



School of Medicine

Department of Family Medicine
and Public Health Sciences



WAYNE STATE
UNIVERSITY
MOLECULAR MEDICINE AND GENETICS

DESIGN AND ANALYSIS OF NON-INFERIORITY TRIALS: SOME FREQUENTIST AND BAYESIAN PERSPECTIVE

Dr. Samiran Ghosh
sghos@med.wayne.edu

**Director of Biostatistics at BERD,
Wayne State University**

Acknowledgement: Dr. Shrabanti Chowdhury, Dr. Santu Ghosh
and Dr. Ram Tiwari (FDA, OSB, CDRH)

FUNDING AND CONFLICTS OF INTEREST

Major Funding is provided by,

- **ME-1409-21410** a PCORI contract as PI
- P30-ES020957 a NIEHS Center grant, as an Associate Director of Biostatistics Core (Co-I)
- HD-089875-01 a U19 NIH grant on HIV Prevention and Treatment Adherence (Co-I)
- R01 HD089000-NICHD grant (Co-I)
- R01 NS079429 NINDS grant (Co-I)
- R21 HS024750 an AHRQ grant (Co-I)

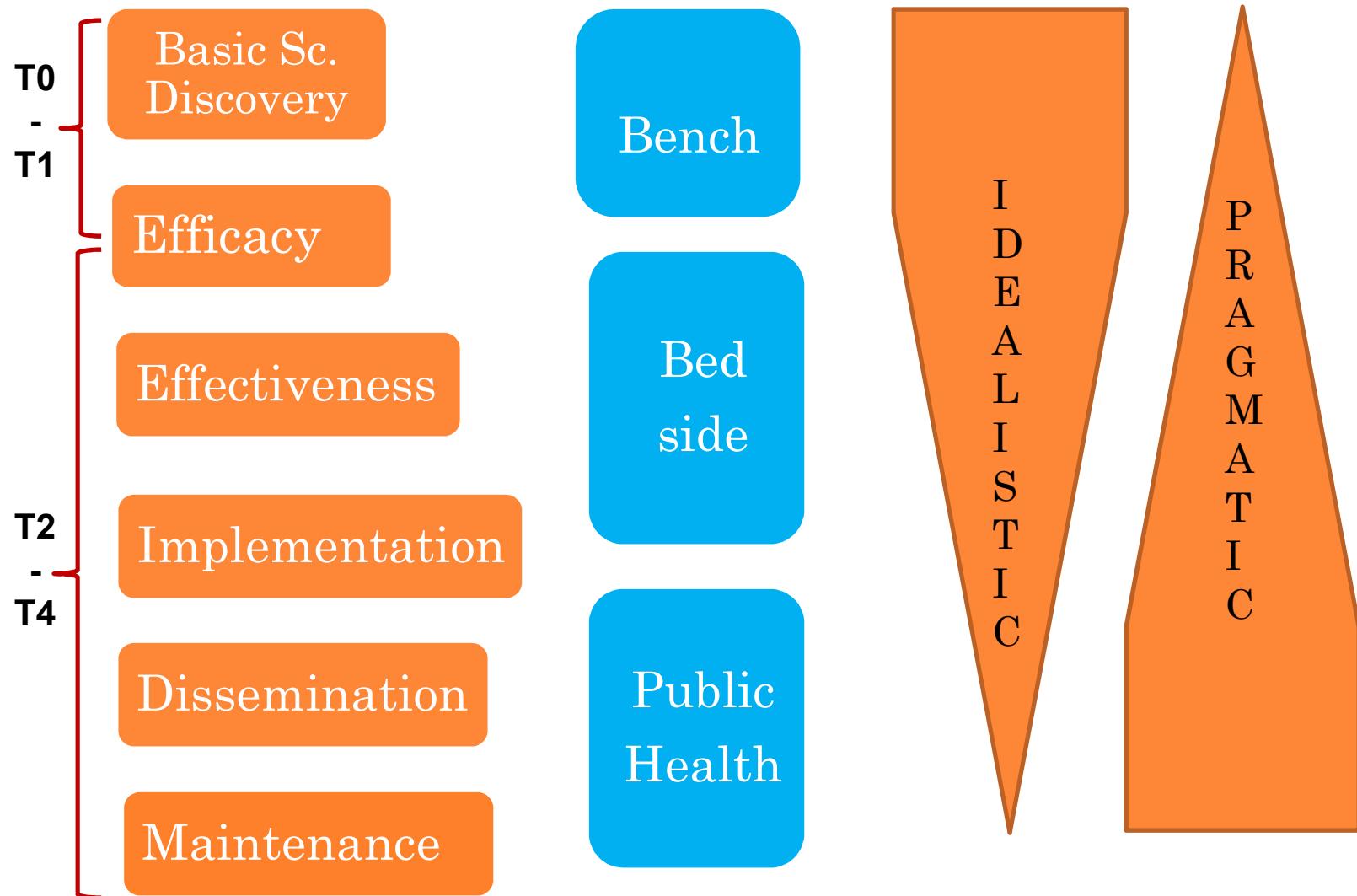
COI

- This should not be taken as the official view of FDA,
rather that of the authors only

