebo, low, medium, high dose (1-sided $\alpha=0.025$)
# Wald test of difference in log HR between low dose and high dose (2-sided $\alpha=0.05$)
Conclusions: Power Results for N = 4000 Dose-Response Design

- High power for detecting dose-response and discriminating TE between low and high mAb doses (25% vs. 75%; 0% vs. 60%; 55% vs. 85%)
- This analysis provides a correlate in terms of average marker level within Low, Medium, High
- Assessment of an individual-level correlate would proceed similarly as for the two-arm design
Summary: Two-Arm vs. Dose-Response Design

• Advantages of dose-response design:
  • Can directly assess an average-marker CoP based on the randomization
  • Can induce greater mAb marker variability among mAb recipients thereby increasing the likelihood of a large gradient in TE and the power to identify an individual-marker CoP
Outline

1. Correlates trial objective and possible designs
2. mAb markers and mAb sieve analysis
3. Sample size for overall treatment efficacy
4. Sample size for correlates of protection
5. Discussion
Goal of the Statistical Trial Design

- **Goal:** To maximize mAb marker variability over time and over mAb recipients, such that there are
  - Large amounts of person-years at risk in putative high protection zones and in putative lower protection zones
  - The HIV diagnostics, marker sampling schedule, marker modeling over time, and assays afford identification of the marker levels present near the time of HIV acquisition events
Integrated Sieve and Correlates Analysis

- Genotypic and phenotypic sieve analysis: Powerful tool to determine whether epitope-specific neutralization is a correlate of protection, and to characterize the specificity of the protection.

- The complementary correlates analysis could identify how treatment efficacy varies with the threshold of marker response, providing a direct input to study endpoints for HIV vaccine trials.
Discussion: Multiple mAbs?

• Pros:
  • Increased marker variability
  • Multiple chances to detect an efficacious mAb

• Cons:
  • Each mAb has distinct biological properties, complicating the interpretation of results that pool over the mAbs

• One option for a multiple mAb trial:
  • 90% power to detect overall TE = 50–60% for each mAb separately
  • Adequate power to detect correlates pooling over the mAbs
Statistical Research Needed to Design an Efficacy Trial

- Expanded design calculations are needed that:
  - Incorporate data on marker variability over time and among mAb recipients [VRC 601, VRC 602, HVTN 104, etc.]
  - Incorporate data on neutralization escape mutations in the mAb footprint to develop epitope specific markers
  - Incorporate models of the mAb time-course that predict an individual’s whole time course from measurements that would be taken in the efficacy trial
- Compare designs by marker variability created from the choice of
  - mAb schedule, dose, schedule × dose
- Select the optimal design
Acknowledgements

Michal Juraska  Allan deCamp  Yunda Huang

Scientific Discussions

- Robert Bailer
- Susan Buchbinder
- Larry Corey
- Barney Graham
- Julie McElrath
- Jim Kublin
- Julie Ledgerwood
- Craig Magaret [+ logo plot]
- David Montefiori
- Yegor Veronin