SOME STATISTICAL CONSIDERATIONS IN THE DESIGN AND ANALYSIS OF EQUIVALENCE STUDIES USING PHARMACODYNAMIC AND CLINICAL ENDPOINTS

Scott Haughie, M.Sc., Senior Director, Biostatistics
IPAC-RS/UF Orlando Inhalation Conference
Approaches in International Regulation
Development Considerations Session
Wednesday March 19th 2014
Aims

- This talk will compare and contrast different approaches to designing and analyzing equivalence studies, namely:
  - Pharmacodynamic (PD) endpoints (Dose scale analysis, Emax model, Dose selection, Linear regression, Relative potency, Relative bioavailability, Power and Sample size)
  - Clinical endpoints (Response scale analysis, Distributional properties, Assay sensitivity, Power and Sample size)
Disclaimer

- Any views expressed or implied by the contents of this presentation are based on my own thoughts.
- As such, any opinions given or communicated do not reflect a particular view or position on the part of Mylan, and are not attributable to Mylan.
Acknowledgments

- I would like to thank Dr. Jonathan Ward (VP and Global Clinical Development Leader for the Respiratory Group at Mylan) for his valuable input to this talk.
Development Context

- Regulatory guidance on requirements for demonstrating bioequivalence (BE) for generic orally inhaled products (OIPs) is a developing area

- Prior to 2009, no specific guidance in EU or US
  - A hindrance to development of generic OIPs
  - 1st draft CHMP (general OIP) guidance in 2009 (Final 2010)
  - 2 draft product-specific guidances from FDA (Office of Generics) in 2013

- Guidance generally describes 3 main areas where equivalence needs to be shown
  - In-vitro equivalence
  - In-vivo systemic bioequivalence via pharmacokinetics (PK)
  - In-vivo local equivalence via PD or clinical endpoints
EMEA/CHMP Approach to Local Equivalence

- EMEA only requires local equivalence studies in the event that PK BE has failed in some respect
  - If PK BE is successfully demonstrated, additional studies are not normally required for the EU
- EMEA recommends assessing Relative Potency (RP) which requires at least 2 doses of test and 2 doses of reference
  - Show dose response in both products
  - Dose scale analysis
  - Also compare results at each dose level (response scale analysis)
- 2009 OIP Guidance
FDA/OGD Approach to Local Equivalence

- FDA requires local equivalence to be demonstrated regardless of PK BE
  - Therefore the FDA requirements usually represent an additional hurdle after PK BE studies which may not be present for the EU
  - For this reason, the FDA requirements feature prominently in this talk

- FDA draft guidance for albuterol in April 2013 (revised June 2013)

- FDA draft guidance for Fluticasone / Salmeterol Combination (FSC) in September 2013
The 2 FDA draft Guidances are product-specific and recommend different approaches

Albuterol guidance recommends a dose scale analysis
- Show dose response (at least in reference product)
- Assess Relative Bioavailability (RBA) of test to reference product

FSC guidance recommends a response scale analysis
- No need to show a dose response

Main differences between the albuterol and FSC draft guidances seem to be based on a real or perceived ability to show a dose response in the reference product
# Summary of main differences between FDA draft Guidances for albuterol and FSC

<table>
<thead>
<tr>
<th>Albuterol MDI (April 2013)*</th>
<th>FSC DPI (Sept 2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suggests 2 alternative PD studies (1 based on PC$_{20}$, 1 based on FEV1 endpoints)</td>
<td>Single study with co-primary FEV1 endpoints</td>
</tr>
<tr>
<td>Crossover studies</td>
<td>Parallel group study</td>
</tr>
<tr>
<td>Limits of 67-150% applied to 90% CI</td>
<td>Limits of 80-125% applied to 90% CI</td>
</tr>
<tr>
<td>Dose scale analysis</td>
<td>Response scale analysis</td>
</tr>
<tr>
<td>Two doses of R (dose response), 1 of T</td>
<td>One dose each of T and R</td>
</tr>
<tr>
<td>Include placebo (build Emax model)</td>
<td>Include placebo (assay sensitivity)</td>
</tr>
<tr>
<td>Bootstrap to calculate CI</td>
<td>Not specified how CI calculated</td>
</tr>
</tbody>
</table>

*includes additional information from draft Orlistat guidance (2010)
Why are 2 different approaches necessary?

