Comparing ophthalmology treatments via the integration of IPD and aggregate-level data:

A new MAIC approach

Jonathan Alsop
Director
Numerus

PSI Annual Conference, 5th June 2018
ABSTRACT
Aim: To review and compare methods for indirect comparison of aflibercept and ranibizumab in patients with diabetic macular edema.

Methods: Post-stratification, inverse probability weighting based on simulated data, weight optimization, and regression model techniques were used to compare pooled individual patient-level data from the RESTORE and RESPOND (ranibizumab 0.5 mg as needed after 3 initial monthly doses) studies with summary-level data from the VIVID and VISTA (aflibercept 2.0 mg every 8 weeks after 5 initial monthly doses, 2q8) studies. The impact of adjusting for up to two baseline characteristics was assessed.

Results: All methods provided similar results. After adjustment for baseline best-corrected visual acuity and central retinal thickness, no statistically significant difference in average gain in baseline best-corrected visual acuity from baseline at month 12 was found between ranibizumab 0.5 mg and aflibercept 2q8.

Conclusions: Weight optimization and regression methods are useful options to adjust for more than one baseline characteristic.
Indirect treatment comparison

Individual Patient Data (IPD)

Aggregate-level Data (AGR)
A ‘baseline’ problem….

- Trial #1
  Individual Patient Data

≠

- Trial #2
  Aggregate Data
What is MAIC?

Matching Adjusted Indirect Comparison

Re-weighting of IPD

Weighted IPD baseline characteristics = AGR baseline characteristics

Weighted analysis of IPD outcomes
c.f. with AGR outcome data
Suppose $i^{th}$ patient has $a_i$ (baseline measure)

**Primary:** find weights $w_i$

\[
\sum_i w_i a_i = \bar{a}_{(2)} \\
\sum_i w_i = n_1 \\
w_i \geq 0
\]

**Target mean in AGR**

**IPD study size**

**Secondary:**

\[
w_i \leq w_{\text{max}}
\]

Restrict weights (avoid possible high leverage)

\[
\frac{(\sum_i w_i)^2}{\sum_i w_i^2}
\]

Maximise Effective Sample Size (ESS)
Methods

- Signorovitch et al. (2010)
- Post-stratification (Di Lorenzo et al., 2011)
- Entropy balancing (Hainmueller, 2012)
- MAICn (Han, 2014)
- Polynomial weighting (Regnier et al. 2015)

Borrowing from historic controls
Survey sampling
(e.g. cell weighting, raking)
Polynomial weighting
Polynomial weighting

Frequency

$\text{Weight}$

$a_i$
Example

Diabetic Macular Edema (DME)

Comparison of two Anti-VEGF treatments
## Studies

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Ranibizumab</th>
<th>Aflibercept</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data type</strong></td>
<td>IPD</td>
<td>AGR</td>
</tr>
<tr>
<td><strong>Sponsor</strong></td>
<td>Novartis</td>
<td>Regeneron</td>
</tr>
<tr>
<td><strong>Studies</strong></td>
<td>RESTORE</td>
<td>VIVID</td>
</tr>
<tr>
<td></td>
<td>RESPOND</td>
<td>VISTA</td>
</tr>
<tr>
<td><strong>Study type</strong></td>
<td>Blinded RCT</td>
<td></td>
</tr>
<tr>
<td><strong>Control arm</strong></td>
<td>Laser photoacoagulation therapy</td>
<td></td>
</tr>
<tr>
<td><strong>Endpoint</strong></td>
<td>$\Delta$ from Baseline to M12/LOCF in visual acuity (letters)</td>
<td></td>
</tr>
</tbody>
</table>
### Treatment confounders

<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>Aflibercept arm (N=286)</th>
<th>Ranibizumab arm (N=190)</th>
<th>Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCVA (letters)</td>
<td>59.1 (11.0)</td>
<td>64.1 (10.3)</td>
<td>-5.0 (-7.0, -3.0)</td>
</tr>
<tr>
<td>CRT (µm)</td>
<td>497 (151)</td>
<td>435 (126)</td>
<td>62 (36, 88)</td>
</tr>
</tbody>
</table>

What would have been effect of Ranibizumab in the VIVID/VISTA population?
Polynomial weighting

$j=\text{baseline parameter, } i=\text{patient}$

1) $a_{ij}$ standardised to lie on $[0,1]$: 
   \[ a_{ij}^* = \frac{a_{ij} - \text{min}(a_{ij})}{\text{max}(a_{ij}) - \text{min}(a_{ij})} \]

2) $w_{ij}^* = \left| \beta_{2j} + \beta_{3j}(a_{ij}^* - \beta_{1j}) + \beta_{4j}(a_{ij}^* - \beta_{1j})^2 + \beta_{5j}(a_{ij}^* - \beta_{1j})^3 + \beta_{6j}(a_{ij}^* - \beta_{1j})^4 \right|

3) $W_i = \frac{\Sigma_j w_{ij}^*}{\Sigma_{ij} w_{ij}^*}$

4) Calculate weighted mean & SD of $a_{ij}$ using $W_i \rightarrow \text{Mw}_j, \text{SDw}_j$

5) Minimise 
   \[ L = \Sigma_j \left[ \frac{(\text{Mw}_j - \text{Mt}_j)^2 + (\text{SDw}_j - \text{SD}_t_j)^2}{\text{SD}_t_j} \right] \] w.r.t. $\beta_{ij}$

