Randomization and Mixture Approaches to Model Building

Nick Holford
Dept Pharmacology & Clinical Pharmacology
University of Auckland, New Zealand
Learn and Confirm Cycle

- Original idea from GE Box (1966)

- Translated to Drug Development

Confirming or Learning?

- **Confirming** tests the Yes/No Hypothesis

- If the question being asked has a **Yes/No** answer then it is a **Confirming** question

- If the question has a **How Much** answer then it is a **Learning** question
Confirming or Learning?

Confirming

• Making sure
• Outcome Expected
• Analysis Assumptions Minimized
  E.g. Randomized Treatment Assignment
• Questions for Drug Approval
  – E.g.
    • Does the drug work?
    • Can it be used safely in renal failure?

Learning

• Exploration
• Outcome Unexpected
• Assumption rich analysis
  – E.g. PKPD model
• Questions for Drug Science
  – E.g.
    • How big an effect does the drug have?
    • What is the clearance in renal failure?

Power

Bias & Imprecision
The Confirming Question

- The most common Clinical Trial Simulation question:
  - “Can the Null Hypothesis be rejected?”

- Frequentist Hypothesis testing requires
  - specification of rejection level
    - Alpha e.g. 0.05 [Type I Error criterion]
  - calculation of a test statistic
    - E.g. likelihood ratio
  - prediction of $P$ associated with test statistic

- How can $P$ be predicted?
The Randomization Test
What is the True P Value?

- A critical issue for population model based analysis
- Assumption that NONMEM OBJ Function Difference is Chi-square distributed may be dubious for confirming trial decisions
- The OBJ Function difference is the Likelihood Ratio Test statistic
- Randomization Test uses computational power to *estimate* the true P value for a given data set and analysis method
Randomization Test

・ Idea is to create a data set that would be expected if the Null Hypothesis was true
・ Simplest case uses a binary covariate e.g. treatment assignment
・ The original data has actual treatment assignment and outcome
・ New data set is identical except for re-randomization of covariate so that the Null Hypothesis will be true (under randomization)
Randomization Test

There are (at least) 3 methods that can be considered for the randomisation of a covariate:

1. Sample from a parametric distribution for the covariate (‘simulation’)
2. Sample with re-sampling from the covariate empirical distribution (like bootstrap)
3. Sample by permutation of the covariate empirical distribution (no resampling)
RT Assignment Methods

Parametric Simulation

```plaintext
if (uran >= 0.5)
   rancov=1
else
   rancov=0
```

Empirical ReSampling

```plaintext
isub=int(nsub*uran)+1
rancov=COV[isub]
```

Empirical Permutation

```plaintext
if (nsub>1) {
   isub=int(nsub*uran)+1
   rancov=COV[isub]
   # remove COV[isub]
   for (i=isub; i<nsub; i++)
      COV[i]=COV[i+1]
   nsub--
} else
   rancov=COV[1]
```
Theophylline and Sex

- Randomized concentration controlled trial of theophylline

- Simple Pharmacodynamic Model

- Does Sex affect Theophylline Emax?
Model for Sex

$PRED$

if (sex.eq.0) then ;female
   fsxemx=THETA(4)
else
   fsxemx=1
endif

E0= THETA(1)*EXP(ETA(1))
EMAX=fsxemx*THETA(2)*EXP(ETA(2))
EC50=THETA(3)*EXP(ETA(3))
Y = E0 + EMAX*THEO/(THEO+EC50) + ERR(1)
# Sex Change

## Original

<table>
<thead>
<tr>
<th>#</th>
<th>ID</th>
<th>TIME</th>
<th>THEO</th>
<th>AGE</th>
<th>WT</th>
<th>SEX</th>
<th>RACE</th>
<th>DIAG</th>
<th>PEFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.01</td>
<td>1</td>
<td>27</td>
<td>57</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.51</td>
<td>20.7</td>
<td>27</td>
<td>57</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>300</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>0.1</td>
<td>21</td>
<td>76</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>1.42</td>
<td>25.5</td>
<td>21</td>
<td>76</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>4.42</td>
<td>26.2</td>
<td>21</td>
<td>76</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>325</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.1</td>
<td>26</td>
<td>67.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>150</td>
</tr>
<tr>
<td>3</td>
<td>1.01</td>
<td>11.7</td>
<td>26</td>
<td>67.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>3.76</td>
<td>7</td>
<td>26</td>
<td>67.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>420</td>
</tr>
</tbody>
</table>