- Demonstrating dose response for an Inhaled Corticosteroid (ICS) for clinical endpoints (FEV1) has serious challenges
  - Fluticasone is the ICS component of FSC
- Biomarkers such as Exhaled Nitric Oxide (eno) have also been explored more recently
  - There are similar limitations with eno for demonstrating local equivalence
    - “Exhaled Nitric Oxide (eno) Versus Adenosine-5'-Monophosphate (amp) Challenge For Demonstrating Dose Response To Inhaled Corticosteroid” by Peter T. Daley-Yates, Daren J. Austin, Jane H. Bentley (ATS 2013)
    - Lee, “Regulatory Science for Generic Orally Inhaled Drug Products” (ISAM 2013)
Is a dose scale analysis OK for albuterol?

- From draft albuterol guidance
  - “Firms are encouraged to consider the conduct of a pilot study to refine the study design (e.g., inclusion and exclusion criteria) and estimate the study power based on intra- and inter-subject variability and slope of the Emax dose-response curve”
  - i.e. a pilot study should be designed to “show a dose response” in the reference product

- What does “show a dose response” mean?
  - The slope obtained has to be steep enough (relative to variability) for a definitive local equivalence study (including test and reference) to be feasible (in terms of patient numbers)
  - For ICS, sample size can be >1000 using eno (Daley-Yates et al)

- But is the situation really different for albuterol?
  - Or indeed salmeterol?
Mylan study for salmeterol component of FSC

- Based on methacholine challenge
- Endpoint: PC$_{20}$
  - Patients breathe a nebulised solution of methacholine for a set duration
  - Increasing concentrations of methacholine are administered until a 20% reduction in FEV$_1$ (aka PC$_{20}$) is achieved
- A single dose of a bronchodilator (e.g. β-receptor agonist) will increase the concentration of methacholine required to reduce FEV$_1$ by 20%.
  - Increasing doses of bronchodilator should further increase the PC$_{20}$, hence a dose response should be identifiable.
Study design

- Randomized, double-blind, placebo and active-controlled, 5-way crossover study
- N=40 asthmatics: screening FEV$_1$ ≥70%, PC$_{20}$ ≤8 mg/mL
- Five treatments: Advair® 100/50 & 200/100 µg; Albuterol 90 & 180 µg; placebo. Single dose.
- Endpoint: Methacholine PC$_{20}$ at 1, 6 and 10 h post-dose
- Centres: 5 Investigators in USA
- Study powered at 90% to detect a statistically significant slope between two doses of Advair® or Albuterol (assuming a doubling of effect).
albuterol log PC$_{20}$ (1 hour post dose)

Log PC$_{20} = -0.3$;
GM = 0.74 mg/mL

Log PC$_{20} = 0.98$;
GM = 2.66 mg/mL

3.59 fold increase from PBO
albuterol log PC$_{20}$ (1 hour post dose)

Log PC$_{20}$ = -0.3;
GM = 0.74 mg/mL

Log PC$_{20}$ = 0.98;
GM = 2.66 mg/mL

Log PC$_{20}$ = 1.23;
GM = 3.42 mg/mL

3.59 fold increase from PBO

1.29 fold increase btw doses
albuterol log PC$_{20}$ emax model (1 hr p.d.)

$E_{max} \text{ (Log PC}_{20}\text{)} = 1.86$

$E_0 \text{ (Log PC}_{20}\text{)} = -0.34$

$\text{ED}_{50} = 39 \text{ mcg}$
albuterol log $PC_{20}$ linear model (1 hr p.d.)