OUTLINE OF THE TALK

1. Efficacy vs. Comparative Effectiveness (CE)
2. Description of the Problem of Non-inferiority
3. What is “Gold Standard” or 3-arm design/trial
4. Some Literature Reviews on various Statistical issues (under *Classical & Frequentist* setup)
5. Fraction Margin vs. Fixed Margin Approach
6. Group Sequential Design in NI
7. *Bayesian* Formulation of NI trial
8. A Conditional Testing Approach
9. Some Applications
10. Some Concluding Remarks

TRANSLATIONAL RESEARCH SPECTRUM



EFFICACY VS. EFFECTIVENESS RESEARCH

- **Efficacy** (de-jure) is the capacity for beneficial change (or therapeutic effect) of a given intervention under *Idealized* condition (or as directed)
- To determine whether a drug or an intervention is efficacious we often proceed via Randomized Controlled Trial (RCT)
- RCT is based on three basic principals of Design of Experiment,
 1. Randomize, 2. Replicate and 3. Blocking
- RCT involves human subjects
- Prospective controlled experiment under strict inclusion-exclusion criteria (hence the term *Idealized*)
- Note also, to declare something efficacious we need a “control” group

COMPARATIVE EFFECTIVENESS RESEARCH

“...comparing the relative benefits and harms among range of available treatments or interventions for a given health condition...” - - - AHRQ

- De-facto - A composite of efficacy and adherence
- Need to compare more than one Active Interventions
- Conducted possibly in a “Real-World” setting

CER can be conducted via,

- Randomized Clinical Trial (RCT)
 1. Tight Internal Validity, Hypothesis Driven, Causality
 2. Costly, Somewhat Artificial Setup
- Observational Study
 1. Cheaper, Higher External Validity
 2. Selection Bias, Confounding etc.

RCT: A BRIEF PRIMER

- Randomized controlled clinical trials (RCTs) are an indispensable source of information about efficacy of treatments in almost any disease area
- RCTs place a strong emphasis on internal validity with randomization, double-blinding, and control or comparison groups
- The goal is to determine whether certain intervention is efficacious compared to a control group
- In the absence of an effective treatment the usefulness of placebo controlled RCTs are uncontroversial
- However in the presence of an established effective regime, placebo controlled RCTs are non-ethical

CER AND NON-INFERIORITY TRIAL

- Implication of placing an Active Comparator arm in RCT is huge
- This gives rise to *Superiority* and *Non-Inferiority* trial
- When Superiority of an Intervention is questionable Non-Inferiority (NI) is often proposed
- NI intervention may be *slightly less* efficacious
- However it may be less toxic, less invasive, cheaper or less debilitating, and hence preferable
- In a Classical Non-inferiority trial placebo arm is absent
- This makes ethical sense but may lead to serious inferential consequence

NON-INFERIORITY(NI) TRIAL

- An Experimental Intervention (E) is compared with another active control (R) (e.g. established regime)
- Done mostly for ethical reason when it is established that R is clinically preferable to a placebo (P) in a population/sub-population of interest
- It requires a clinically acceptable *margin* ($\Delta > 0$)
- Primarily interested in testing (e.g. the mean μ),

$$\text{NI: } H_0 : \mu_E - \mu_R \leq -\Delta \text{ vs. } H_1 : \mu_E - \mu_R > -\Delta$$

(Some authors define NI margin as $\delta < 0$, in that case $\delta = -\Delta$)

