# Non-Inferiority Clinical Trials to Establish Effectiveness*

Dr. Mark Rothmann
U.S. Food and Drug Administration
CDER

24th Annual Biopharmaceutical Applied Statistics Symposium
October 25, 2017

# Outline

- Introduction to Non-inferiority Trials – FDA Guidance
  - What is a Non-inferiority Trial?
  - The Non-inferiority Margin
  - Statistical Inference
- Missing Data and Non-inferiority Trials
  - Historical Studies
  - Estimand
  - Prevention of Missing Data

# Introduction to Non-inferiority Trials – FDA Guidance

# Reference

- *Guidance for Industry: Non-Inferiority Clinical Trials to Establish Effectiveness,* U.S. FDA CDER/CBER, November 2016

# What is a Non-Inferiority Trial?

- An active-controlled trial where the intent is to show that a test product is not worse than the control product by a pre-specified amount on a meaningful endpoint

# Assay Sensitivity

- A demonstration a test product is superior to a placebo or active control is entirely interpretable without further assumptions (other than lack of bias)
- A conclusion of non-inferiority (NI) depends on knowing the active control had its expected effect in the NI study. This is called 'assay sensitivity.'
  - Showing a small difference between treatments, may mean that the products are similarly effective or similarly ineffective.
- Without a placebo arm, assessing assay sensitivity relies on external information, giving NI studies similar characteristics of a historically controlled trial.

# Superiority Hypothesis

- Placebo-controlled trial

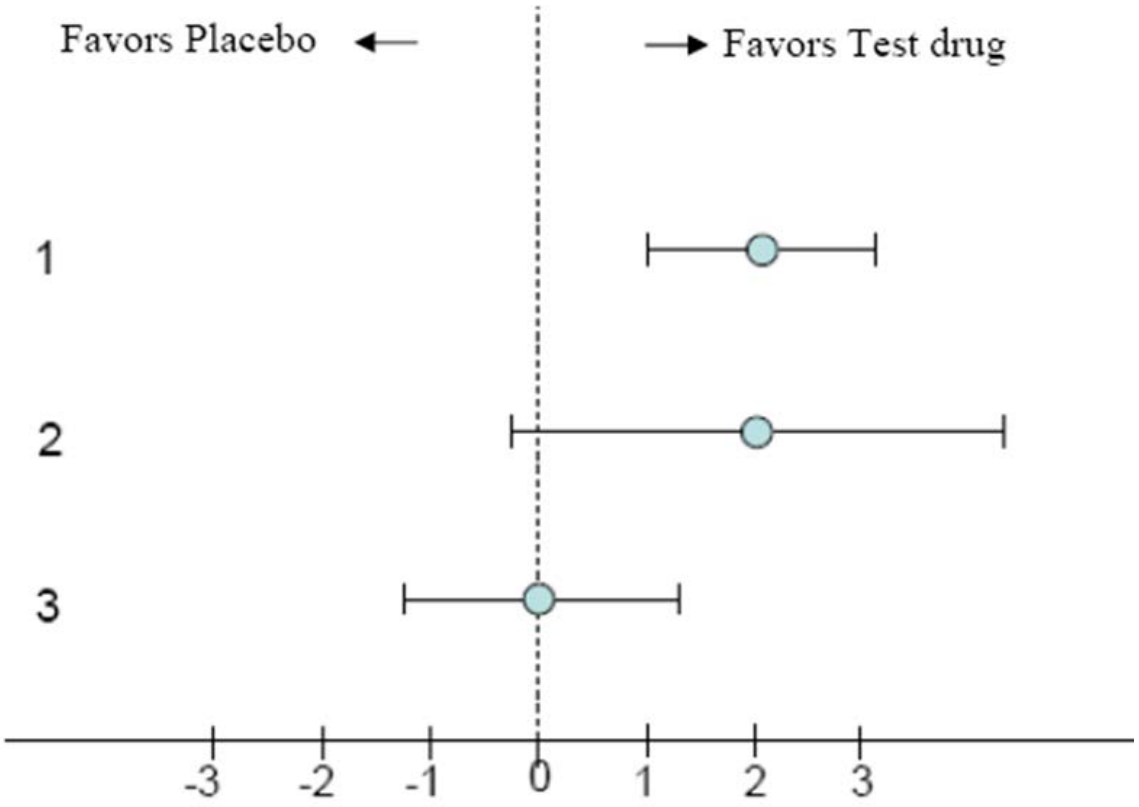
$$H_o: T \leq P; \quad T - P \leq 0$$

$$H_a: T > P; \quad T - P > 0$$

T: Test Product

P: Placebo

# Possible Results
# Point Estimate and 95% CI



# Goal of a Non-inferiority test

- show effect of the test drug (T) is not inferior to the effect of the active control (C) by a specified amount, called the non-inferiority margin, or M.