Log $PC_{20} = 0.97 + 0.0026$ (dose)

$\Delta$ Log $PC_{20} = 0.0026$ (\Delta dose) = 0.234

Fold change $PC_{20} = \exp(0.234) = 1.26$
Points to note

- ISAM poster (Kandala et al) indicates that optimum sample size is when the chosen test dose is equal to the \( ED_{50} \) of the reference dose response \( emax \) curve
  - 39 mcg is the \( ED_{50} \)
- Very little difference between \( emax \) model and linear model in 90 – 180 mcg range
- FDA albuterol draft guidance recommends doses 90 and 180 mcg albuterol
  - 90 mcg to be the dose of the test product
  - More than twice the \( ED_{50} \)
- Sample size based on Mylan data
  - Approx 300-400 patients depending on power (80% or 90%)
## Comparison to literature for albuterol (PC$_{20}$ Geometric means in mg/mL)

<table>
<thead>
<tr>
<th>Source</th>
<th>Time post dose</th>
<th>Placebo</th>
<th>albuterol 90 µg</th>
<th>albuterol 180 µg</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mylan study</td>
<td>1 h</td>
<td>0.74</td>
<td>2.66</td>
<td>3.42</td>
<td>1.29</td>
</tr>
<tr>
<td>Creticos</td>
<td>15 min</td>
<td>0.53</td>
<td>3.44</td>
<td>6.40</td>
<td>1.86</td>
</tr>
<tr>
<td>Ahrens (Spiros®)</td>
<td>15 min</td>
<td></td>
<td>7.50</td>
<td>10.15 (270 µg)</td>
<td>1.35</td>
</tr>
<tr>
<td>Ahrens (Spiros®)</td>
<td>15 min</td>
<td></td>
<td>7.50</td>
<td>9.60 Interpolated</td>
<td>1.28</td>
</tr>
<tr>
<td>Ahrens (Ventolin®)</td>
<td>15 min</td>
<td></td>
<td>5.55</td>
<td>12.27 (270 µg)</td>
<td>2.21</td>
</tr>
<tr>
<td>Ahrens (Ventolin®)</td>
<td>15 min</td>
<td></td>
<td>5.55</td>
<td>8.00 interpolated</td>
<td>1.44</td>
</tr>
<tr>
<td>Parameswaran (Proventil®)</td>
<td>10 min</td>
<td>1.7</td>
<td>14.8 (100 µg*)</td>
<td>26.1 (200 µg*)</td>
<td>1.76</td>
</tr>
<tr>
<td>Parameswaran (Ventolin®)</td>
<td>10 min</td>
<td>1.9</td>
<td>14.5 (100 µg*)</td>
<td>27.5 (200 µg*)</td>
<td>1.89</td>
</tr>
</tbody>
</table>

*salbutamol*
### Comparison to literature for salmeterol (PC$_{20}$ Geometric means in mg/mL)

<table>
<thead>
<tr>
<th>Source</th>
<th>Time post dose</th>
<th>Placebo</th>
<th>salmeterol 50 µg</th>
<th>salmeterol 100 µg</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mylan Study</td>
<td>1 h</td>
<td>0.74</td>
<td>2.76</td>
<td>3.69</td>
<td>1.34</td>
</tr>
<tr>
<td>Derom</td>
<td>1 h</td>
<td>3.8</td>
<td>14.4</td>
<td>24.0</td>
<td>1.67</td>
</tr>
<tr>
<td>Higham</td>
<td>30 min</td>
<td>0.8</td>
<td>3.1 (25 µg)</td>
<td>5.5</td>
<td>1.77</td>
</tr>
<tr>
<td>Higham</td>
<td>2 h</td>
<td>0.95</td>
<td>6 (25 µg)</td>
<td>8.8</td>
<td>1.47</td>
</tr>
<tr>
<td>Palmqvist</td>
<td>1 h</td>
<td>1.48</td>
<td>6.66</td>
<td>8.41 (250 µg)</td>
<td>1.26</td>
</tr>
</tbody>
</table>

*salbutamol*
Data suggests that dose scale analysis may not actually be feasible for albuterol using \( \text{PC}_{20} \)

- Similar issues as seen with ICS using eno may arise as pilot studies are attempted and data generated.

