FIRST SESSION

INTRODUCTION TO SUBGROUP ANALYSIS

Brian D M Tom
SUBGROUP ANALYSIS

• A secondary analysis that explores the possibility of differential treatment (intervention) effects within the study population

• That is, we are interested in determining whether a treatment effect (or lack of it) depends on certain baseline characteristics (e.g. demographic or clinical features) of patients

• For example, the treatment may have a strong positive effect for young patients and no effect (or a negative effect) for old patients
WHY DO IT?

• Heterogeneity – Patients recruited do not form a homogeneous sample

• Scientific and ethical obligation to do so

• Doctors don’t treat “average” patients, they treat individuals

• Prior suspicion or evidence
  • For example, effect of chemotherapy on the survival experience of lung cancer patients might differ depending on whether patients are smokers or non-smokers

• Hypothesis generation - to inform future studies

• However, subgroup analysis/analyses need to be approached with care!
DIFFICULTIES

- Not enough statistical power/false negatives
  - powered to detect overall treatment effect
- Multiplicity/false positives (spurious findings)
  - many baseline variables (e.g. ISIS-2 & astrological birth sign)
  - continuous variables (many ways to form subgroups)
- Data-dredging
  - not a priori/pre-specified
  - reporting only “interesting” findings
- Incorrectly done

• Illustrated in Matthews and Altman (BMJ, 1996)

• Investigating the effect of antenatal steroids for preventing neonatal respiratory distress syndrome (NRDS)

• In particular, interested in determining whether there was a differential effect of treatment on babies whose mothers did or did not develop pre-eclampsia
<table>
<thead>
<tr>
<th>NRDS</th>
<th>Pre-eclampsia</th>
<th>No pre-eclampsia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Steroid</td>
<td>Placebo</td>
</tr>
<tr>
<td>Proportion</td>
<td>21.2%</td>
<td>27.3%</td>
</tr>
<tr>
<td>Standard error</td>
<td>7.1%</td>
<td>7.8%</td>
</tr>
<tr>
<td>Number</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Difference in proportion</td>
<td>6.1%</td>
<td>6.2%</td>
</tr>
<tr>
<td>Standard error</td>
<td>10.5%</td>
<td>2.7%</td>
</tr>
<tr>
<td>P-value</td>
<td>0.57</td>
<td>0.021</td>
</tr>
</tbody>
</table>
HOW NOT TO DO IT!

• Compare P-values from separate analyses

• $P = 0.57$ vs $P = 0.02$

• Therefore there is *no evidence* of a treatment effect for mothers with pre-eclampsia, whilst there is *evidence* of a treatment effect in mothers without pre-eclampsia

• Incorrectly claim that this establishes a difference in the effect of steroids between mothers with pre-eclampsia and mothers without pre-eclampsia

• Not sufficient for the treatment effect to be significant in one subgroup and not in the other – no evidence $\neq$ no effect!
Note

- Treatment effects in the two subgroups are 6.1% & 6.2%
  - almost the same effect size
- Standard errors are 10.5% and 2.7%
  - no. of mothers with pre-eclampsia = 66
  - no. of mothers without pre-eclampsia = 529
• Statistically test for a treatment by covariate interaction

• That is, formally test the
  - Null hypothesis that the true treatment effects are equal in the two subgroups; versus
  - Alternative hypothesis that the true treatments effects are unequal in the two subgroups

• Test statistic or by constructing a confidence interval for the difference in treatment effects
  - Difference in treatment effects = 6.2 – 6.1 = 0.1%
  - Standard error for this difference = 10.9%
• Test statistic p-value of 0.99
• 95% CI of –21.3% to 21.5%
• Therefore no evidence of a statistically significant interaction
OTHER CONSIDERATIONS

• In the example, we defined the treatment effect in terms of difference in proportions. However, we could have equally defined the treatment effect in terms of the odds ratio

• Additive v multiplicative models – Different scales of measurement

• Possibility of getting a statistically significant interaction effect for one model and not for the other

• Discussion by Everitt and Smith (Psych Med, 1979) of the disagreement between Tennant & Bebbington (Psych Med, 1978) and Brown & Harris (Psych Med, 1978)
• Also, in many analyses, a continuous variable is dichotomised or categorised before investigating its interaction with treatment on the response

• Implications of doing this categorisation
  - Mismodelling the relationship
  - Bias, efficiency, coverage and power
CONCLUSIONS

• Subgroup analyses should be undertaken with care
• Preferably, pre-specified
• Should not, in general, place undue emphasis on these findings, which commonly lack statistical power
• A degree of scepticism is required, until results are confirmed through replication
• There are issues in the area of subgroup analysis still to be resolved fully


