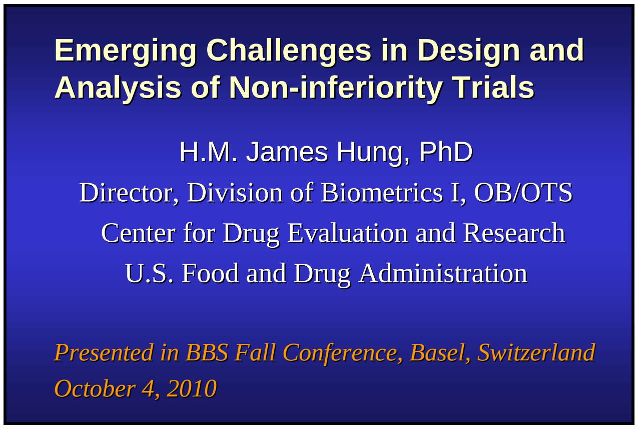
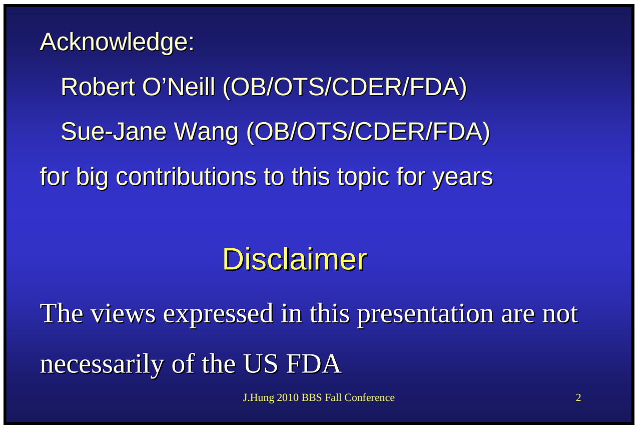


Emerging Challenges in Design and Analysis of Non-inferiority Trials

H.M. James Hung, PhD
Director, Division of Biometrics I, OB/OTS
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

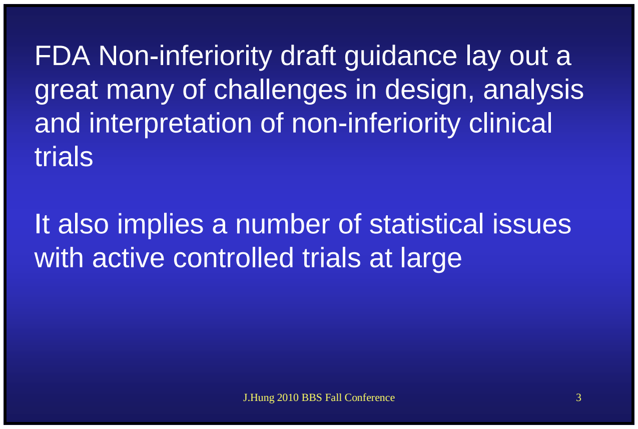
Presented in BBS Fall Conference, Basel, Switzerland
October 4, 2010



Acknowledge:
Robert O’Neill (OB/OTS/CDER/FDA)
Sue-Jane Wang (OB/OTS/CDER/FDA)
for big contributions to this topic for years

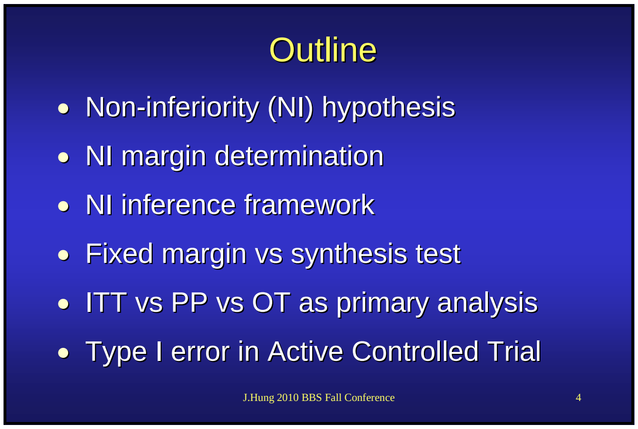
Disclaimer

The views expressed in this presentation are not necessarily of the US FDA



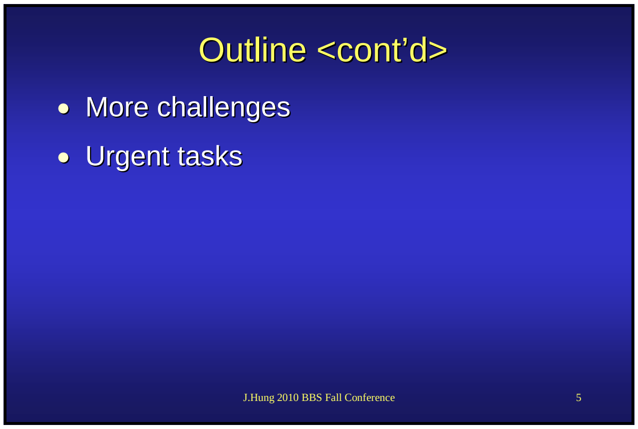
FDA Non-inferiority draft guidance lay out a great many of challenges in design, analysis and interpretation of non-inferiority clinical trials

It also implies a number of statistical issues with active controlled trials at large



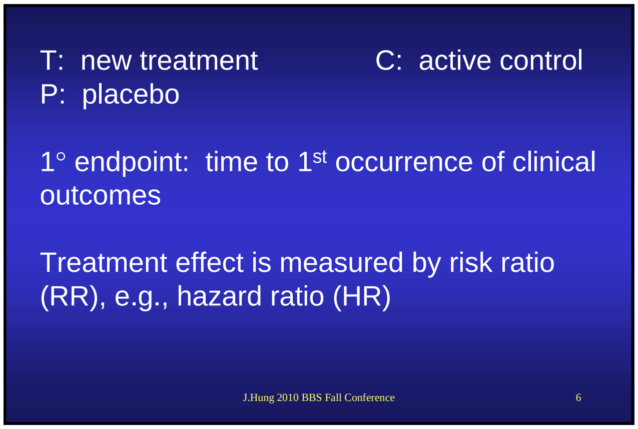
Outline

- Non-inferiority (NI) hypothesis
- NI margin determination
- NI inference framework
- Fixed margin vs synthesis test
- ITT vs PP vs OT as primary analysis
- Type I error in Active Controlled Trial



