Statistical design and data monitoring for personalized medicine interventions

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

59th WSC
Hong Kong, China
28 August 2013
Collaborators

Junghnam Joo, PhD
Korea National Cancer Center

Nancy Geller, PhD
National Heart, Lung and Blood Institute

Stephen Kimmel, MD, MSCE
University of Pennsylvania

Jonas Ellenberg, PhD
University of Pennsylvania
Personalized medicine interventions

- Customized health care tailored to individual patient
- Increasing availability of low-cost genetic testing
- **Pharmacogenetics**: Variation in genes that regulate drug effects
- Goal: Improve drug safety and efficacy
- Use genotype information to determine safe and effective drug/dose
- Can genotype-guided drug therapy improve patient care/outcomes?
- Evaluate within a randomized controlled trial

☆ Statistical design is complicated by knowledge that some participants may not respond to intervention due to their genetic makeup
Warfarin

- 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
- Several patient factors known to influence warfarin response
  - Clinical factors: Age, race, body size, other medications, ...
  - Genetic polymorphisms: CYP2C9, VKORC1, ...

B French (bcfrench@upenn.edu)
Personalized medicine interventions
59th WSC
COAG study

Clarification of Optimal Anticoagulation through Genetics (NCT00839657)

- **Objective**: Compare two approaches for warfarin therapy initiation based on algorithms that predict ultimately required stable dose
  - Genotype-guided dosing: Clinical information and relevant genotypes
  - Clinical-guided dosing: Only clinical information

- **Hypothesis**: By choosing an initiation dose that is more likely to be an individual’s ultimately required stable dose, the degree of improper anticoagulation that is common early in therapy can be reduced

- **Setting**: Multi-center, double-blind, randomized controlled trial

- **Outcome**: Percentage of time participants spend within therapeutic INR range (PTTR) during the first 4 weeks of therapy
  - Linear interpolation between INR values [Rosendaal et al., 2003]
  - Reasonably continuous and symmetric
  - Standard deviation 20–30%
PTTR calculation
Key design considerations

1. Targeted or untargeted design
2. Estimate of minimum detectable difference
3. Type-1 error rate for primary subgroup analysis
4. Planned interim analyses and monitoring

★ A circuitous route to a power calculation for a two-sample $t$ test
Targeted or untargeted design

- Genetic information could be used prior to randomization to identify participants who may be unresponsive to intervention [Simon, 2008]
- **Targeted design**: Study eligibility is restricted to participants who are predicted to be responsive based on their genetic characteristics
- Individuals with certain *CYP2C9* and *VKORC1* variants may not benefit from genotype-guided warfarin dosing [Anderson et al., 2007]
- By excluding potentially unresponsive participants, a targeted design may require a smaller sample size [Simon and Maitouram, 2004]
- Cost-benefit considerations to determine a practical design
  - Cost of genetic screening for eligibility
  - Cost of enrolling potentially unresponsive participants
Untargeted design

- Dose study: COAG study participants would receive warfarin therapy regardless of their CYP2C9 and VKORC1 variants
- If we excluded participants who may not benefit from genotype-guided dosing, we would be unable to evaluate our assumptions
  - 40% will possess a single genetic variant
  - Difference in PTTR between dosing groups will be 0%
- All participants are genotyped prior to randomization, so that much of the cost is already incurred in screening
- Including unresponsive participants enhances study generalizability
Minimum detectable difference

Accommodate differential effectiveness of genotype-guided dosing between groups defined by genetic variants (in \textit{CYP2C9} or \textit{VKORC1})

\begin{align*}
(1) \quad \text{PTTR}_G &= 0.4 \times 73\% \times 1 + 0.6 \times 61\% \times 1.15 = 71.3\% \\
(2) \quad \text{PTTR}_C &= 0.4 \times 73\% + 0.6 \times 61\% = 65.8\% \\
& \quad \text{Minimum detectable difference} = 5.5\% \\
\end{align*}