# Non-inferiority Hypotheses

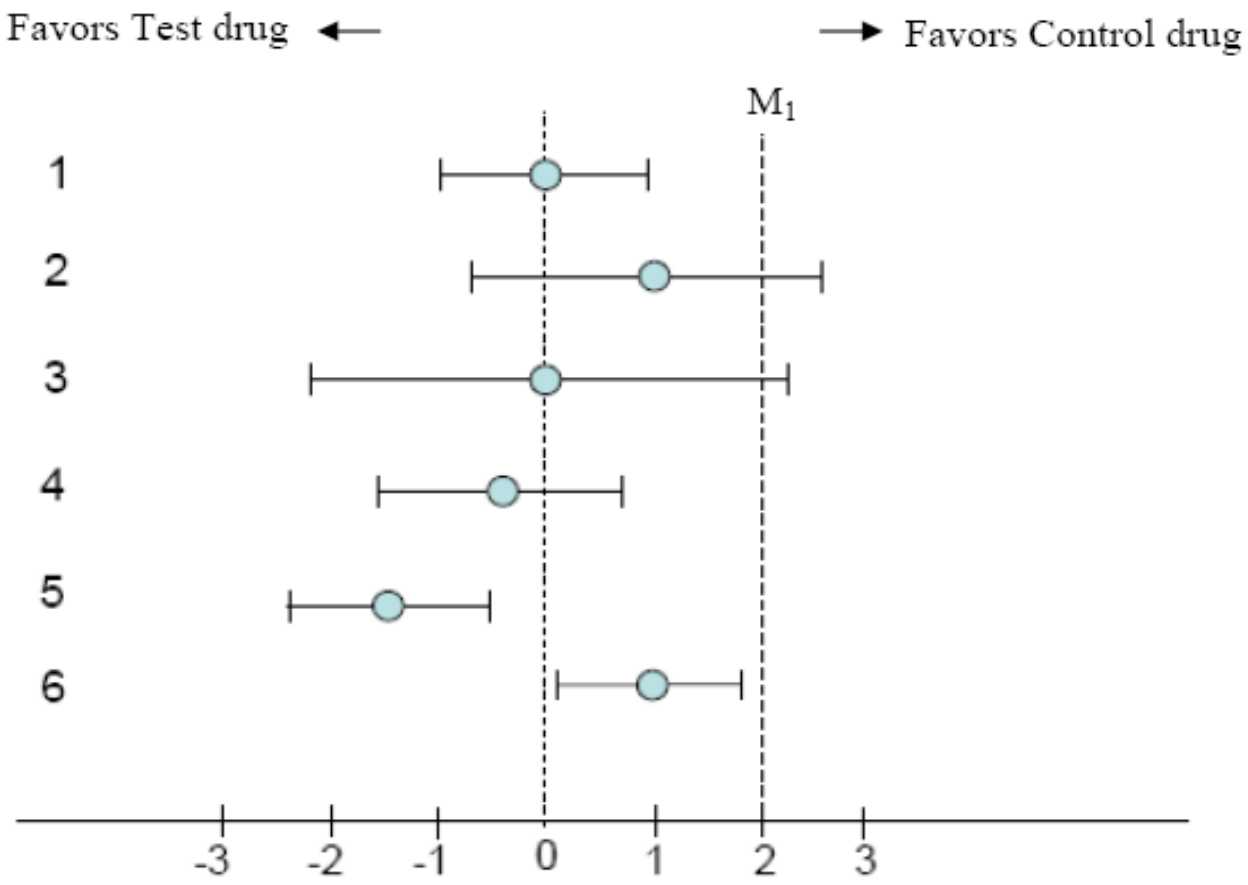
$H_o$: $C - T \geq M$ (T inferior to C by M or more)

$H_a$: $C - T < M$ (T inferior to C by less than M)

# Choices for M

- Set it equal to entire known effect of active control relative to placebo, based on past randomized controlled trials
- With this choice for M, called $M_1$, and assuming control drug attains this level of efficacy in the NI study, a finding of non-inferiority means that the test drug has an effect greater than 0.
- A more usual choice is to set M equal to a value smaller than $M_1$, called $M_2$, for which the test product is not unacceptably worse than the active control, which may be based on clinical judgment.