Outline <cont'd>

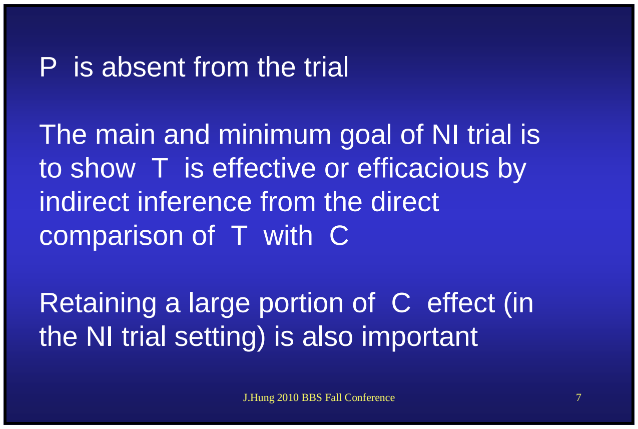
- More challenges
- Urgent tasks



T: new treatment C: active control
P: placebo

1° endpoint: time to 1st occurrence of clinical outcomes

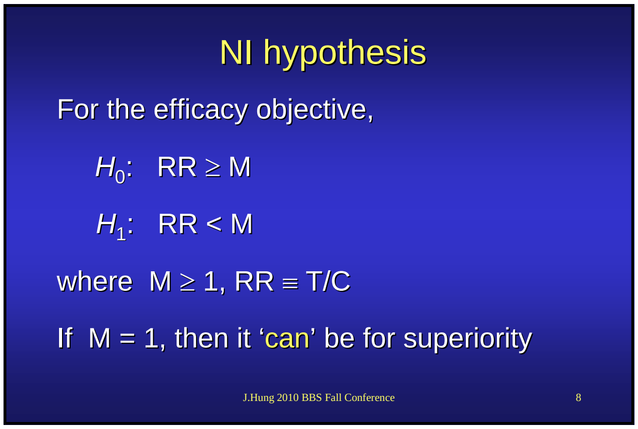
Treatment effect is measured by risk ratio (RR), e.g., hazard ratio (HR)



P is absent from the trial

The main and minimum goal of NI trial is to show T is effective or efficacious by indirect inference from the direct comparison of T with C

Retaining a large portion of C effect (in the NI trial setting) is also important



NI hypothesis

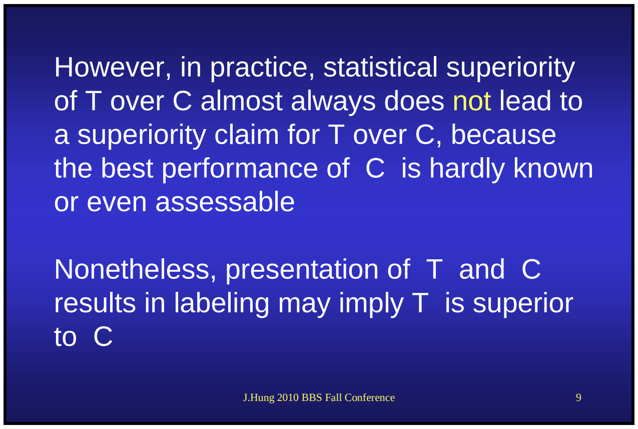
For the efficacy objective,

H_0: RR ≥ M

H_1: RR < M

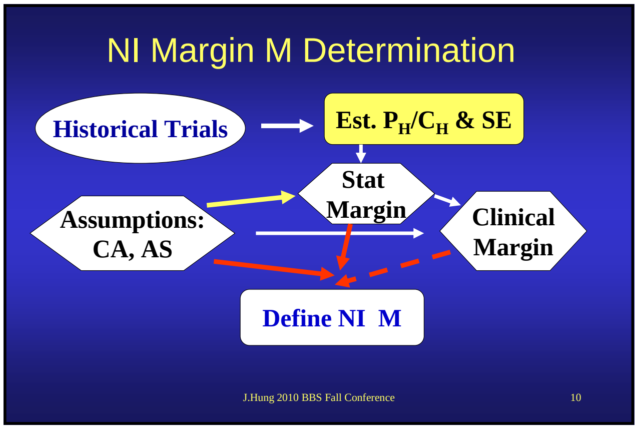
where M ≥ 1, RR ≡ T/C

If M = 1, then it 'can' be for superiority



However, in practice, statistical superiority of T over C almost always does not lead to a superiority claim for T over C, because the best performance of C is hardly known or even assessable

Nonetheless, presentation of T and C results in labeling may imply T is superior to C



NI Margin M Determination

Historical Trials

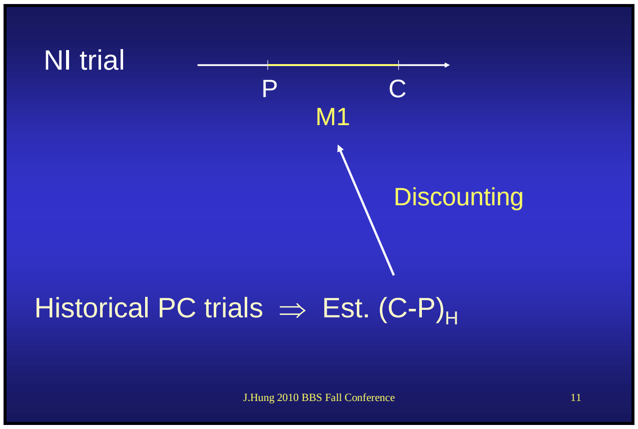
Est. P_H/C_H & SE

Stat Margin

Assumptions: CA, AS

Clinical Margin

Define NI M



NI trial

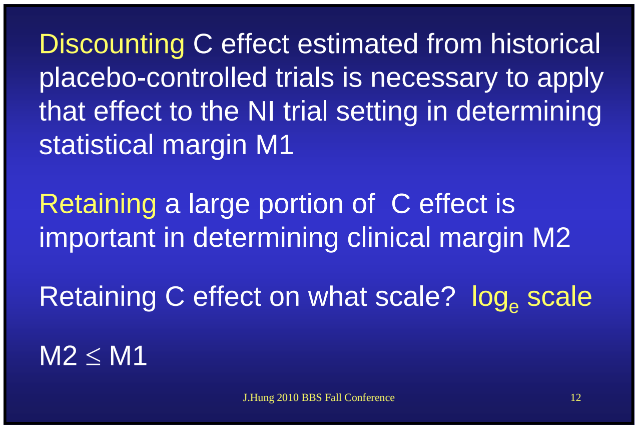
P

C

M1

Discounting

Historical PC trials ⇒ Est. $(C\text{-}P)_H$

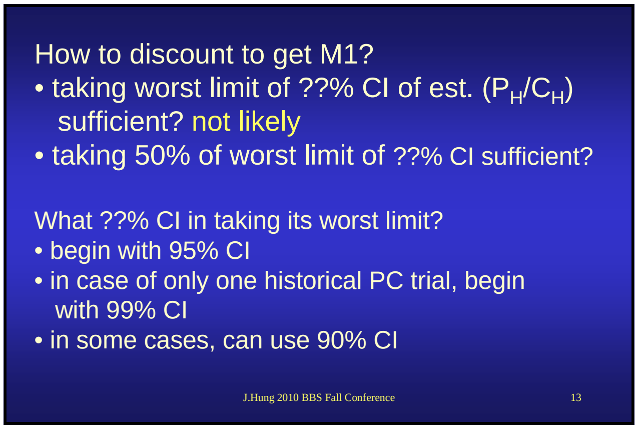


Discounting C effect estimated from historical placebo-controlled trials is necessary to apply that effect to the NI trial setting in determining statistical margin M1