$\text{Mt}_j, \text{SD}_t_j = \text{target mean & SD (AGR)}$
## Baseline Matching

<table>
<thead>
<tr>
<th>Source</th>
<th>Baseline BCVA (letters)</th>
<th>Baseline CRT (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Laser Mean (SD)</td>
<td>Anti-VEGF Mean (SD)</td>
</tr>
<tr>
<td>VIVID/ VISTA (target AGR)</td>
<td>60.2 (10.8)</td>
<td>59.1 (11.0)</td>
</tr>
<tr>
<td>RESTORE/ RESPOND*</td>
<td>59.8 (9.8)</td>
<td>61.5 (9.5)</td>
</tr>
<tr>
<td>RESTORE/ RESPOND (weighted IPD)</td>
<td>60.2 (10.8)</td>
<td>59.1 (11.0)</td>
</tr>
</tbody>
</table>

*Laser Mean (SD)* & *Anti-VEGF Mean (SD)*

*VIVID/VISTA inclusion/exclusion criteria applied.*
Weights

![Bar chart showing frequency of weights (wi) ranging from 0.00 to 5.50. The most frequent weight is around 0.50 with a frequency of 180, followed by 1.00 with a frequency of 60. Weights above 2.00 have lower frequencies.]
### Effect on BCVA Change to M12

<table>
<thead>
<tr>
<th>Source</th>
<th>Anti-VEGF vs. Laser Mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIVID/VISTA</td>
<td>10.0 (8.3, 11.7)</td>
</tr>
<tr>
<td>RESTORE/RESPOND (raw IPD)</td>
<td>7.6 (5.3, 9.8)</td>
</tr>
<tr>
<td>RESTORE/RESPOND (weighted IPD)</td>
<td>10.3 (7.9, 12.7)</td>
</tr>
</tbody>
</table>
SAS Code (brute force approach)

* Make multiple copies of original dataset;
proc surveyselect data=d1 out=d2 method=srs sampsize = 368 reps=50000 seed=1;
run;

* Simulate beta weighting parameters for each baseline variable;
data d3;
do Replicate=1 to 50000;
  a1=ranuni(0); b11=ranuni(0); b12=ranuni(0);b13=ranuni(0);b14=ranuni(0);*BCVA;
  a2=ranuni(0); b21=ranuni(0); b22=ranuni(0);b23=ranuni(0);b24=ranuni(0); *CRT;
  output;
end;
run;

* Calculate weight;
data d4;
merge d2 d3;
by Replicate;
t1=(BCVA-min1)/(max1-min1);
t2=(CRT-min2)/(max2-min2);
w1=abs(b11*(t1-a1)+b12*(t1-a1)**2+b13*(t1-a1)**3+b14*(t1-a1)**4);
w2=abs(b21*(t2-a2)+b22*(t2-a2)**2+b23*(t2-a2)**3+b24*(t2-a2)**4);
wo=w1+w2;* Overall weight;
run;
* Calculate weighted means and SDs;
proc means data=d5 noprint;
  var BCVA CRT;
  weight wo
by Replicate trt;
output out=ss2 mean=m1 m2 std=s1 s2;
run;

* Compare vs. AGR: target means / SDs, loss function;
data ss3;
  set ss2;
  if trt='Ranibizumab' then do;mean1=59.1;std1=11.0;mean2=497;std2=151; end;
  if trt='Laser' then do;mean1=60.2;std1=10.8;mean2=509;std2=153; end;
dm1=(m1-mean1)**2;
dm2=(m2-mean2)**2;
sm1=(s1-std1)**2;
sm2=(s2-std2)**2;
r1=sqrt(dm1)/std1; r2=sqrt(sm1)/std1;
r3=sqrt(dm2)/std2; r4=sqrt(sm2)/std2;
loss=r1+r2+r3+r4; * Loss function;
run;
Polynomial weighting vs. Signorovitch

Effective Sample Size:

N = 368
Signorovitch = 221
Polynomial = 229
Polynomial Weighting: Weaknesses

- Cannot match against what is not reported
- Lack of distribution overlap between IPD and AGR
- Weights are estimates (bootstrap CIs?)
- How to choose maximum weight?
- How to maximise ESS?
- Best trade off between matching AGR & max ESS?
Matching vs. Effective Sample Size

IPD: 1, 2, 6, 7

AGR target: mean=4, SD=2.94

How to weight data, i.e. choose $w_1, w_2, w_3, w_4$?

Simple(!) solution: $w_1=w_2=w_3=w_4=1$  
(ESS=4.0)

Suboptimal solution: $w_1=0.57, w_2=1.64, w_3=0.37, w_4=1.42$  
(ESS=3.1)
Polynomial Weighting: Benefits

- ‘Convergence’ appears to be good
- Allows maximum weights to be specified

Flexible (user-defined) loss function:

- weight some baseline characteristics more heavily than others
- weight some summary stats more others (e.g. mean > SD)
- match against multiple summary stats (e.g. mean, median, SD, proportions, IQR, other percentiles)
- could incorporate ESS maximisation
MAIC: Software

Signorovitch
- SAS: NLPNRA subroutine and GRD option in PROC IML
- R: Optim function

Malangone and Sherman
- SAS

Polynomial weighting
- SAS: Brute force, Proc IML?
- R: Optim function