PhRMA PISC Expert Team White Paper: Toward a consistent standard of evidence when evaluating the efficacy of an experimental treatment from a randomized, active-controlled Trial

Patrick Peterson, Ph.D.
Eli Lilly and Company

BASS XV (Nov. 3, 2008)

PISC Team (Non-Inferiority / Active-Controlled Trials)

- Steven Snapinn (Amgen) << Chair >>
- Kevin Carroll (Astra Zeneca)
- Christy Chuang-Stein (Pfizer)
- Yu-Yun Ho (Novartis)
- Qi Jiang (Amgen)
- Gang Li (J&J)
- Patrick Peterson (Lilly)
- Yong-Cheng Wang (J&J)
- Matilde Sanchez (Arenapharm)
- Rick Sax (Astra Zeneca)

Outline

- Demonstrating efficacy in an active controlled (AC) trial (We did not consider safety-related issues of AC trials)
- Background & review of non-inferiority (NI) methods
  - Issues with fixed-margin approaches and inconsistent standards of evidence
- The case for “one standard of evidence” and the use of the synthesis method to apply this standard
- Logical inconsistencies if treating NI differently
- Key assumptions of the synthesis method
  - Adjusting for possible violations
- Conclusions / Q&A / Discussion

Focus of our Team's Efforts:

- Demonstrating an experimental treatment (T) is sufficiently effective for regulatory approval
- We assume the following for purposes of our inquiry:
  - Pivotal trial compares T to an active control treatment (C)
  - Inclusion of a placebo arm is either unethical or impractical
  - We will be able to demonstrate that T is a sufficiently safe treatment for the intended indication

Active Controlled (AC) Trial without Placebo Arm

- Possible outcomes of our AC trial:

1) T statistically superior to C
2) C statistically superior to T
3) No statistical difference observed between T and C

Although (1) is the most desirable outcome, a prudent researcher should plan for (2) and/or (3).