Retaining a large portion of C effect is important in determining clinical margin M2

Retaining C effect on what scale? $\log_e$ scale

$M2 \leq M1$

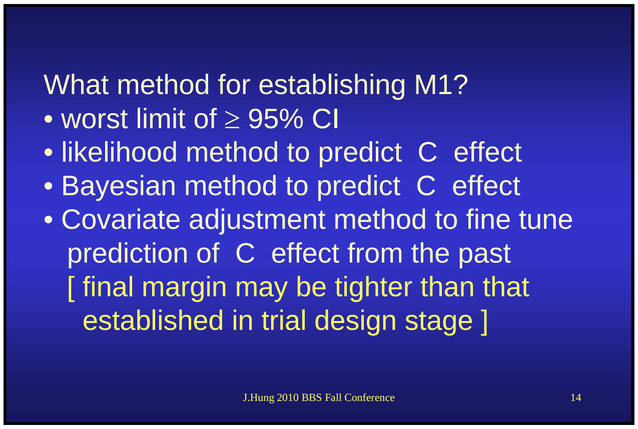


How to discount to get M1?

- taking worst limit of ??% CI of est. (P_H/C_H) sufficient? not likely
- taking 50% of worst limit of ??% CI sufficient?

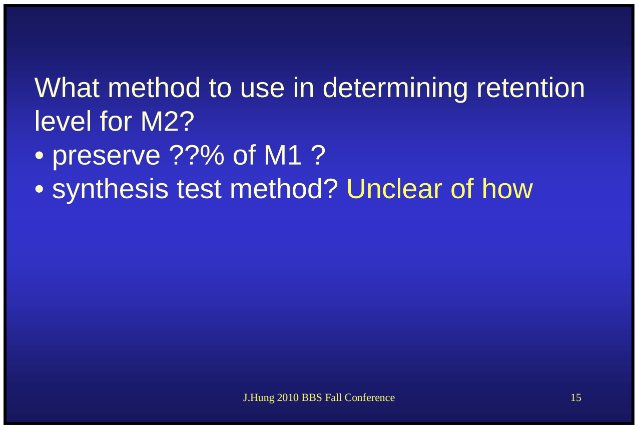
What ??% CI in taking its worst limit?

- begin with 95% CI
- in case of only one historical PC trial, begin with 99% CI
- in some cases, can use 90% CI



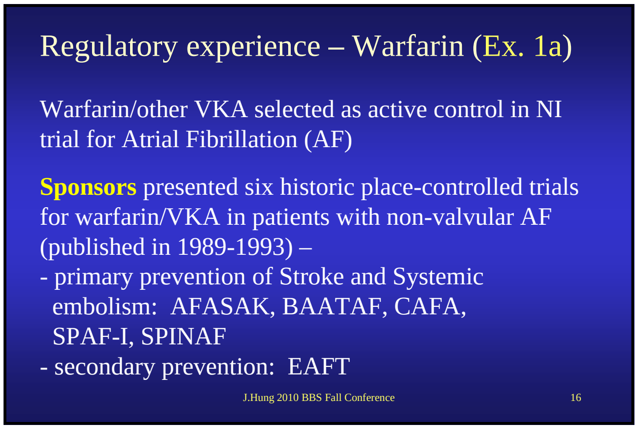
What method for establishing M1?

- worst limit of ≥ 95% CI
- likelihood method to predict C effect
- Bayesian method to predict C effect
- Covariate adjustment method to fine tune prediction of C effect from the past
  [final margin may be tighter than that established in trial design stage]



What method to use in determining retention level for M2?

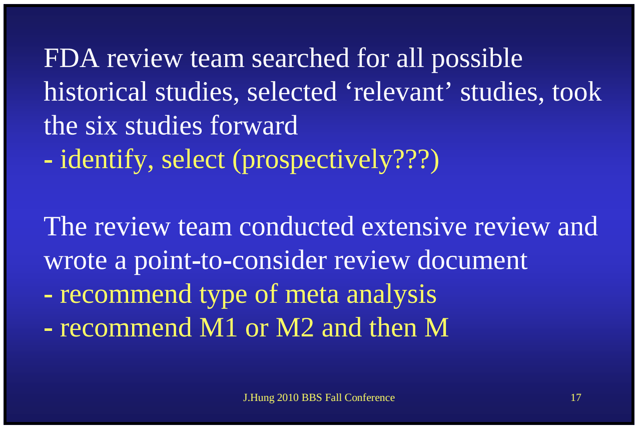
- preserve ??% of M1 ?
- synthesis test method? Unclear of how



Regulatory experience – Warfarin (Ex. 1a)

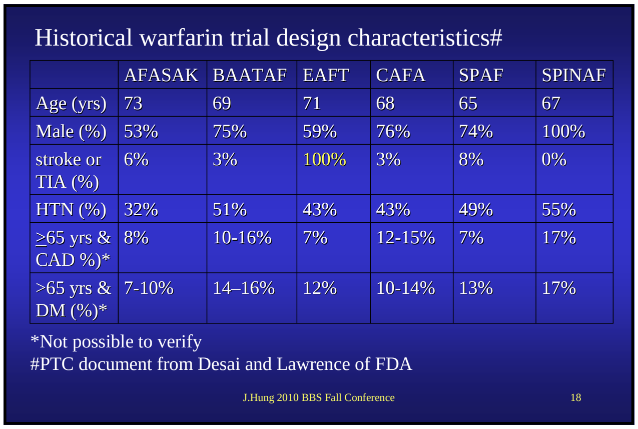
Warfarin/other VKA selected as active control in NI trial for Atrial Fibrillation (AF)

Sponsors presented six historic place-controlled trials for warfarin/VKA in patients with non-valvular AF (published in 1989-1993) –
- primary prevention of Stroke and Systemic embolism: AFASAK, BAATAF, CAFA, SPAF-I, SPINAF
- secondary prevention: EAFT



FDA review team searched for all possible historical studies, selected ‘relevant’ studies, took the six studies forward
- identify, select (prospectively???)