# Possible Results with margin $M_1$ Point Estimate and 95% CI (C-T)



# Reasons for Using a Non-Inferiority Design

- An available treatment provides an important benefit (e.g., prevents serious harm, such as death or irreversible morbidity)
- Comparative effectiveness is truly needed to understand risk-benefit

# Policy statement

In certain circumstances, however, it may be important to consider whether a new product is less effective than available alternative therapies, when less effectiveness could present a danger to the patient or to the public. For example, it is essential for public health protection that a new therapy be as effective as alternatives that are already approved for marketing when:

1. the disease to be treated is life-threatening or capable of causing irreversible morbidity (e.g., stroke or heart attack); or

2. the disease to be treated is a contagious illness that poses serious consequences to the health of others (e.g., sexually transmitted disease).

- April 1995, President Clinton and Vice President Gore [Reinventing Regulation of Drugs and Medical Devices, part of the National Performance Review]

# The Non-Inferiority Margin

- The definitions used to describe these two versions of M are:

$M_1$ = the entire effect of the active control assumed to be present in the NI study

$M_2$ = the largest clinically acceptable difference (degree of inferiority) of the test drug compared to the active control

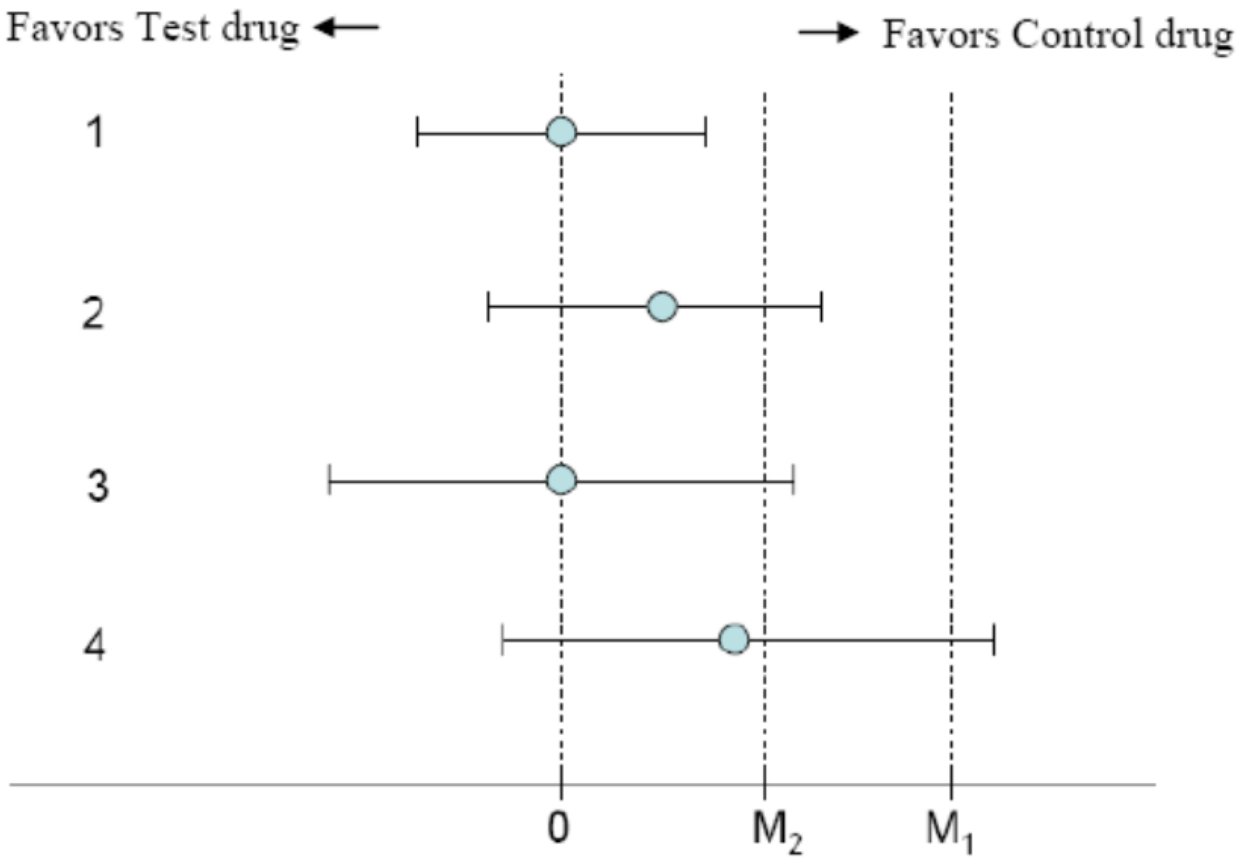
# The Non-Inferiority Margin

- M1
  - estimated based on historical experience of the active control drug. Its relevance to the current NI trial is based on
    - (1) assessment that the effect of the active control in the non-inferiority trial is similar to that estimated in the past (the constancy assumption), and
    - (2) assessment of the quality of the NI trial.
      - Are there aspects that could reduce a difference between the active control and the new drug ("bias toward no difference" which is a "bias toward the alternative").

# The Non-Inferiority Margin

- The choice of M2 is a matter of clinical judgment
  - M2 can never be greater than M1

# Possible Results of an NI Study Showing Active Control – Test Drug Differences Point Estimate and 95% CI



# Assay Sensitivity and Choice of $M_1$

- Assay sensitivity is the ability of the trial to have detected a difference between treatments of a specified size.
- Assay sensitivity means that had the study included a placebo, the underlying effect of the control would have been at least M1.

# Assay Sensitivity and Choice of $M_1$

- The choice of M1, and the conclusion that a trial has assay sensitivity (i.e., the active control would have had an effect of at least M1), is based on