Data suggests that dose scale analysis could not be used for salmeterol either.

Issues applicable to FDA albuterol (using Emax and RBA) are also applicable to EMEA (using linear and RP).

Response scale analysis (as proposed with FSC guidance) may be more appropriate.

However, some questions remain on response scale.
Some questions on the FSC draft guidance

- FEV1 is usually considered to be Normally distributed and usually analyzed using methods which assume this property
  - For example: t-test, ANOVA, ANCOVA
  - Comparison between treatments usually analyzed using differences
- In the Sept 2103 draft Guidance, FDA propose analyzing test versus reference ratios using 80-125% equivalence criteria (90% confidence interval)
- Does this imply an analysis on the log scale?
Contrast PK BE with BE for clinical endpoint

- September 2013 FDA draft Guidance on PK BE:
  - “Equivalence based on: AUC and $C_{\text{max}}$ for fluticasone propionate and salmeterol. The 90% confidence intervals (CIs) for the geometric mean T/R ratios of AUC and $C_{\text{max}}$ should fall within the limits of 80.00–125.00

- September 2013 FDA draft Guidance on BE for CE:
  - “Equivalence based on: T/R ratio for the primary endpoints. The 90% CIs for the T/R ratios for the primary endpoints should fall within the limits of 80.00-125.00%.”

- Spot the difference: no mention of Geometric Mean!
Should limits be symmetric?

- For Normally distributed variables, the difference is also Normally distributed.
- Not the case for the ratio of two Normals.

**Diagram: Distribution of the Ratio of two Normals**

- 10000 studies
- 2 groups
- N=100 per group
- True ratio=1
- CV=92%
- Min=0.56
- Max=1.68
- 5%=0.8
- 95%=1.24
Assay sensitivity

- Very important in equivalence studies
- "Assay sensitivity is a property of a clinical trial defined as the ability to distinguish an effective treatment from a less effective or ineffective treatment" – ICH E10
- From Sept 2013 draft FDA Guidance:
  - “To ensure adequate study sensitivity, the T and R products should both be statistically superior to placebo (p<0.05) with regard to the BE study primary endpoints”
- Should assay sensitivity be assessed using T – P and R – P differences?
Sample size and power

- Approximate formula exists for sample size based on ratio of two means [depends on coefficient of variation (CV)]
- Can also be performed using simulations
  - Advantages
    - Apply actual analysis method to simulated data
    - Where there is a choice of analysis method, you can compare performance
- Two main methods for calculating variance of a ratio
  - Fieller’s theorem (also used for linear case of RP and RBA)
  - Delta method
- Boostrapping was recommended for PD endpoint and dose scale analysis
  - Is this the case for clinical endpoint?
Assessing variability

- The proposed BE for clinical endpoint study does require a large sample size
  - But a thorough literature search is required to identify studies most like that proposed

- Large range of CVs and effect sizes which depend on
  - Date (90s versus 00s)
  - Run-in type (ICS / placebo)
  - Population (asthma severity)
  - Combination / mono products

- For PK and PD BE, we usually assume true T/R ratio is 0.95 (or 1.05)
  - What assumption should be made about true T/R ratio for clinical endpoint?
Overall Conclusions

- Dose scale analysis for local equivalence
  - Serious limitations using eno for ICS
  - This has been acknowledged in current draft FDA guidance for FSC which proposes a response scale analysis
  - Feasibility may also be an issue for albuterol despite current draft FDA guidance recommending it
  - Similar issues arise for salmeterol, so this also supports the response scale approach for FSC

- Response scale analysis for local equivalence
  - Should be more achievable but large sample size still likely
  - Questions of detail need to be understood in order to design studies and allow prospectively planned analyses to be carried out

- Guidance needs to be achievable
  - Otherwise will be as big a block to generic OIP development as no guidance