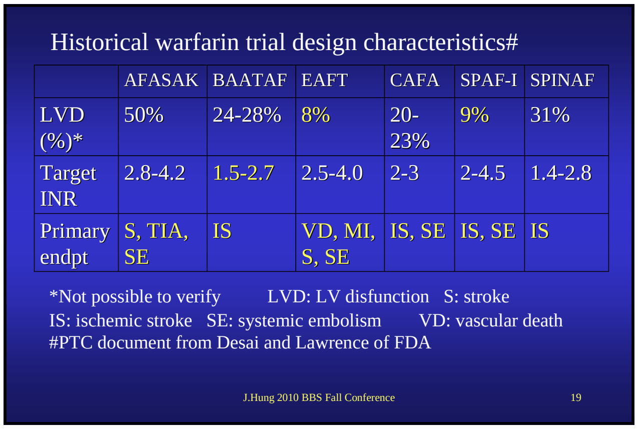
The review team conducted extensive review and wrote a point-to-consider review document
- recommend type of meta analysis
- recommend M1 or M2 and then M



Historical warfarin trial design characteristics#

	AFASAK	BAATAF	EAFT	CAFA	SPAF	SPINAF
Age (yrs)	73	69	71	68	65	67
Male (%)	53%	75%	59%	76%	74%	100%
stroke or TIA (%)	6%	3%	100%	3%	8%	0%
HTN (%)	32%	51%	43%	43%	49%	55%
≥65 yrs & CAD %)*	8%	10-16%	7%	12-15%	7%	17%
>65 yrs & DM (%)*	7-10%	14–16%	12%	10-14%	13%	17%

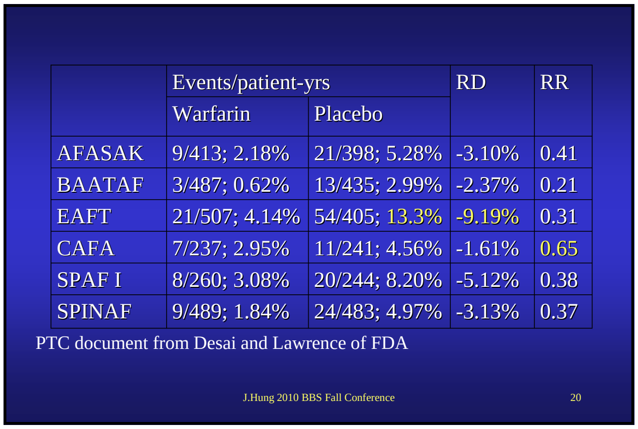
*Not possible to verify
#PTC document from Desai and Lawrence of FDA



Historical warfarin trial design characteristics#

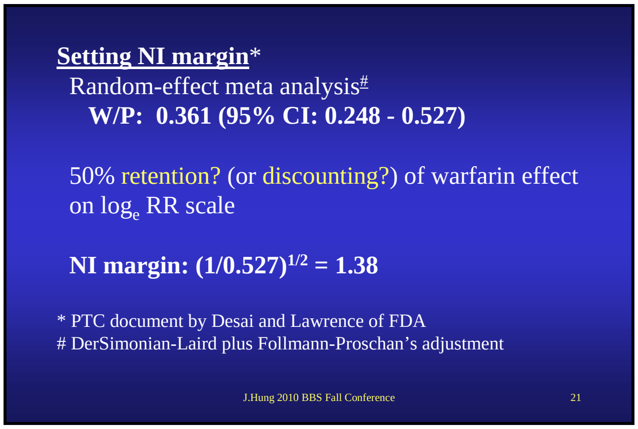
	AFASAK	BAATAF	EAFT	CAFA	SPAF-I	SPINAF
LVD (%)*	50%	24-28%	8%	20-23%	9%	31%
Target INR	2.8-4.2	1.5-2.7	2.5-4.0	2-3	2-4.5	1.4-2.8
Primary endpt	S, TIA, SE	IS	VD, MI, S, SE	IS, SE	IS, SE	IS

*Not possible to verify LVD: LV disfunction S: stroke
IS: ischemic stroke SE: systemic embolism VD: vascular death
#PTC document from Desai and Lawrence of FDA



	Events/patient-yrs		RD	RR
	Warfarin	Placebo		
AFASAK	9/413; 2.18%	21/398; 5.28%	-3.10%	0.41
BAATAF	3/487; 0.62%	13/435; 2.99%	-2.37%	0.21
EAFT	21/507; 4.14%	54/405; 13.3%	-9.19%	0.31
CAFA	7/237; 2.95%	11/241; 4.56%	-1.61%	0.65
SPAF I	8/260; 3.08%	20/244; 8.20%	-5.12%	0.38
SPINAF	9/489; 1.84%	24/483; 4.97%	-3.13%	0.37

PTC document from Desai and Lawrence of FDA



Setting NI margin*

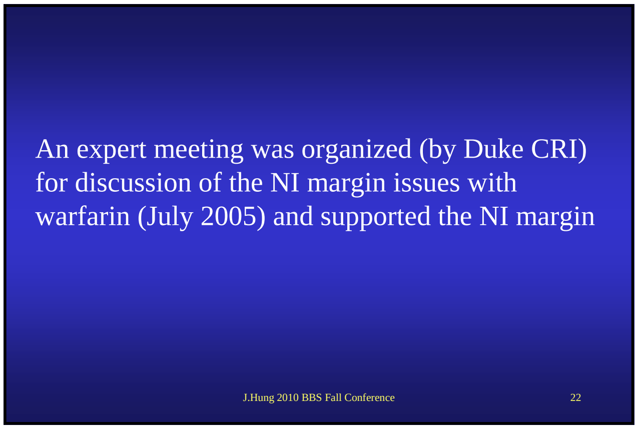
Random-effect meta analysis[#]

W/P: 0.361 (95% CI: 0.248 - 0.527)

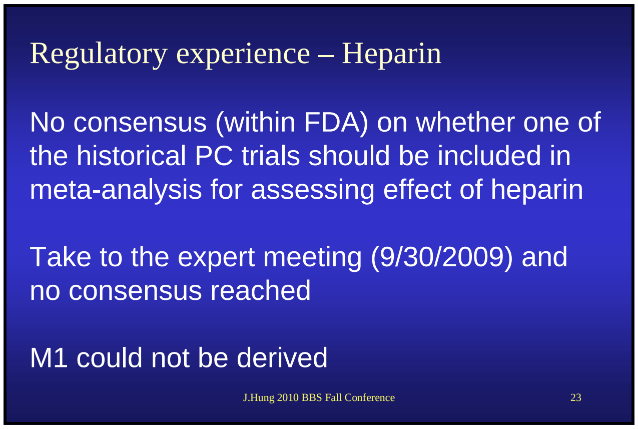
50% retention? (or discounting?) of warfarin effect on $\log_e$ RR scale

NI margin: $(1/0.527)^{1/2}$ = 1.38

* PTC document by Desai and Lawrence of FDA
# DerSimonian-Laird plus Follmann-Proschan’s adjustment



An expert meeting was organized (by Duke CRI) for discussion of the NI margin issues with warfarin (July 2005) and supported the NI margin

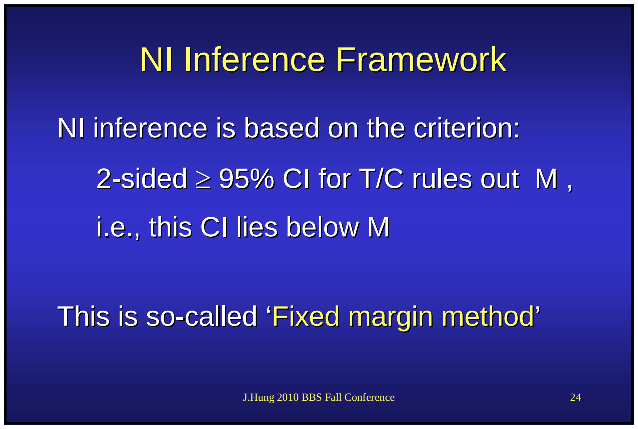


Regulatory experience – Heparin

No consensus (within FDA) on whether one of the historical PC trials should be included in meta-analysis for assessing effect of heparin