  1) historical evidence of sensitivity to drug effects
  2) similarity of the new NI trial to the historical trials (the constancy assumption)
  3) the quality of the new trial (ruling out defects that would tend to minimize differences between treatments).

# Historical evidence of sensitivity to drug effects (HESDE) (ICH E-10)

- Appropriately designed, conducted trials in the past that used the control treatment (or one or more pharmacologically similar products) regularly showed treatment superior to placebo (or some other reference).
- The estimate of the size of the effect must take the variability of past results into account

# Similarity of NI trial to historical studies and "constancy assumption"

- NI study is sufficiently similar to the past studies with respect to all important study design and conduct features that might influence the active control effect. For example, similar in
  - characteristics of the patient population (e.g., entry criteria)
  - important concomitant treatments
  - definitions and ascertainment of study endpoints
  - dose of active control
  - analytic approaches

# Quality of the Non-inferiority trial

- With NI studies, it is believed that sloppy study conduct may introduce a bias towards no difference (which is in the alternative hypothesis of non-inferiority)
- Addressing Missing Data

# Statistical Inference

- Fixed margin method (95%-95% method)
- Synthesis method
- Neither or other

# Fixed margin method (95%-95% method)

- The 95% confidence interval of the estimated "average" effect of the control based on the historical studies and
- The 95% confidence interval for C-T from the NI study.

# Example

- The fixed margin approach on 35-day mortality rates applied to a new thrombolytic product, reteplase, for treatment of acute myocardial infarction.
- Meta-analysis of the results from available placebo controlled trials of streptokinase, the active comparator (control) for the NI study,
  - Point estimate for the effect on 35-day mortality = 2.6% with a 95% lower bound of 2.1% (i.e., $M_1$).
- Clinical decision: new thrombolytic should rule out a loss of more than half of the benefit of streptokinase to be regarded as an acceptable alternative.
  - $M_2$ = 2.1%/2 = 1.05%
- NI analysis: does the 95% CI (one sided for this particular case) of the difference in mortality rates exclude an increase of 1.05%?
- The INJECT study accomplished this, and the product was approved for marketing.
  - Upper limit of 0.71% <1.05%

# Synthesis Method

- Synthesizes data from historical trials and current NI trial, reflecting the variability in both data sources.
  For M = 0, synthesis test statistic is

$$\frac{\left(\widehat{C_H - P_H}\right) + \left(\widehat{T_N - C_N}\right)}{\sqrt{SE_H^2 + SE_N^2}}$$