- Population proportion of 0.4 and 0.6 for 1 and 0, \( > 1 \) variants
- Mean PTTR of 73\% and 61\% for 1 and 0, \( > 1 \) variants
- Relative difference in mean PTTR between dosing groups of 0\% and 15\% for 1 and 0, \( > 1 \) variants
- Minimum detectable difference between dosing groups is 5.5\% [French et al., 2010]
Type-1 error rate

- **Primary subgroup**: Participants whose dose calculated from the genotype and clinical dose-initiation algorithms differs by \( \geq 1 \) mg
  - Known at randomization; not a post-randomization selection
  - Participants with larger expected differences in initial dose should have larger differences in PTTR between dosing groups
  - May drive the difference between dosing groups in full cohort

- How to allocate type-1 error rate \((\alpha)\) between the full cohort \((\alpha_A)\) and primary subgroup \((\alpha_S)\) analyses?
  - \(\alpha = 0.05, \alpha_A = 0.04\) fixed
  - \(\alpha_A + \alpha_S = \alpha\) may be conservative; Bonferroni-type adjustment
  - **Alpha allocation**: Consider the correlation between tests to inflate \(\alpha_S\) [Alosh and Hugue, 2009; Joo et al., 2010]
  - Based on final observed data; design used conservative adjustments
Alpha allocation

- $Y = \text{PTTR}$
- $A = \text{full cohort}; \ S = \text{primary subgroup}$
- $G = \text{genotype dosing group}; \ C = \text{clinical dosing group}$
- $p = \text{relative size of primary subgroup}$

\begin{align*}
Z_A &= \sqrt{\frac{n_{AG} n_{AC}}{n_{AG} \sigma_{AG}^2 + n_{AC} \sigma_{AC}^2}} \left( \bar{Y}_{AG} - \bar{Y}_{AC} \right) \\
Z_S &= \sqrt{\frac{n_{SG} n_{SC}}{n_{SG} \sigma_{SG}^2 + n_{SC} \sigma_{SC}^2}} \left( \bar{Y}_{SG} - \bar{Y}_{SC} \right) \\
\text{Cov}(Z_A, Z_S \mid H) &= \sqrt{\frac{p \sigma_{SG}^2}{\sigma_A^2}} = \sqrt{p \gamma}
\end{align*}

[Joo et al., 2010]
Alpha allocation

Under the null hypothesis $H$

(6) $\alpha = P(|Z_A| > z_{1-\alpha_A/2} \text{ or } |Z_S| > z_{1-\alpha_S/2} \mid H)$

$= \alpha_A + \alpha_S - P(|Z_A| > z_{1-\alpha_A/2} \text{ and } |Z_S| > z_{1-\alpha_S/2} \mid H)$

so that for fixed $\alpha$ and $\alpha_A$, $\alpha_S$ may be inflated

<table>
<thead>
<tr>
<th>$\alpha_A$</th>
<th>$p$</th>
<th>$\gamma$</th>
<th>$\alpha_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
<td>0.6</td>
<td>1.0</td>
<td>0.0200</td>
</tr>
<tr>
<td>1.1</td>
<td></td>
<td></td>
<td>0.0222</td>
</tr>
<tr>
<td>1.2</td>
<td></td>
<td></td>
<td>0.0245</td>
</tr>
<tr>
<td>1.3</td>
<td></td>
<td></td>
<td>0.0275</td>
</tr>
<tr>
<td>1.4</td>
<td></td>
<td></td>
<td>0.0313</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td></td>
<td>0.0363</td>
</tr>
</tbody>
</table>
Interim analyses and monitoring

- Internal pilot study [Wittes and Brittain, 1990]
  - Estimate PTTR standard deviation from portion of pre-planned sample
  - Possibly increase sample size based on estimated standard deviation
  - Restricted design: Not typically necessary to increase type-1 error rate
  - No change in sample size recommended by DSMB

- Conditional power analysis [Proschan and Hunsberger, 1995]
  - Requested by DSMB due to suboptimal recruitment rate
  - Probability of a significant difference at the conclusion of the study given current difference and assumed difference from future data
    - Null hypothesis (no difference)
    - Alternative hypothesis (minimum detectable difference)
    - Current difference
  - No change in sample size recommended by DSMB