Take to the expert meeting (9/30/2009) and no consensus reached

M1 could not be derived



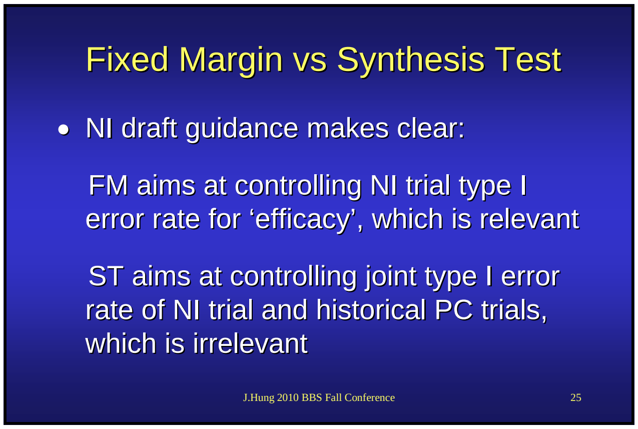
NI Inference Framework

NI inference is based on the criterion:

2-sided ≥ 95% CI for T/C rules out M ,

i.e., this CI lies below M

This is so-called ‘Fixed margin method’

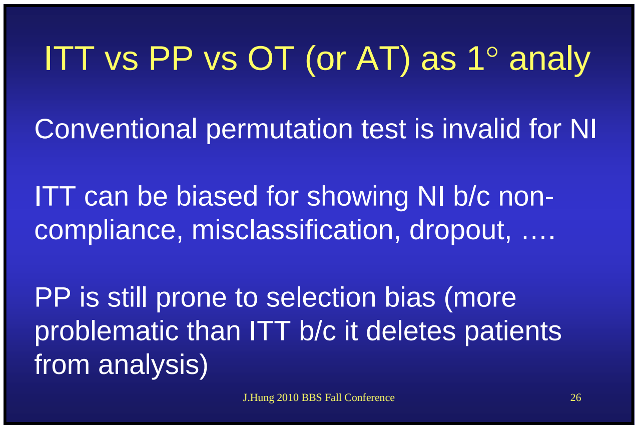


Fixed Margin vs Synthesis Test

- NI draft guidance makes clear:

  FM aims at controlling NI trial type I error rate for ‘efficacy’, which is relevant

  ST aims at controlling joint type I error rate of NI trial and historical PC trials, which is irrelevant

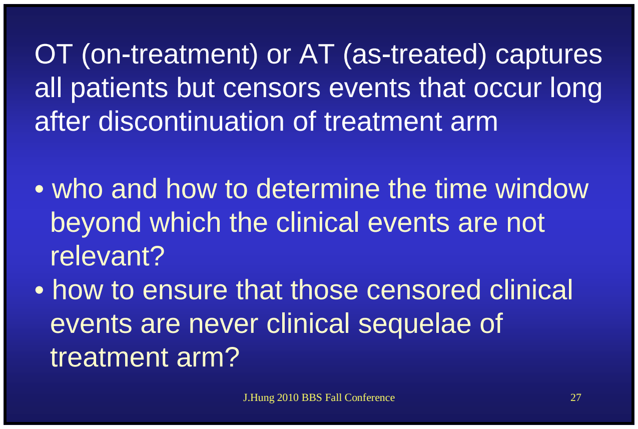


ITT vs PP vs OT (or AT) as 1° analy

Conventional permutation test is invalid for NI

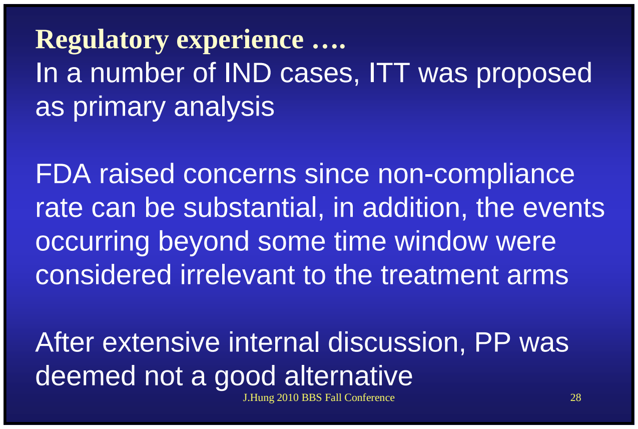
ITT can be biased for showing NI b/c non-compliance, misclassification, dropout, ….

PP is still prone to selection bias (more problematic than ITT b/c it deletes patients from analysis)



OT (on-treatment) or AT (as-treated) captures all patients but censors events that occur long after discontinuation of treatment arm

- who and how to determine the time window beyond which the clinical events are not relevant?
- how to ensure that those censored clinical events are never clinical sequelae of treatment arm?

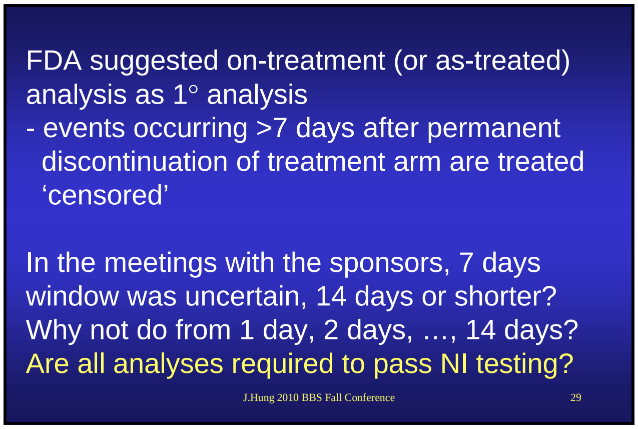


Regulatory experience ….