- Where *H* signifies from historical studies and *N* signifies from NI trial. $SE_H$ and $SE_N$ are the standard errors from historical studies and NI study. Reject null hypothesis for large values of the test statistic

- Although 95%-95% method is mathematically equivalent to comparing the test statistic

$$\frac{\left(\widehat{C_H - P_H}\right) + \left(\widehat{T_N - C_N}\right)}{SE_H + SE_N}$$

  to the 97.5$^{th}$ percentile of a normal or other distribution

- Second formula has a larger denominator. Compared to synthesis method, 95%-95% method is conservative. Both methods rely crucially on the constancy assumption

# Selection of the Active Control

- Often selected as the product that has the largest observed treatment effect
  - Issue that which is observed the best tends to overstate its truth (i.e., represents a “random high”)

# Evaluating Effect of Active Control – Possible Methods

- Issue: Dealing with random highs
- Fixed Effects Meta-Analysis (or some other weighted analysis)
  - Weights based on common/identical effect
  - Include all products in the same pharmacologic class
- Hierarchical Modeling
  - Assume exchangeable effects

# Missing Data and Non-inferiority Trials

# Missing Data – Historical Studies Used

- Amount and Nature of Missing Data
  - No control over this
  - May not be known
- Addressing Missing Data
  - Try to account/address in the evaluation of the effect of the active control
    - Ex: Weight Management

# Estimand

- Most appropriate endpoint as the primary endpoint
- Intention-to-Treat (ITT)
  - All measurements regardless of adherence
- NI margin (evaluation of the effect of the active control based on the selected estimand)

# Prevention of Missing Data

- References
- Subjects
- Investigators
- Protocol

# References


- National Research Council. The prevention and treatment of missing data in clinical trials. National Academies Press, Washington, 2010
- Fleming, TR. Addressing missing data in clinical trials. *Ann Intern Med* 2011;154:113-117.

# Subjects

- Limit burden and inconvenience of data collection
- Increase incentives for participation and completion
- Educate subjects during the informed consent process on the scientific importance of their data even if they discontinue treatment, as well as the detrimental effect missing data have on the confidence and interpretation of the results

# Ex: Influence of Site Investigators

- A Particular Clinical Trial
  - A site with over 30 subjects had every subject measured for the primary endpoint
  - A site with over 25 subjects had only 3 subjects measured for the primary endpoint

# Investigators (1 of 2)

- Select investigators with good track record for enrolling and following subjects for complete data.
- Educate investigators in the necessity to maximize data capture.
- Signed investigator agreements on their commitment to continue follow-up efforts even after subjects discontinued treatment and initiate other interventions.

# Investigators (2 of 2)

- Tie payment of investigators to follow-up.
- Monitor data collection during the trial.
  - Poorly performing sites can receive further training, site visits or even site closure.

# Ex: Influence of investigators – T2DM

- For HbA1c change from baseline to 6 months in T2DM
  - Missing data varies from 4% to 30%
- Reduction in missing data over time across studies involving the same experimental therapy
  - Two studies with approx. 5% missing data

# Protocol (1 of 2)

- Distinction between reasons for nonadherence versus reasons for non-retention.
- “Withdrawal of consent” should be properly used
  - only when a subject no longer wishes to participate in the trial and no longer authorizes further collection of their outcome data.

# Protocol (2 of 2)

- Recognize the negative effect of substantial levels of missing data cannot be addressed by increasing sample size.
- Provide performance standards that should be met to achieve high quality of trial conduct, including targeted levels of data capture.

# Addressing Missing Data in the Analysis (1 of 2)

- Method for addressing missing data in the analysis should not encourage missing data
- Some methods of imputation fall under the alternative hypothesis
  - E.g., BOCF
- For repeated measures, the NI margin only applies to the landmark/time of interest
  - Problems with LOCF, BOCF

# Addressing Missing Data in the Analysis (2 of 2)

- Lack of a placebo arm may prevent some types of ways of addressing missing data
- May not be possible to use the same or analogous way to address missing data as the historical studies
- Impute based on a model considering "wash out" – "return to baseline"
  - Would have some of the concerns of BOCF
- Imputation under the NI null hypothesis (may be appropriate sometimes and sometimes not)

# Ex: Address MD in the analysis

- For HbA1c change from baseline to 6 months in T2DM
  - Have had an active-controlled study where the control had 12% missing data and experimental has 27% missing data

Thank you

Questions?