Combining evidence: Purpose is everything

Stephen Senn
Acknowledgements

Many thanks for the invitation

This work is partly supported by the European Union’s 7th Framework Programme for research, technological development and demonstration under grant agreement no. 602552. “IDEAL”

Some of this work is joint with Artur Araujo & Steven Julious at Sheffield University
N-of-1 studies

- Studies in which patients are repeatedly randomised to treatment and control
- Increased efficiency because
  - Each patient acts as own control
  - More than one judgement of effect per patient
- However, only possible for chronic diseases
- Possible randomisation in $k$ cycles of treatment
- Implies $2^k$ possible sequences
A simulated example

• Twelve patients suffering from a chronic rare respiratory complaint
  • For example cystic fibrosis
• Each patient is randomised in three pairs of periods, comparing two treatments A and B
• Adequate washout is built in to the design
• Thus we have $12 \times 3 \times 2 = 72$ observations altogether
• Efficacy is measured using forced expiratory volume in one second ($\text{FEV}_1$) in ml
• How should we analyse such an experiment?
Trellis Plot of FEV per cycle grouped in patients

(c) Stephen Senn
Possible objectives of an analysis

• Is one of the treatments better?
  • Significance tests
• What can be said about the average effect in the patients that were studied?
  • Estimates, confidence intervals
• What can be said about the average effects in future patients?
• What can be said about the effect of a given patient in the trial?
• What can be said about a future patient not in the trial?
Two different philosophies

Randomisation philosophy
• The patients in a clinical trial are taken as fixed
• The population about which inference is made is all possible randomisations
• The patients don’t change, only the pattern of assignments of treatments change

Sampling philosophy
• The patients are regarded as a sample from some possible population of patients
• This is usually handled by adding error terms corresponding to various components of variance
• This approach is much more common
Is one of the treatments better?

Significance tests

Rothamsted School

• Leading statisticians such as Fisher, Yates, Nelder, Bailey

• Developed analysis of variance not in terms of linear models but in terms of symmetry

• High point was John Nelder’s theory of general balance (1965)

General Balance

1) Establish and define block structure
2) Establish and define treatment structure
3) Given randomisation the analysis then follows automatically

Here the block structure is Patient/Cycle GenStat®
Patient(Cycle) SAS®

The treatment structure is Treatment
The general balance approach

**BLOCKSTRUCTURE** Patient/Cycle

**TREATMENTSTRUCTURE** Treatment

ANOVA[FPROBABILITY=YES;NOMESSAGE=residual] Y

. 

**Analysis of variance**

Variate: FEV₁ (mL)

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>s.s.</th>
<th>m.s.</th>
<th>v.r.</th>
<th>F pr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient stratum</td>
<td>11</td>
<td>1458791.</td>
<td>132617.</td>
<td>10.04</td>
<td></td>
</tr>
<tr>
<td>Patient.Cycle stratum</td>
<td>24</td>
<td>316885.</td>
<td>13204.</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>Patient.Cycle.<em>Units</em> stratum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>641089.</td>
<td>641089.</td>
<td>50.57</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Residual</td>
<td>35</td>
<td>443736.</td>
<td>12678.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>2860501.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparing two models

The first is without a patient by treatment interaction

NB Analysis with proc glm of SAS®

The second is with a patient by treatment interaction*

* This is analogous to a fixed effects meta-analysis

---

**Table 1: Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type II SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>patient</td>
<td>11</td>
<td>1458791.444</td>
<td>132617.404</td>
<td>10.46</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>patient*cycle</td>
<td>24</td>
<td>316884.667</td>
<td>13203.528</td>
<td>1.04</td>
<td>0.4479</td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>641089.389</td>
<td>641089.389</td>
<td>50.57</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

---

**Table 2: Parameter Estimates**

| Parameter     | Estimate  | Standard Error | t Value | Pr > |t| |
|---------------|-----------|----------------|---------|------|---|
| mean effect   | 188.722222| 26.5394469     | 7.11    | <.0001 |
Two simple analyses

Based on 36 pairs (3 per patient)

Based on 12 averages (one per patient)

\[ 7.11^2 = 50.57 \]
Two more difficult questions

The average effects in future patients?
• This may require a mixed effects model
• Allow for a random treatment-by-patient interaction
  • The possibility that there may be variation in the effect from patient to patient
• Strong assumptions may be involved

The average effect for a given patient?
• The same random effect model can be used to predict long-term average effects for patients in the trial
• A weighted estimate is used whereby the patient’s only results are averaged with the general result
Any damn fool can analyse a clinical trial and frequently does
Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td>1772.67</td>
</tr>
<tr>
<td>Residual</td>
<td>23685</td>
</tr>
</tbody>
</table>

Fit Statistics

-2 Res Log Likelihood          457.7
AIC (Smaller is Better)        461.7
AICC (Smaller is Better)       462.1
BIC (Smaller is Better)        462.6

Solution for Fixed Effects

| Effect   | Estimate | Error  | DF | t Value | Pr > |t| |
|----------|----------|--------|----|---------|------|--|
| Intercept| 188.72   | 28.8838| 11 | 6.65    | <.0001|

Not how the analysis using proc mixed is identical as regards inference about the mean effect to the summary measures approach using one contrast per patient.
Morals

• Design is crucial
• Analysis depends on *purpose*
• And also on design and vice versa
• Results depend on philosophical framework
• Calculation is difficult, yes, but so is thinking
To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.

RA Fisher