## Permuted

<table>
<thead>
<tr>
<th>#</th>
<th>ID</th>
<th>TIME</th>
<th>THEO</th>
<th>AGE</th>
<th>WT</th>
<th>SEX</th>
<th>RACE</th>
<th>DIAG</th>
<th>PEFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.01</td>
<td>1</td>
<td>27</td>
<td>57</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.51</td>
<td>20.7</td>
<td>27</td>
<td>57</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>300</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>0.1</td>
<td>21</td>
<td>76</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>1.42</td>
<td>25.5</td>
<td>21</td>
<td>76</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>4.42</td>
<td>26.2</td>
<td>21</td>
<td>76</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>325</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>0.1</td>
<td>26</td>
<td>67.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>150</td>
</tr>
<tr>
<td>3</td>
<td>1.01</td>
<td>11.7</td>
<td>26</td>
<td>67.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>3.76</td>
<td>7</td>
<td>26</td>
<td>67.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>420</td>
</tr>
</tbody>
</table>
NONMEM and RT

- Run Null and Alternate Models with Original Data
- Compute Original Data delta OBJ
  \[ dOBJ_{org} = \text{original null OBJ minus original alternate OBJ} \]
- Run Alternate Model with many (1000+) randomized data sets
- Compute Randomized Data delta OBJ
  \[ dOBJ = \text{original null OBJ minus randomized data set OBJ} \]
- Sort on \( dOBJ \) and find quantile corresponding to \( dOBJ_{org} \)
WFN nmrt

- Any model/data
  - Care with paths for user defined $SUB$
- WFN command:
  
  ```
  nmrt SEX theopdsex 1 1000
  ```
- SEX is an example of the covariate that will be permuted to generate null data sets
- Results in theopd.rt_SEX directory in theopd.txt
Null Distribution (FOCE)

-2.63 -1.53 -0.44 0.66 1.76 2.86 3.95 5.05 6.15 7.25 8.34 9.44 10.54

deltaOBJ

dOBJorg 5.608
P=0.024
### True P and Critical DOBJ Values

<table>
<thead>
<tr>
<th>Rep</th>
<th>Dobj</th>
<th>Obj</th>
<th>P Chisq</th>
<th>P Quant</th>
<th>Quant/ChisQ</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>890</td>
<td>5.625</td>
<td>5795.706</td>
<td>0.017706</td>
<td>0.023139</td>
<td>1.306831</td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>5.608</td>
<td>5795.723</td>
<td>0.017879</td>
<td>0.024145</td>
<td>1.350485</td>
<td>Reject</td>
</tr>
<tr>
<td>464</td>
<td>5.568</td>
<td>5795.763</td>
<td>0.018292</td>
<td>0.025151</td>
<td>1.374994</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>4.163</td>
<td>5797.168</td>
<td>0.041316</td>
<td>0.049296</td>
<td>1.193135</td>
<td></td>
</tr>
<tr>
<td>894</td>
<td>4.161</td>
<td>5797.17</td>
<td>0.041365</td>
<td>0.050302</td>
<td>1.216048</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>4.146</td>
<td>5797.185</td>
<td>0.041733</td>
<td>0.051308</td>
<td>1.229431</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>864</td>
<td>3.911</td>
<td>5797.42</td>
<td>0.047971</td>
<td>0.055332</td>
<td>1.153445</td>
<td></td>
</tr>
<tr>
<td>265</td>
<td>3.828</td>
<td>5797.503</td>
<td>0.050403</td>
<td>0.056338</td>
<td>1.11775</td>
<td></td>
</tr>
<tr>
<td>545</td>
<td>3.777</td>
<td>5797.554</td>
<td>0.051962</td>
<td>0.057344</td>
<td>1.103582</td>
<td></td>
</tr>
</tbody>
</table>

Original Data dOBJ org FOCE = 5.608
Theophylline Emax and Sex

http://wfn.sourceforge.net/wfnrt.htm
Theophylline Emax and Sex

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>Randomization Method</th>
<th>Successful Runs %</th>
<th>$\chi^2$ P</th>
<th>Quantile P</th>
<th>DOBJ(0.05,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOCE(^1)</td>
<td>Parametric</td>
<td>99.7</td>
<td>0.018</td>
<td>0.028</td>
<td>4.506</td>
</tr>
<tr>
<td>5.608</td>
<td>Non-Parm</td>
<td>99.7</td>
<td>0.018</td>
<td>0.033</td>
<td>4.570</td>
</tr>
<tr>
<td></td>
<td>Permutation</td>
<td>99.4</td>
<td>0.018</td>
<td>0.024</td>
<td>4.163</td>
</tr>
<tr>
<td>FO(^1)</td>
<td>Parametric</td>
<td>99.7</td>
<td>0.051</td>
<td>0.073</td>
<td>4.407</td>
</tr>
<tr>
<td>3.816</td>
<td>Non-Parm</td>
<td>99.7</td>
<td>0.051</td>
<td>0.080</td>
<td>4.741</td>
</tr>
<tr>
<td></td>
<td>Permutation</td>
<td>99.5</td>
<td>0.051</td>
<td>0.065</td>
<td>4.115</td>
</tr>
</tbody>
</table>

- **DOBJ** FOCE $\chi^2$ P closer to true P value?
- **Method** Permutation closer to $\chi^2$ P?

