PrEP Study Design Considerations for a Long-Acting Agent

Robert Cuffe (VHL), Britt Stancil (GSK)

06 May 2013
LAA specific Issues Influencing Phase 3 Design

• Guidance on target safety database
  • Different regulatory sources
  • Population/PK profile/regulatory history are influential
  • Impacts duration, total N, randomisation ratio

• Safety monitoring issues
  • Management of pregnancy/adverse drug reactions after LAA
  • Access to high level supportive care
  • Impacts population/site selection

• Data supporting estimates of PrEP efficacy
  • Ex vivo, preclinical, target tissue PK
  • Impact effect size & N
Randomization

HIV neg and high risk

LAA
All arms receive condoms, counselling and regular testing

Comparator

• Primary endpoint: time to HIV-1 Infection
• Duration: 2 year total (1 year recruitment, min 1 year follow-up)
• Assume 15% drop-out rate

Q1: Population
MSM / women / serodiscordant couples?

Q2: Comparator
TDF/FTC or non-pharmaceutical intervention (NPI)?
Assume 90% efficacy vs. NPI*
Assume 40-80% efficacy vs. TDF/FTC (depending on population)

* With an 80% assumption for sensitivity analyses
General PrEP Sample Size Considerations

- Effect size
- Minimum threshold
- Power
- Signal
- Target number of events
- Incidence rate
- Total study duration
- Follow-up duration
- Target Accrual
Effect of RR and incidence rate upon N

<table>
<thead>
<tr>
<th>Annual Incidence (%)</th>
<th>Risk Reduction (%)</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV Events</td>
<td></td>
<td>161</td>
<td>50</td>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td>0.5</td>
<td>N=</td>
<td>28,000</td>
<td>10,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>N=</td>
<td>2,500</td>
<td>950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>N=</td>
<td>480</td>
<td>260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>N=</td>
<td>320</td>
<td>170</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table Displays Total Number of Subjects Needed to Enroll Active: SOC 1:1

Power=90%, \( \alpha=0.025 \)

Assume LAA risk reduction 80-90% vs. NPI, LAA risk reduction 40-80% vs. TVD, 2 year study (1 year recruitment), \( H_0: RR=1 \), N is rounded to two significant digits
POPULATIONS
Population 1: Serodiscordant Couples

Assumptions

- Incidence rate 0.5%
- Truvada use available
- Risk reduction 60% (incidence rate 0.2% on LAA)

Sample size

<table>
<thead>
<tr>
<th>Total Duration (month)</th>
<th>Time for Recruitment (month)</th>
<th>Risk Reduction vs TVD (number of events required)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>40% (161 events)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60% (50 events)</td>
</tr>
<tr>
<td>24</td>
<td>12</td>
<td>28495</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10128</td>
</tr>
<tr>
<td>24</td>
<td>15</td>
<td>30972</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11008</td>
</tr>
<tr>
<td>48</td>
<td>30</td>
<td>15534</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5521</td>
</tr>
</tbody>
</table>
Population 2: High Risk Women

Assumptions
- Incidence rate 6% (on TVD or NPI)
- Risk Reduction: 90% (0.6% incidence on LAA)
- 15% drop out rate

Total Number of Subjects Needed to Enroll

<table>
<thead>
<tr>
<th>Total Duration (month)</th>
<th>Time for Recruitment (month)</th>
<th>Risk Reduction vs TVD or NPI (number of events required)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>80% (17 events)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>90% (8 events)</td>
</tr>
<tr>
<td>24</td>
<td>12</td>
<td>315</td>
</tr>
<tr>
<td></td>
<td></td>
<td>169</td>
</tr>
</tbody>
</table>

What is the minimum number of events needed for a clinically meaningful demonstration of efficacy?
When is the **right time for trials in women?**

- Extrapolation from men to women is difficult
  - Theoretically – biology & behaviour
  - Clinical trials less compelling for women than men

- Unmet need arguably greater

- Risk of giving LAA with limited safety data to WCBP

- NPI less contentious
  - Smallest study to demonstrate efficacy and effectiveness
  - Open-label study assessing risk mitigation possible
  - Inclusion of TVD makes less difference to assumptions
Population 3: MSM (assuming 4% risk)

Assumptions:
- 15% dropout rate, 12 month recruit + 12 month f/u

vs TVD:
- TVD incidence: 2%
- Risk reduction 80%
  (0.4% incidence on LAA)

vs NPI:
- NPI incidence: 4%
- Risk reduction 90%
  (0.4% incidence on LAA)

Table Displays Total Number of Subjects Needed to Enroll
Active: SOC 1:1

<table>
<thead>
<tr>
<th>Annual Incidence (%)</th>
<th>Risk Reduction (%)</th>
<th>60</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV Events</td>
<td>50</td>
<td>17</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>N= 2515</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>N= 475</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Power=90%, \( \alpha=0.025 \)
NPI vs. TVD (where choice available)

- **TVD**
  - Demonstrated superiority
  - Relevant question for approval

- **NPI**
  - Fastest proof of efficacy
  - Can assess risk migration
  - Relevant question locally

*angle of scales reflects PowerPoint defaults and in no way prejudices the appropriate ethical balance*
STATISTICAL COMPLEXITIES
## Blending heterogeneous groups

- E.g. different comparators in different jurisdictions (TVD vs. NPI)
- Need to ensure representative data from both populations

<table>
<thead>
<tr>
<th>Annual Incidence (%)</th>
<th>Risk Reduction (%)</th>
<th>60</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV Events</td>
<td></td>
<td>50</td>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>N=</td>
<td>2515</td>
<td>955</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>N=</td>
<td>475</td>
<td>255</td>
<td></td>
</tr>
</tbody>
</table>

Table Displays Total Number of Subjects Needed to Enroll
Active: SOC 1:1

Power=90%, $\alpha=0.025$
Different patterns and effects of non-adherence

--- --- --- --- --- --- --- --- --- --- --- --- --- -

Clinic visits

...could bias

1° endpoint: observe non-adherent QD subjects longer

2° analyses: relationship between adherence and response

relationship differs by arm

quality of measurement differs by arm

Solution for 1° endpoint

1. Flexibility in visit scheduling
2. Allow LAA subjects to refuse treatment on-study
3. Seroconversion survey at study end (incl. LTFU)
4. Sensitivity analyses
Interim monitoring

• Formal interim efficacy analyses add little to total N
  • e.g. analyse after 17/29 events adds <2% to sample size,
  • could stop after 16 months
• “Minimum N” already a question for full analysis set

• Constant DMC monitoring enables
  • Balance of events if required
  • Fair adherence on an active comparator
Conclusions

- Still assumption heavy, but N for p<0.05 << N for other qs
  - N Serodiscordant >> MSM (vs. TVD) > MSM (vs. NPI) >/= Women

- Comparator for MSM a complex discussion
  - NPI faster and can assess risk migration
  - accommodation of heterogeneity

- Further details will be resolved at a later date
  - balance between time & N
  - interim monitoring for early conclusion
  - analysis plan for different reflection of adherence
  - blinding...
  - age limits...