In a number of IND cases, ITT was proposed as primary analysis

FDA raised concerns since non-compliance rate can be substantial, in addition, the events occurring beyond some time window were considered irrelevant to the treatment arms

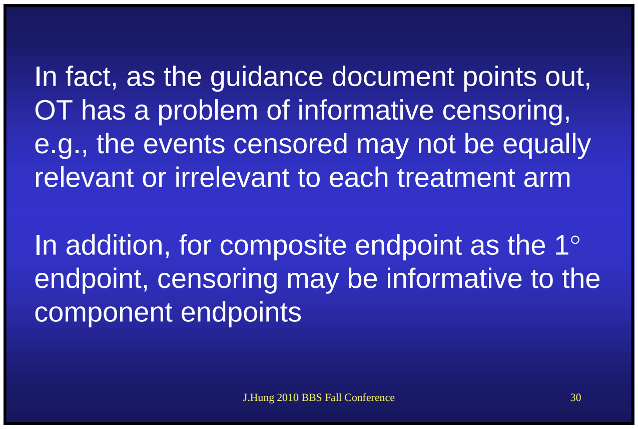
After extensive internal discussion, PP was deemed not a good alternative



FDA suggested on-treatment (or as-treated) analysis as 1° analysis

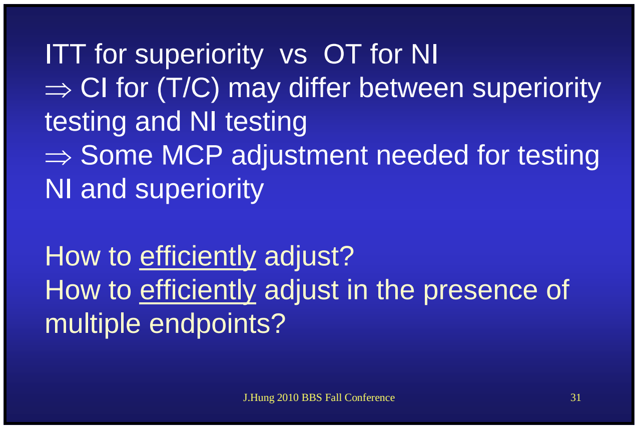
- events occurring >7 days after permanent discontinuation of treatment arm are treated ‘censored’

In the meetings with the sponsors, 7 days window was uncertain, 14 days or shorter? Why not do from 1 day, 2 days, …, 14 days? Are all analyses required to pass NI testing?



In fact, as the guidance document points out, OT has a problem of informative censoring, e.g., the events censored may not be equally relevant or irrelevant to each treatment arm

In addition, for composite endpoint as the 1° endpoint, censoring may be informative to the component endpoints



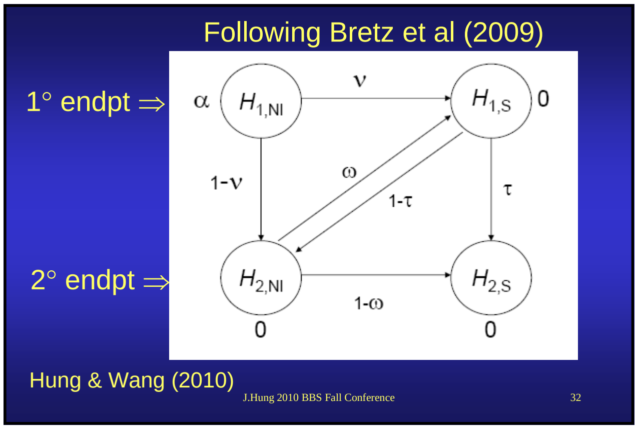
ITT for superiority vs OT for NI

⇒ CI for (T/C) may differ between superiority testing and NI testing

⇒ Some MCP adjustment needed for testing NI and superiority

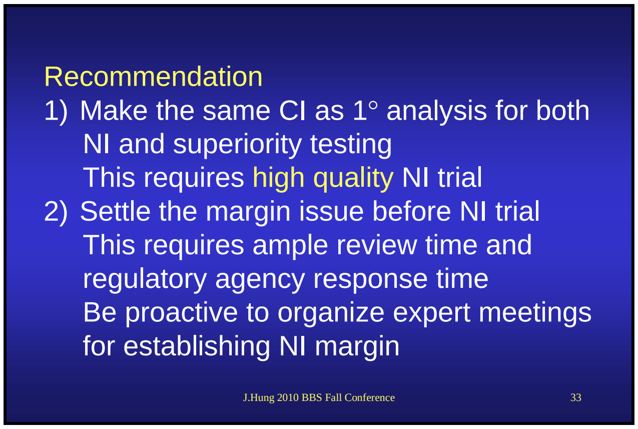
How to efficiently adjust?

How to efficiently adjust in the presence of multiple endpoints?



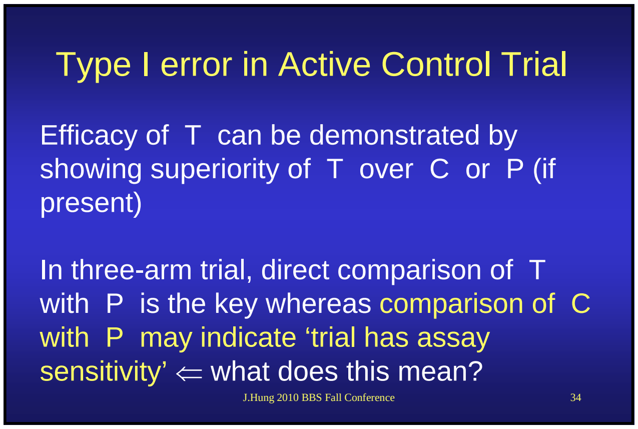
Following Bretz et al (2009)

Hung & Wang (2010)



Recommendation

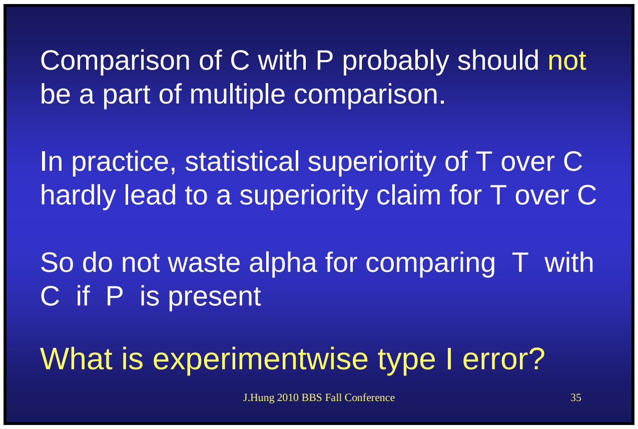
1) Make the same CI as 1° analysis for both NI and superiority testing
   This requires high quality NI trial
2) Settle the margin issue before NI trial
   This requires ample review time and regulatory agency response time
   Be proactive to organize expert meetings for establishing NI margin



Type I error in Active Control Trial

Efficacy of T can be demonstrated by showing superiority of T over C or P (if present)

In three-arm trial, direct comparison of T with P is the key whereas comparison of C with P may indicate 'trial has assay sensitivity' ⇐ what does this mean?

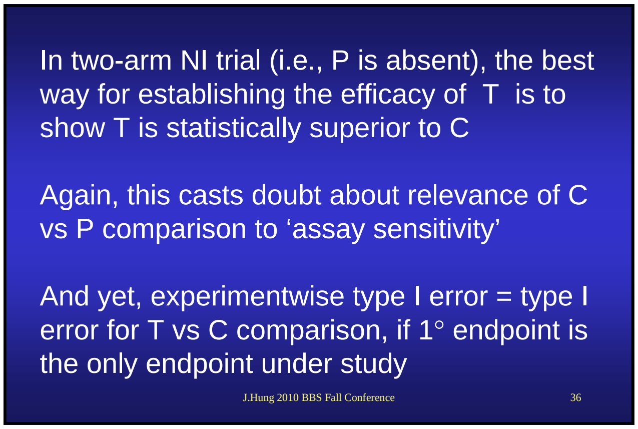


Comparison of C with P probably should not be a part of multiple comparison.