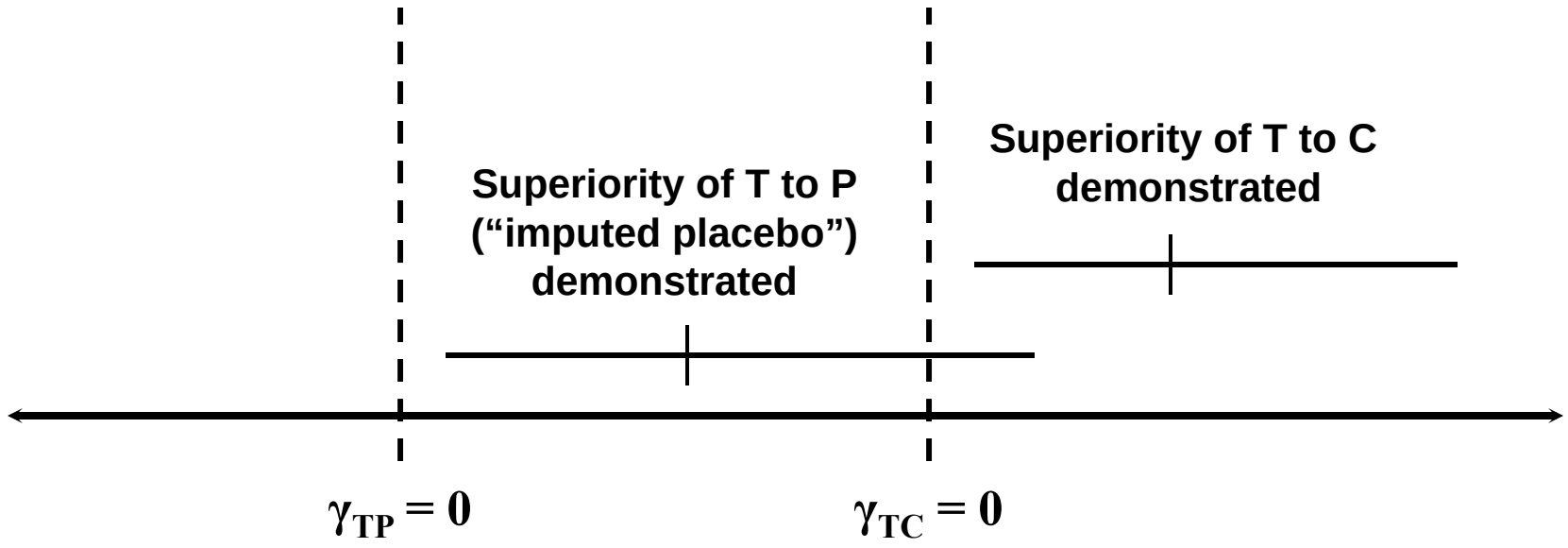
Relevant Parameters for the AC Trial

- For our AC trial, we can still define:
  - γ_{TC} = parameter for efficacy of T over C *(estimated in AC trial)*
  - γ_{CP} = parameter for efficacy of C over P *(from historical trial)*
  - γ_{TP} = parameter for efficacy of T over P *(not directly estimated)*

<< Assume increasing values of these represent greater efficacy >>

So a test to show that $\gamma_{TP} > 0$ (*i.e.* that T has greater efficacy than P) corresponds to the traditional "gold standard" for demonstrating the efficacy of a new treatment.

Illustration of Two Tests for Efficacy Using the 95% Confidence Interval Estimate for γ_{TC}



Two Types of Methods for Analyzing an AC Trial (1)

- Both require an historical estimate of γ_{CP}

  (For our discussion we will consider γ_{CP} as denoting the efficacy of C relative to either placebo or some prior standard of care)

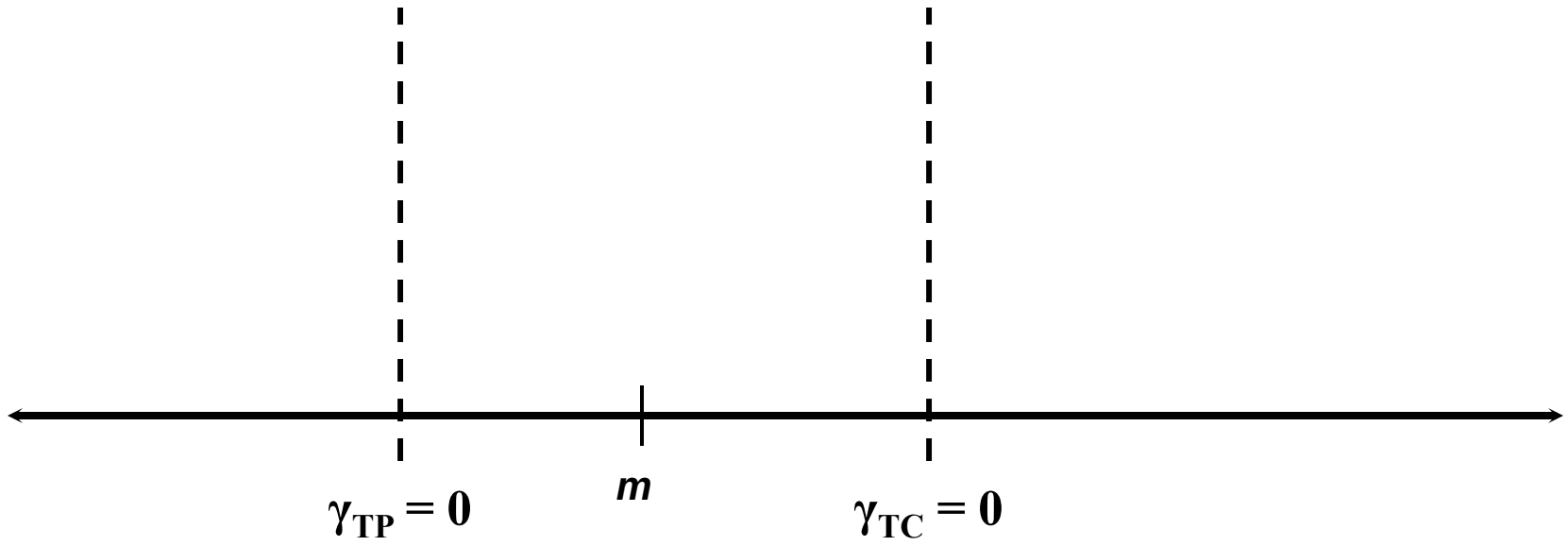
  - Fixed margin approach:

    - Use historical estimate of γ_{CP} (and possibly its standard error as well) to determine a fixed numerical margin *m* against which to test

      H_0: $\gamma_{TC} = m$ versus H_A: $\gamma_{TC} > m$

      - One example: The "95-95", or "two-confidence-interval" method sets *m* equal to (–1/2) the lower 95% confidence bound for γ_{CP}

The margin *m*, when considered on the scale of γ_{TC} , is a number less than 0:



Two Types of Methods for Analyzing an AC Trial (2)

- In general, fixed margin methods do not treat the historical estimate for γ_{CP} as a random variable.
  - Because of this, Rothmann and colleagues* showed that the "95-95" method typically leads to much lower-than-nominal α levels.

- Synthesis approach:
  - Assume γ_{CP} is constant (the so-called "constancy assumption") and then treat the historical estimate of γ_{CP} as a random variable.
  - Assume stochastic independence between historical data and the AC trial; then the following estimates can be a basis for analysis:
    - $\hat{\gamma}_{TP} = \hat{\gamma}_{TC} + \hat{\gamma}_{CP}$
    - $\hat{\text{Var}}\,[\hat{\gamma}_{TP}] = \hat{\text{Var}}\,[\hat{\gamma}_{TC}] + \hat{\text{Var}}\,[\hat{\gamma}_{CP}]$

* Stats in Med 22:239-264, 2003

FDA's "Preservation of Effect" Approach

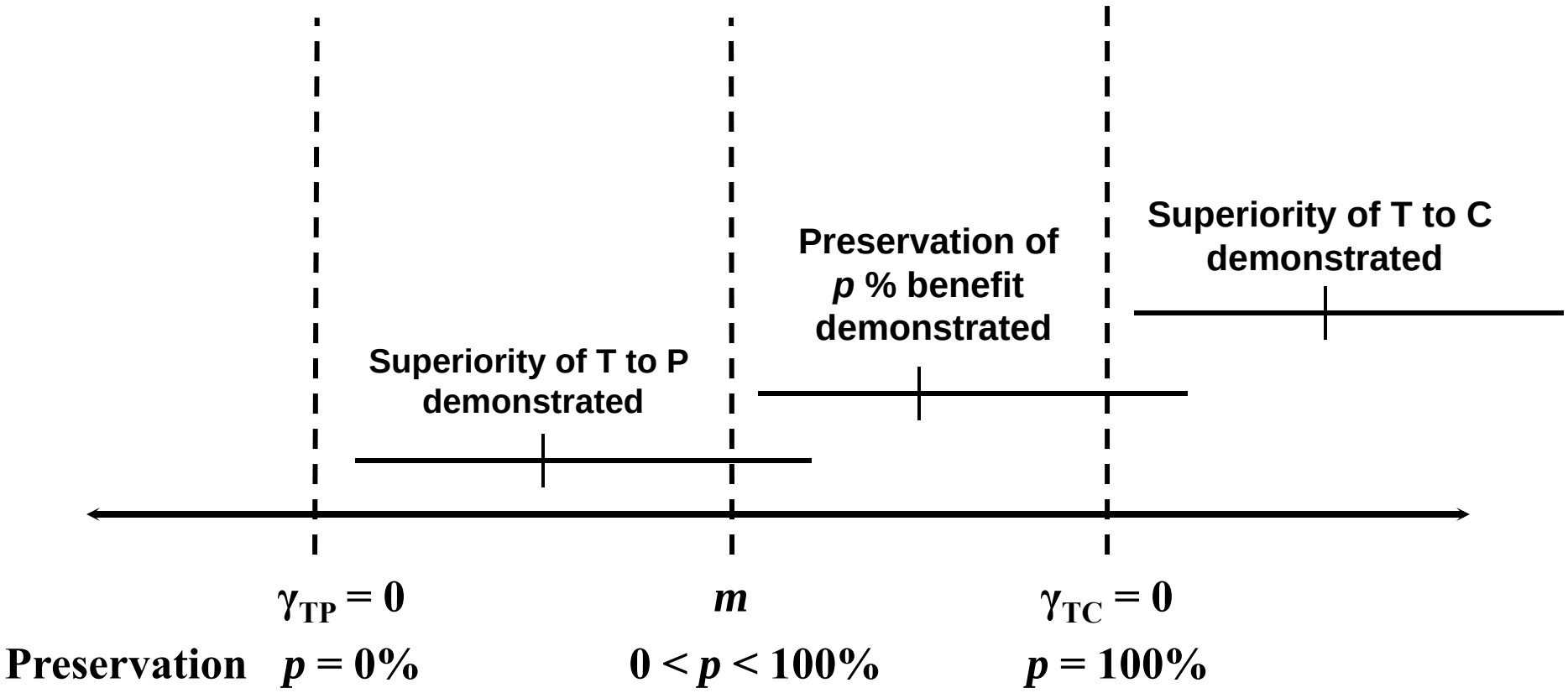
- FDA has typically rejected the "gold standard" (showing $\gamma_{TP} > 0$) when evaluating AC trials --- instead requiring that T preserve some fixed fraction of the efficacy benefit that C has over P:
  - *e.g.* Show that $\gamma_{TP} > (0.5)\ \gamma_{CP}$ << 50% preservation >>

