

How to Determine Whether Two Behavioral Interventions are Equivalent

*Application of Experience
from the Pharmaceutical Industry*

Brian L. Wiens

Amgen Inc.

Thousand Oaks, California

Introduction

- Methods developed in context of pharmaceutical industry
 - Heavily regulated
- Methods should be applicable in behavioral sciences with minimal modification
- Statistical methods are "easy"
- Other issues are difficult

Motivation

Why would I want to do this?

- Compare a new behavioral intervention to a standard. It may be:
 - Easier
 - Cheaper
 - Quicker
 - Another language
 - A different population

Background

- For pharmaceuticals, the gold standard is comparison versus placebo (no intervention)
- When placebo cannot be used, still must show that new treatment is better than placebo
- Solution: indirect comparison versus active control (treatment with known benefit)

Definitions

- Noninferiority
 - The test treatment is not markedly worse than the control treatment (may be better)
- Equivalence
 - The two treatments are not markedly different
- NB: Simply ruling out a statistically significant difference is not sufficient
 - $p > 0.05$ does not imply equality

Statistical Methods

- Equivalence
 - $H_0: |\mu_1 - \mu_2| \geq \delta_0$
 - $H_1: |\mu_1 - \mu_2| < \delta_0$
- Rejecting the null hypothesis is required to make the desired conclusion
- Blackwelder, 1982

Statistical Methods (cont.)

- $z = \frac{(|\bar{x}_{\text{control}} - \bar{x}_{\text{test}}|) - \delta_0}{\text{se}(\bar{x}_{\text{control}} - \bar{x}_{\text{test}})}$ is asymptotically normal under mild regularity conditions
- Or, compare upper bound of confidence interval to δ_0 and lower bound to $-\delta_0$
- For noninferiority, drop the absolute value signs

Statistical Methods (Example)

- Proportion of subjects who "benefit"
- $\delta_0 = 0.10$ for noninferiority
- Observed proportions:
 - Control group: 68/100 benefit
 - Test group: 65/100
- 95% CI: 0.030 ± 0.185 or $(-0.155, 0.215)$
- Upper bound exceeds $\delta_0 \Rightarrow$ not noninferior

Difficult Issues

- Choosing a margin
 - Some unfortunate terminology*
 - Assay sensitivity
 - Putative placebo effect
 - Constancy assumption
- Validity of historical control

Choosing a Margin

- Much has been written
 - Little has been decided
- Start with noninferiority
 - SJ Wang
 - HM J Hung
 - R Temple
 - S Ellenberg

Choosing a Margin (cont.)

- Must rule out placebo
 - δ_0 must be smaller than difference between control and placebo/no intervention
 - "Putative placebo" estimated from historical comparison of control to placebo
 - "Constancy assumption": effect of control versus placebo is same in all studies
 - "Assay sensitivity": placebo would not be shown noninferior/equivalent to control

Choosing a Margin (cont.)

- Basic framework
 - Active control compared to placebo/no intervention in multiple trials
 - Mean treatment effect and variability are considered
 - Choose margin (δ_0) such that it is smaller than treatment effect in most/all historical comparisons

Choosing a Margin (cont)

- Putative placebo may be insufficient
- If margin is substantial, a "noninferior" treatment could be substantially worse
- Consider implications of the difference
 - E.g., mortality: reduce putative placebo-based margin by 50%
- Do not increase putative placebo-based margin, only decrease it

Choosing a Margin (cont)

- For equivalence follow same basic steps
- If neither treatment is accepted, then putative placebo is not calculable
- If one treatment is accepted, upper bound may be different from lower bound (putative placebo in one direction only)
- If both treatments have established effect, putative placebo not relevant

Validity of Historical Control

- Historical comparison of active treatment to placebo must be relevant
 - Same inclusion and exclusion criteria, evaluation methods, study population, etc.
- May not be available due to changes
 - Population is new or changed
 - New supportive treatments are available

Other Issues

- Any statistical issue must be addressed for noninferiority/equivalence studies
 - Multiple comparisons
 - Missing data
 - Subgroup analyses
 - Model mis-specification
 - Etc.

Summary

- Statistical methods are straightforward
- Other issues are difficult
- Experience from pharmaceutical industry may be instructive