A Decision Making Framework For Utilising External Control Arms

Basel Biometric Section – May 2019
Chris Harbron, MCO, Roche
A Randomised Clinical Trial

Randomisation ensures no selection bias and underpins validity of statistical comparisons.
A Naïve Real World Data Control Arm

In the absence of a randomised control arm, naïve comparison of recruited treated patients with a RWD control is likely to be biased due to differences between populations.
A Real World Data Control Arm With Propensity Scoring

Propensity scoring the RWD Control to the treated arm will reduce biases.

However a combination of unmeasured confounders and systematic differences may still leave population and/or measurement differences between the recruited treated patients and the RWD control arm.
Real World Data Control Arms

How much additional uncertainty would using a matched RWD control incur compared to using a randomised control?
Why Might a Randomised and a RWD Control Arm Differ?

**Study Specific Biases**

- **The process of being in a clinical trial**
  - Patient selection
  - Site selection
  - Higher levels of attention

- **Measurement biases**
  - More events and data captured in clinical trials
  - Variables measured in different ways
  - Some variables (e.g. ECOG) captured in clinical trials but not in clinical practice
  - Clinical trial data may be collected on a more regular basis

- **Unmeasured confounders – Unknown unknowns**
  - We see that absolute results in clinical trials often vary more than we may expect
  - An external control may also have unexpected differences
  - No absolute bound on the size of these differences

**Systematic Biases**
Idea – Surrogacy & Surrogate Threshold Effect

Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation, Burzykowski & Buyes, Pharmaceutical Statistics 2006;5;173–186

Note: RWD data shown in this presentation is artificial data
We May Expect Different Endpoints / Indications to Exhibit Different Levels of Variability

Potential Relationship For RCT & RWD Overall Survival

Potential Relationship For RCT & RWD PFS

May expect as the methodology for collecting an outcome is refined, or as the understanding of key covariates for an indication grows allowing improved propensity scoring models to be fitted, that the variability will decrease.
However, Surrogacy isn’t a Perfect Analogy to our Situation

<table>
<thead>
<tr>
<th>Classic Surrogacy</th>
<th>Comparing Control Arms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliberately different endpoints</td>
<td>Endpoints designed to be similar</td>
</tr>
<tr>
<td>Same patients</td>
<td>Different patients</td>
</tr>
<tr>
<td>Same treatments</td>
<td>But same patients in treated arm</td>
</tr>
<tr>
<td></td>
<td>Same treatments</td>
</tr>
</tbody>
</table>

**Plotting:** $T_{TRT} \nu T_{CTRL} \ \text{against} \ S_{TRT} \nu S_{CTRL}$

If T & S are unrelated, these will be unrelated

**Plotting:** $T_{TRT} \nu T_{CTRL\_RAND} \ \text{against} \ T_{TRT} \nu T_{CTRL\_RWD}$

The common term $T_{TRT}$ will induce a correlation

**Issue:** Most of the variability in such a plot, will represent variability in the performance of the novel treatment comparators. No reason to think the difference between control arms should be related to how good the treatment being compared to is. For our aim of understanding the relationship between the different flavours of controls this is a distraction.
For a More Rigorous Examination of Bias & Variability, Rotate Through 45% -> Bland-Altman Plot

Unlike surrogacy, as units on both axes are on the same scales, taking differences and averages makes sense

- Hazard Ratio v Randomised Control Arm
- Hazard Ratio v RWD Control Arm

- This is a nuisance parameter
- This is what we are interested in

\( T_T \text{ v } T_{C,RCT} \)

\( T_{C,RCT} \text{ v } T_{C,RWD} \)

\( T_T \text{ v } 0.5(T_{C,RCT} + T_{C,RWD}) \)
Proposal: Focus Directly on the Differences Between RCT & RWD

- **Treatment**
- **Control**
- **Propensity Weighted RWD Control**
- **RWD Control**

**Propensity score to weight the RWD Control to the treated arm**

**Compare the matched RWD control to the randomised control arm**

**Characterise this distribution over a range of studies**

![Graph](image)
Once Characterised : How Do We Use This Distribution?

Using the Approximation :

<table>
<thead>
<tr>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log(\text{HR}<em>{\text{Trt} v \text{Rand Cont}}) \approx \log(\text{HR}</em>{\text{Trt} v \text{RWD Cont}}) - \log(\text{HR}_{\text{RWD Cont} v \text{Rand Cont}})$</td>
</tr>
</tbody>
</table>

**Systematic Bias**

<table>
<thead>
<tr>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[\log(\text{HR}<em>{\text{Trt} v \text{Rand Cont}})] \approx E[\log(\text{HR}</em>{\text{Trt} v \text{RWD Cont}})] - E[\log(\text{HR}_{\text{RWD Cont} v \text{Rand Cont}})]$</td>
</tr>
</tbody>
</table>

**Study Specific Bias**

<table>
<thead>
<tr>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Var}[\log(\text{HR}<em>{\text{Trt} v \text{Rand Cont}})] \approx \text{Var}[\log(\text{HR}</em>{\text{Trt} v \text{RWD Cont}})] + \text{Var}[\log(\text{HR}_{\text{RWD Cont} v \text{Rand Cont}})]$</td>
</tr>
</tbody>
</table>

**Standard Analysis**

**Distribution of CT v RWD differences**
Comparing Back To Surrogacy

**Surrogacy**

Prediction Variance: \( \text{Var}(\beta + b_0 | \mu_{s0}, \alpha_0, \vartheta) \approx f\{\text{Var}(\hat{\alpha}_{s0}, \hat{\alpha}_0)\} + f\{\text{Var}(\hat{\vartheta})\} + (1 - R_{\text{trial}}^2)\text{Var}(b_0) \)

**External Controls**

\( \text{Var}[\text{Log}(HR \text{ Trt v Rand Cont})] \approx \text{Var}[\text{Log}(HR \text{ Trt v RWD Cont})] + \text{Var}[\text{Log}(HR \text{ RWD Cont v Rand Cont})] \)

Variability of assessed comparison
Residual differences between RCT & RWD Controls

No model, assumed constant
What Would be the Impact on the Analysis?

• To reflect the increased uncertainty:
  – A widening of confidence intervals, which will lead to:
    – a decrease in the alpha level required for statistical significance in the unadjusted analysis, or
    – an adjustment to the Minimal Detectable Difference, requiring a larger effect size

• A shift of the estimate of treatment effect to reflect any systematic biases:
  – To be conservative, propose an asymmetric approach
  – Don’t adjust if RWD control is on average a worse outcome than RCT control
  – Adjust if RWD control is on average a better outcome than RCT control
What Would be the Impact on the Analysis?

- Treatment vs RWD Control Comparison
- Additional Variability Associated With Non-Randomised Comparison
- Overall Assessment Of Treatment Effect
Considerations & Points For Discussion

• Variances are highly variable to estimate
  - Consider using $t$-distributions rather than normal to capture this uncertainty

• May have a non-normal distribution of comparisons
  - For example with one or two studies showing larger differences where matching has failed
  - Consider using mixture distributions to model the variability

• If the propensity scoring was to fail to deliver comparable populations in one study out of twenty, that is already spending our type 1 error
  - How many historical studies would we need to demonstrate the robustness of the relationship?

• Potential applications in internal and regulatory decision making
  - How may the required levels of evidence differ between these situations
Doing now what patients need next