1=Compaq Visual Fortran 6.6 Update A

©NHG Holford, 2005, all rights reserved.
Randomization Test of SEX and WT

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Parameter</th>
<th>Obj</th>
<th>dOBJ</th>
<th>Randomization Test Probability</th>
<th>PPV² Emax</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>-</td>
<td>5801.3</td>
<td>5.599</td>
<td>0.024</td>
<td>0.27</td>
</tr>
<tr>
<td>Sex</td>
<td>Emax</td>
<td>5795.7</td>
<td></td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Weight⁴</td>
<td>Emax, E0</td>
<td>5813.5</td>
<td>0.235</td>
<td>0.635</td>
<td>0.43</td>
</tr>
<tr>
<td>Sex Weight⁴</td>
<td>Emax, Emax, E0</td>
<td>5813.3</td>
<td></td>
<td>0.43</td>
<td></td>
</tr>
</tbody>
</table>

1=NONMEM V 1.1 FOCE
2=Population Parameter Variability (sqrt(ω²))
3=Compaq Visual Fortran 6.6 Update C
4=Allometric model on Emax and E0

©NHG Holford, 2005, all rights reserved.
Likelihood Ratio Test Results

- Seems to be a smaller Emax in women
  Against biological expectation (no effect)

- Adding weight worsens objective function and PPV suggesting no effect of weight on PEFR
  Against biological expectation (increase with size)
Mixed up about Sex? Is there another Way?

- A Bayesian approach is to estimate the probability of one model being favoured over another model

- NONMEM can use a mixture model to estimate this probability

Thanks to Steve Duffull for suggesting this approach
Mixture on Models

$PRED$

IF (MIXNUM.EQ.1) THEN ; sex effect model
  IF (SEX.NE.1) THEN
    FSXM=THETA(4) ; female
  ELSE
    FSXM=1 ; male
  ENDIF
ELSE
  FSXM=1 ; no effect of sex
ENDIF

E0 = THETA(1)*EXP(ETA(1))
EMAX=FSXM*THETA(2)*EXP(ETA(2))
EC50=THETA(30*EXP(ETA(3))

Y = E0 + EMAX*THEO/(THEO+EC50) + EPS(1)

$MIX$

NSPOP=2
P(1)=THETA(5) ; Prob of sex effect model
P(2)=1-THETA(5) ; Prob of no sex effect
Mixture Model Method

- Specify both models in the same problem
- Estimate the mixture probability for each model
- Bootstrap the mixture model
- Examine distribution of mixture probability
Step 1: Probability of Weight on Emax and E0
1000 Permutations/Bootstraps

Null Distribution
Weight on Emax and E0
(Permuted on Weight)

Test Distribution
Weight on Emax and E0
Step 2: Probability of Sex on Emax
With and Without Weight on Emax and E0
1000 Permutations/Bootstraps

Null Distribution
With Weight on Emax and E0
(Permuted on Sex)

Test Distribution
With Weight on Emax and E0
## Fractional Change of Emax in Females
### 1000 Bootstraps

<table>
<thead>
<tr>
<th></th>
<th>Weight on Emax and E0</th>
<th>Without Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td>0.96</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td>0.72 – 1.24</td>
<td>0.06-1.01</td>
</tr>
</tbody>
</table>
Mixture Probability Results

- Distribution when weight is included is very different from the null when any weight effect has been removed by permutation

- No support for a sex effect
  *Confirms biological expectation*

- Suggests a weight effect on Emax and E0
  *Confirms biological expectation*
Statistics

- “Significance”
  - Count of Mixture Prob > 0.5
- “Weight of Evidence”
  - Average of Mixture Probability
- Odds
  - For = Weight/(1-Weight)
  - Against=(1-Weight)/Weight
## Mixture Model Odds

**SEX and WT**

<table>
<thead>
<tr>
<th></th>
<th>Mixture P Significance</th>
<th>Mixture P Weight</th>
<th>Odds For</th>
<th>Odds Against</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Weight</td>
<td>0.86</td>
<td>0.83</td>
<td>4.75</td>
<td>0.21</td>
</tr>
<tr>
<td>Weight on Emax</td>
<td>0.73</td>
<td>0.73</td>
<td>2.73</td>
<td>0.37</td>
</tr>
<tr>
<td>Weight on Emax and E0</td>
<td>0.45</td>
<td>0.52</td>
<td>1.07</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Conclusion

- Females appear to have a smaller theophylline Emax for PEFR.
- This difference is no longer supported when body size is used to explain between subject differences in Emax and baseline PEFR.
- Mixture models are an alternative to the randomization test for model building.