In practice, statistical superiority of T over C hardly lead to a superiority claim for T over C

So do not waste alpha for comparing T with C if P is present

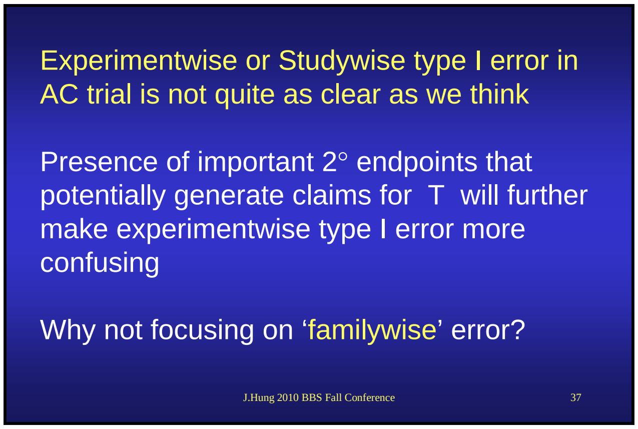
What is experimentwise type I error?



In two-arm NI trial (i.e., P is absent), the best way for establishing the efficacy of T is to show T is statistically superior to C

Again, this casts doubt about relevance of C vs P comparison to 'assay sensitivity'

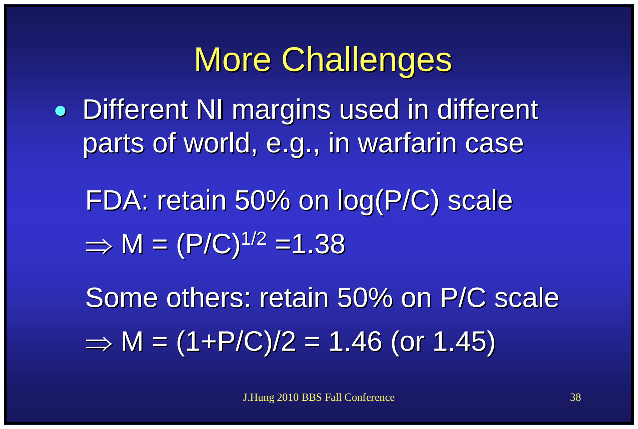
And yet, experimentwise type I error = type I error for T vs C comparison, if 1° endpoint is the only endpoint under study



Experimentwise or Studywise type I error in AC trial is not quite as clear as we think

Presence of important 2° endpoints that potentially generate claims for T will further make experimentwise type I error more confusing

Why not focusing on 'familywise' error?



More Challenges

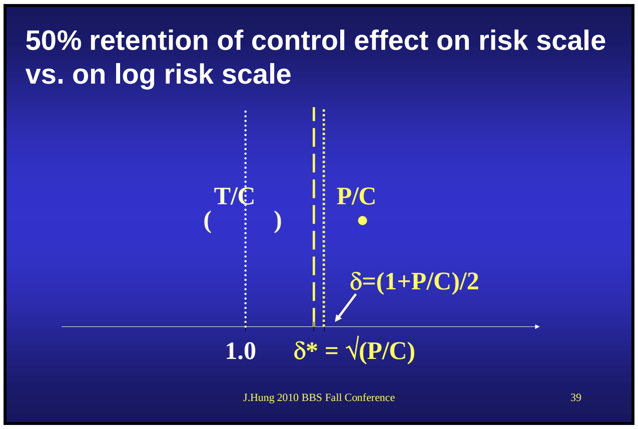
- Different NI margins used in different parts of world, e.g., in warfarin case

  FDA: retain 50% on log(P/C) scale

  $\Rightarrow$ M = $(P/C)^{1/2}$ =1.38

  Some others: retain 50% on P/C scale

  $\Rightarrow$ M = (1+P/C)/2 = 1.46 (or 1.45)



50% retention of control effect on risk scale vs. on log risk scale

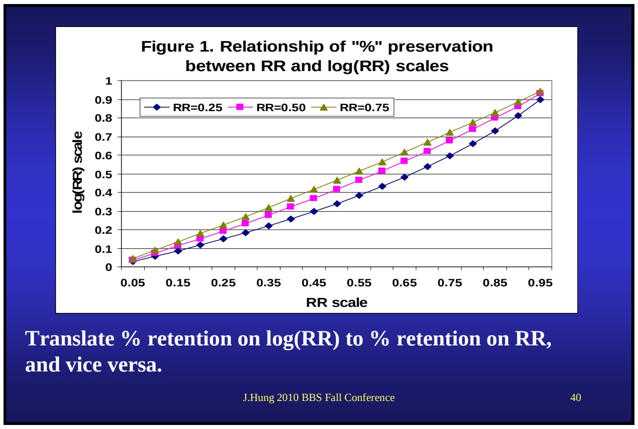
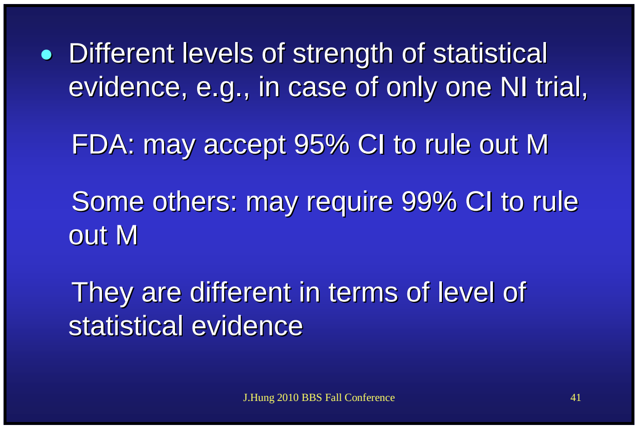


Figure 1. Relationship of "%" preservation between RR and log(RR) scales

Translate % retention on log(RR) to % retention on RR, and vice versa.