    FDA has been inconsistent over how to test for preservation:
    - Synthesis approach used in some cases (*e.g.* Rothmann method*)
    - 95-95 or other fixed margin methods applied in other instances

If $\hat{V}ar\ [\hat{\gamma}_{TC}]$ can be adequately approximated as a function of the AC trial size, then the synthesis test of preservation will correspond to a fixed margin test (where the margin *m* is a function of the preservation percentage to be tested).

* Stats in Med 22:239-264, 2003

%-Preservation Illustrated by the 95% Confidence Interval Estimate for γ_{TC}



The Case for "One Standard of Evidence"

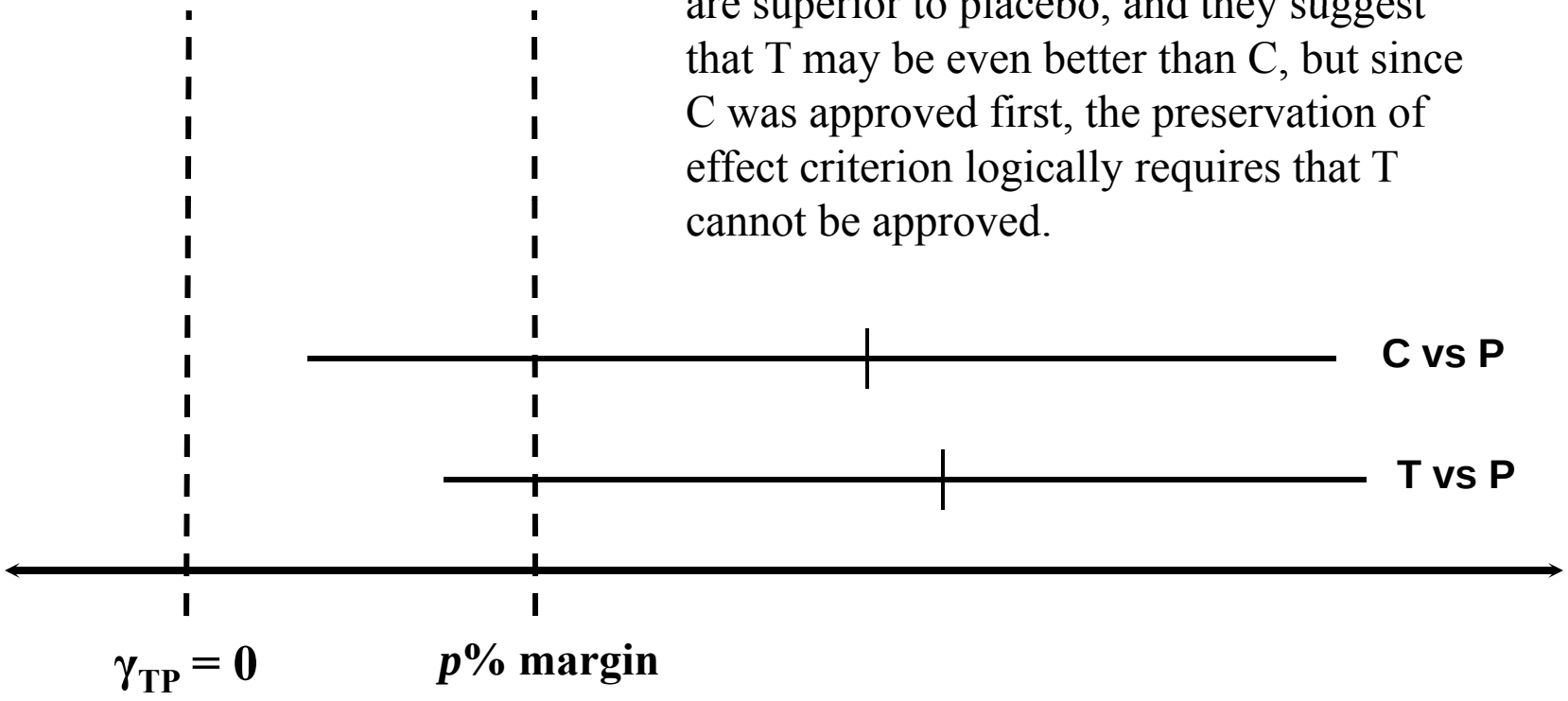
- The traditional standard of evidence for efficacy of a new treatment T is:
  → Statistically significant evidence that $\gamma_{TP} > 0$
- Why should an arbitrarily higher standard of evidence ($\gamma_{TP} > y > 0$) be applied when an AC trial has been performed?
- The preservation margin is necessarily arbitrary, in the sense that there will be values below the margin for which there is no meaningful clinical difference in efficacy from a value above the margin.
  - Preserving less than *p*% does **not** imply that T is an ineffective treatment.
  - In contrast, "$\gamma_{TP} = 0$" has a definite objective clinical meaning.
- Requiring a higher standard of evidence for AC trials institutes a regulatory bias in favor of the first drug to be approved.