- No interim monitoring for efficacy or formal stopping rules
COAG sample size

Sample size calculations

- Proportion with a single genetic variant 0.4–0.6
- $\alpha_A = 0.04$
- Drop-out rate of 10%

<table>
<thead>
<tr>
<th>Standard deviation of PTTR</th>
<th>20%</th>
<th>25%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Power</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Power</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Power</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ 0.4</td>
<td>5.5%</td>
<td>550</td>
<td>750</td>
</tr>
<tr>
<td>Δ 0.5</td>
<td>4.6%</td>
<td>792</td>
<td>1050</td>
</tr>
<tr>
<td>Δ 0.6</td>
<td>3.7%</td>
<td>1238</td>
<td>1642</td>
</tr>
</tbody>
</table>

Final sample size: 1022
**COAG sample size**

Sample size calculations

- Proportion with a single genetic variant 0.4–0.6
- $\alpha_A = 0.04$
- Drop-out rate of 10%

### Standard deviation of PTTR

<table>
<thead>
<tr>
<th>$\Delta$</th>
<th>20% Power</th>
<th>25% Power</th>
<th>30% Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>550 80%</td>
<td>750 90%</td>
<td>860 90%</td>
</tr>
<tr>
<td>0.5</td>
<td>792 80%</td>
<td>1050 25%</td>
<td>1238 25%</td>
</tr>
<tr>
<td>0.6</td>
<td>1238 90%</td>
<td>1642 90%</td>
<td>1932 90%</td>
</tr>
</tbody>
</table>

Final sample size: 1022
COAG sample size

Sample size calculations

- Proportion with a single genetic variant 0.4–0.6
- $\alpha_A = 0.04$
- Drop-out rate of 10%

<table>
<thead>
<tr>
<th>Standard deviation of PTTR</th>
<th>20% Power</th>
<th>25% Power</th>
<th>30% Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>80%</td>
<td>90%</td>
<td>80%</td>
</tr>
<tr>
<td>0.4</td>
<td>5.5%</td>
<td>550</td>
<td>750</td>
</tr>
<tr>
<td>0.5</td>
<td>4.6%</td>
<td>792</td>
<td>1050</td>
</tr>
<tr>
<td>0.6</td>
<td>3.7%</td>
<td><strong>1238</strong></td>
<td>1642</td>
</tr>
</tbody>
</table>

Final sample size: **1022**
Sensitivity analysis

Suppose that the group with 0, >1 variants includes participants who are unresponsive, perhaps due to no difference between predicted initial doses

\[
PTTR_G = 0.4 \times 73\% \times 1 + 0.6 \times 61\% \times [1 + (0.15 \times d)]
\]

such that the relative difference between groups is diluted by \((1 - d)\)

<table>
<thead>
<tr>
<th>(d)</th>
<th>(\Delta)</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>3.8%</td>
<td>65%</td>
</tr>
<tr>
<td>0.8</td>
<td>4.4%</td>
<td>77%</td>
</tr>
<tr>
<td>0.9</td>
<td>4.9%</td>
<td>86%</td>
</tr>
<tr>
<td>1.0</td>
<td>5.5%</td>
<td>93%</td>
</tr>
</tbody>
</table>

Preliminary data supports \(d = 0.91\) [IWPC, 2009]
Summary

Key design considerations in a personalized medicine intervention

1. Targeted or untargeted design
   - Exclude or include potentially unresponsive participants?
   - Cost-benefit of screening versus enrolling and generalizability

2. Estimate of minimum detectable difference
   - Population distribution of relevant allelic variants
   - Differential effectiveness of intervention across subpopulations

3. Type-1 error rate for primary subgroup analysis
   - Select subgroup most likely to benefit from intervention
   - Exploit correlation to inflate type-1 error rate: $\alpha_S > \alpha - \alpha_A$

4. Planned interim analyses and monitoring
   - Monitor key assumptions for sample size calculations?
   - Interim analyses and stopping rules?
References

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