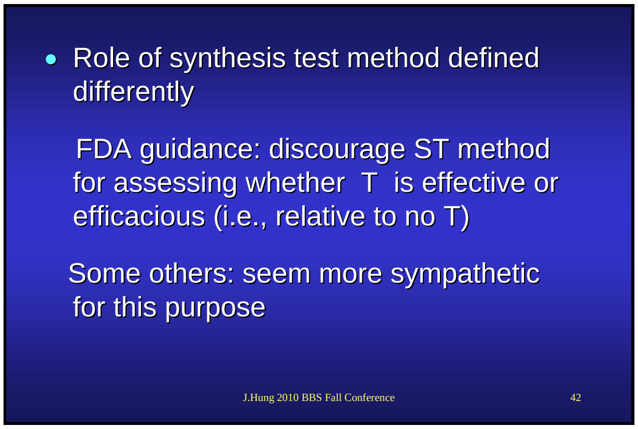


- Different levels of strength of statistical evidence, e.g., in case of only one NI trial,

FDA: may accept 95% CI to rule out M

Some others: may require 99% CI to rule out M

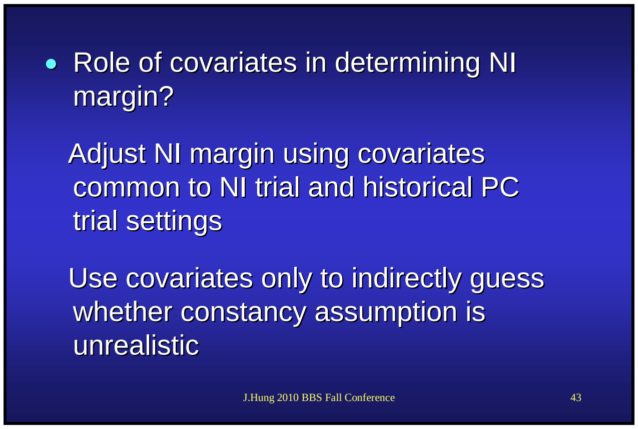
They are different in terms of level of statistical evidence



- Role of synthesis test method defined differently

FDA guidance: discourage ST method for assessing whether T is effective or efficacious (i.e., relative to no T)

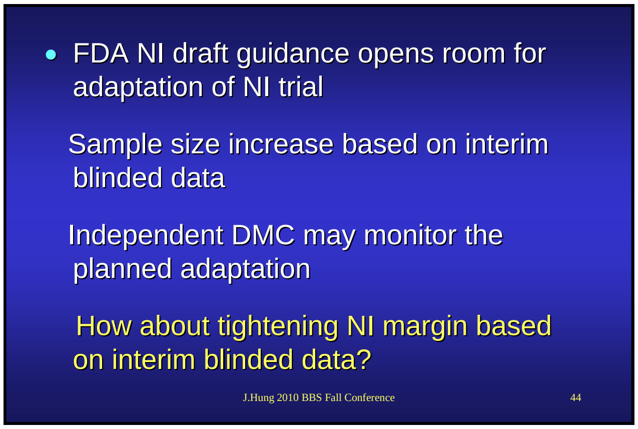
Some others: seem more sympathetic for this purpose



- Role of covariates in determining NI margin?

Adjust NI margin using covariates common to NI trial and historical PC trial settings

Use covariates only to indirectly guess whether constancy assumption is unrealistic

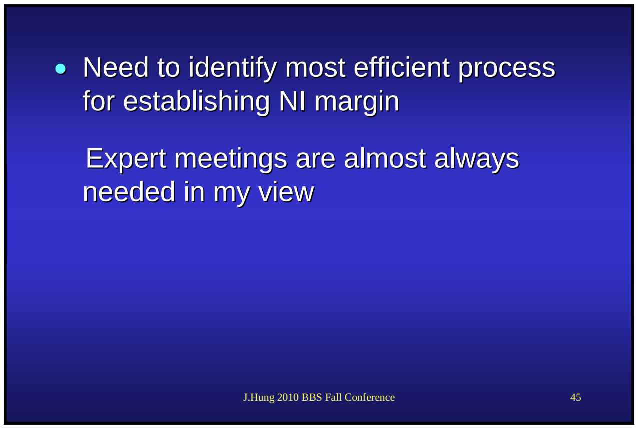


- FDA NI draft guidance opens room for adaptation of NI trial

Sample size increase based on interim blinded data

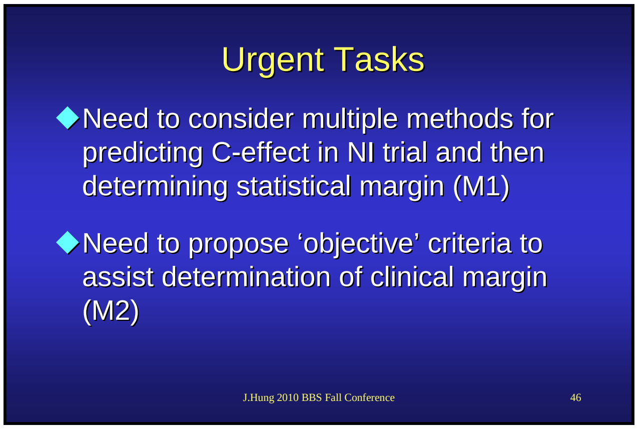
Independent DMC may monitor the planned adaptation

How about tightening NI margin based on interim blinded data?



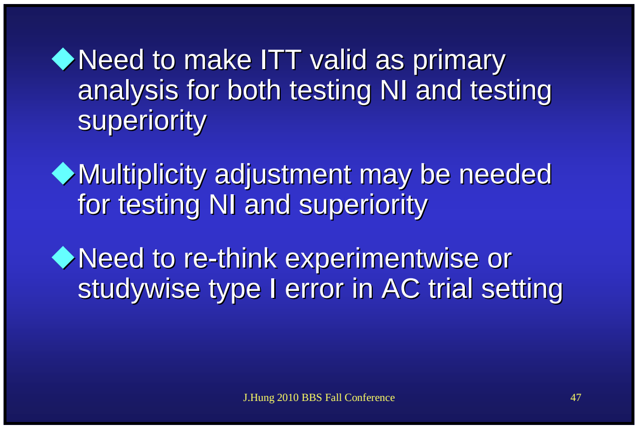
- Need to identify most efficient process for establishing NI margin

  Expert meetings are almost always needed in my view

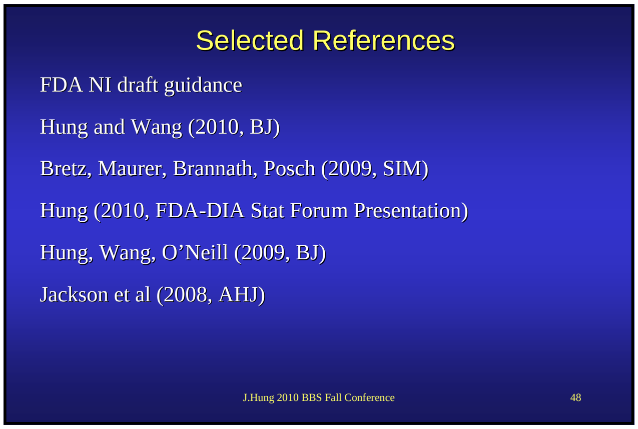


Urgent Tasks

- Need to consider multiple methods for predicting C-effect in NI trial and then determining statistical margin (M1)
- Need to propose ‘objective’ criteria to assist determination of clinical margin (M2)



- Need to make ITT valid as primary analysis for both testing NI and testing superiority
- Multiplicity adjustment may be needed for testing NI and superiority
- Need to re-think experimentwise or studywise type I error in AC trial setting



Selected References

FDA NI draft guidance

Hung and Wang (2010, BJ)

Bretz, Maurer, Brannath, Posch (2009, SIM)

Hung (2010, FDA-DIA Stat Forum Presentation)

Hung, Wang, O’Neill (2009, BJ)

Jackson et al (2008, AHJ)