The Case for One Standard, continued:

- The regulatory bias inherent in applying the preservation criterion leads to logical inconsistencies, as illustrated by the following example:
  - Suppose C and T were both evaluated in placebo-controlled trials, but C was approved first
  - Assume $\gamma_{TP} > \gamma_{CP}$
  - FDA's requirement that T preserve *p*% of C's benefit over placebo will in many plausible instances lead to rejection of T (even though T may be a better drug than C !)

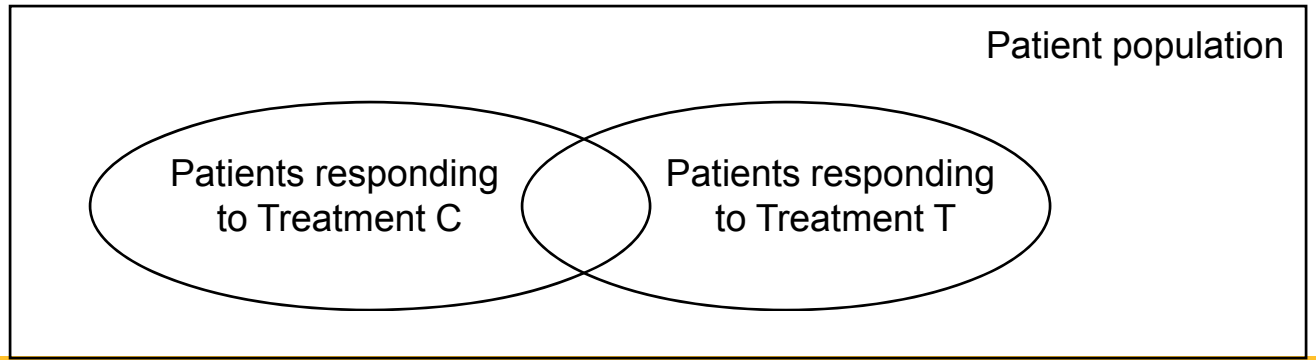
Hypothetical Example: 95% CIs for efficacy relative to P

The results here indicate that both C and T are superior to placebo, and they suggest that T may be even better than C, but since C was approved first, the preservation of effect criterion logically requires that T cannot be approved.



We All Lose with the Preservation Standard

- Our hypothetical example highlights unfairness and intellectual inconsistency in current regulatory practice.
  - Not only do pharmaceutical companies lose out in this framework: Physicians and patients may never get access to effective treatments excluded by these practices.
  - Two treatments with similar statistical efficacy may have considerably different efficacy for individual patients:



Example: Metastatic Bladder Cancer

- Randomized trial* of 2-drug regimen Gemzar + cisplatin compared to a 4-drug regimen MVA + cisplatin
- An earlier randomized trial** showed MVA + cisplatin was statistically superior to single-agent cisplatin
- Applying our notation: T = Gemzar and C = MVA
  - We will be asking whether Gemzar, when added to cisplatin, preserves enough of MVA's effect when added to cisplatin
    - γ_{TC} = log hazard ratio (MVA over Gemzar)
    - γ_{CP} = log hazard ratio (no treatment over MVA)
    - γ_{TP} = log hazard ratio (no treatment over Gemzar)

* Von der Masse et al: JCO, 18:3068-3077

** Loehrer et al: JCO, 10:1066-1073 (1992)

Metastatic Bladder Cancer, continued:

- Point estimates:

  $\hat{\gamma}_{CP}$ = 0.421 (with variance = 0.0181) from **

  $\hat{\gamma}_{TC}$ = -0.039 (with variance = 0.0147) from *

  $\hat{\gamma}_{TP}$ = 0.382 (with variance = 0.0328) from synthesis analysis

  → Rothmann/Synthesis method estimated a 90.7% preservation of benefit, but the lower 95% bound on this was only 11.7%

- So the "preservation of 50% of benefit" criterion was not met. However, Gemzar+cisplatin was statistically superior to cisplatin alone (two-sided p-value = 0.035).

* Von der Masse et al: JCO, 18:3068-3077

** Loehrer et al: JCO, 10:1066-1073 (1992)

More on the Bladder Cancer Example

- Assuming constancy of the parameter γ_{CP} across the trials:
  - The Gemzar+Cisplatin combination improves survival over single-agent Cisplatin (2-sided p = 0.035).
  - Survival with Gemzar+Cisplatin treatment was estimated to be similar to that with MVA+Cisplatin (estimated = hazard ratio = 0.96).
- So why do the test for preservation of effect?

Key Assumptions of the Synthesis Method

- Constancy: The parameter γ_{CP} is constant for all sources of data used in the analysis.
  - Note that γ_{CP} is a "between-treatment" parameter
    - The constancy assumption is much weaker than requiring a fixed patient population or constant within-treatment parameters
- Assay sensitivity: The active-controlled trial adequately measures the efficacy outcome and is capable of finding any differences between treatments in that outcome.
  - Assay sensitivity is more of an issue for "softer" efficacy outcomes.

Adjusting for Deviations from Assumptions

- Assay Sensitivity
  - departures defined quantitatively as $a = E(\hat{\gamma}_{TC}) - \gamma_{TC}$
- Constancy
  - departures defined quantitatively as $c = E(\hat{\gamma}_{CP}) - \gamma_{CP}$

- Snapinn and Jiang* have proposed an adjusted synthesis analysis for testing H_0: $\gamma_{TP} = 0$ vs H_A: $\gamma_{TP} > 0$ that is structurally equivalent to the Rothmann geometric test statistic:

$$\text{Reject } H_0 \text{ if } ((1-w)\hat{\gamma}_{CP} + \hat{\gamma}_{TC}) / \text{sqrt}\{(1-w)^2 \hat{\text{Var}}[\hat{\gamma}_{CP}] + \hat{\text{Var}}[\hat{\gamma}_{TC}]\} > 1.96$$

where w controls type-1 error optimally when $w = (a + c)/(\gamma_{CP} + c)$.

<< Snappin and Jiang suggest $w = 0.3$ should be sufficient for moderate departures (corresponds to $a = \gamma_{CP}/8$ and $c = \gamma_{CP}/4$) >>

* Stats in Med (2008) 27:371-381

Conclusions

- One standard of evidence (superiority to placebo) should be maintained regardless of the type of trial.
- The synthesis approach can test for superiority to placebo when an AC trial has been performed, accounting for the variability within both the AC trial and the historical data.
- The adjusted synthesis analysis can be applied in instances where deviations from constancy or assay sensitivity are of concern.
  - The adjustment factor *w* is mathematically equivalent to the preservation percentage in the Rothmann test statistic.
  - But setting *w* as high as 50% appears to be unreasonably conservative for most practical situations.