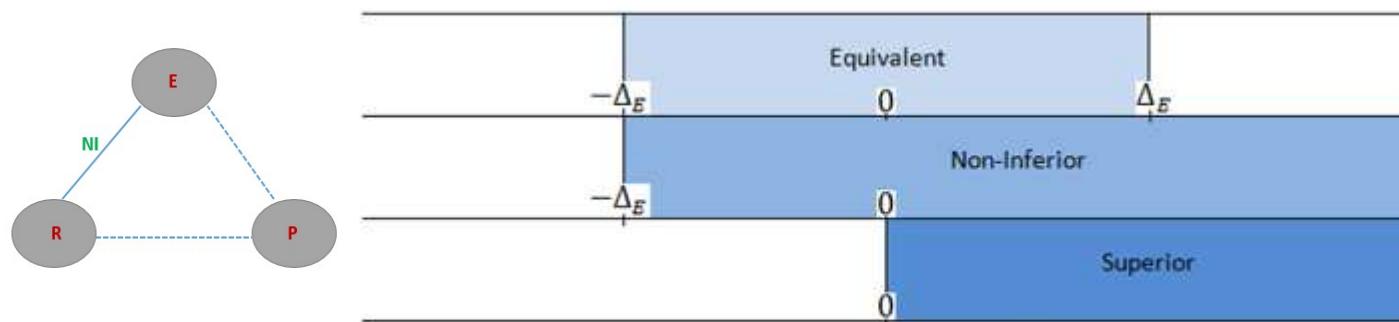
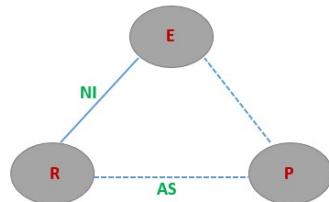


Figure 1: Illustrates the Equivalence, Non-Inferiority and Superiority intervals.

ASSUMPTIONS IN NI TRIAL

- Since in NI-trial placebo arm is absent, we are making implicit assumption of *constancy* and *assay sensitivity*
- Constancy: The historic difference between the R and P are still holds in the current trial. i.e. we still reject the null hypothesis in the presentt setup
- Assay Sensitivity (AS): The ability of current trial to distinguish an effective treatment from a less effective or ineffective intervention (e.g. placebo)
- In short in the absence of Placebo arm we are hoping we will still reject the null hypothesis if we can test,



$$H_0 : \mu_R - \mu_P \leq \gamma \quad \text{vs.} \quad H_1 : \mu_R - \mu_P > \gamma,$$

where $\gamma \geq 0$ is a known constant

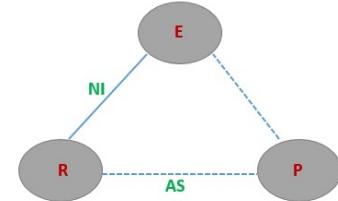
10

- This validation must be done externally

TWO-ARM NI TRIAL: ISSUES

- Many factors may affect AS: Poor disease diagnosis, endpoint selection and timing, poor adherence, loss to follow-up, prior medication/exposure etc.
- Constancy condition is also not testable without a concurrent placebo group
- To avoid these assumptions, if it is ethically OK, EMA recommended adding a placebo arm for internal validation
- Resulting design is a 3-arm “Gold Standard design” which does not require *External Validation*
- Note:- We still need to worry about the margin $\Delta(>0)$
- ICH (E9) gives some general guideline but still the solution is elusive and far from consensus

THREE-ARM NI TRIAL



Issues to consider in 3-arm trial

- Ethical Concern: Not all trials can have a placebo arm
- There is the added difficulty of evaluating **two distinct co-objectives** (NI and AS)
- It may result in a large trial which is Infeasible

Methods Development in NI trial

- For 3-arm trial Pigeot et al. (2003) first proposed a method of choosing “ Δ ” as a *fraction of difference* between R and P in the classical setup under the assumption of homogeneity and normality
- It avoids directly specifying a Δ upfront with a clever trick which is also known as “*effect retention*” approach

BASIC LINEAR MODEL

Consider a one-way fixed effect linear model,

$$X_{li} = \mu_l + e_{li}, \quad l = E, R, P, \quad i = 1, \dots, n_l$$

WLOG we assume larger mean implies better efficacy.

The NI testing problem at hand, $H_0 : \mu_E - \mu_R \leq -\Delta$ vs. $\mu_E - \mu_R > -\Delta$
where $-\Delta = f(\mu_R - \mu_P)$ is a negative fraction in $(-1, 0)$. Hence

$$H_0 : \mu_E - \mu_R \leq f(\mu_R - \mu_P) \text{ vs. } H_1 : \mu_E - \mu_R > f(\mu_R - \mu_P),$$

after some algebra and putting, $\theta = 1 + f$ ($f = 0 \Rightarrow$ superiority)

$$H_0 : \mu_E - \theta\mu_R - (1 - \theta)\mu_P \leq 0 \text{ vs. } H_1 : \mu_E - \theta\mu_R - (1 - \theta)\mu_P > 0$$

13

Note:- This approach makes the assumption $\mu_R - \mu_P \geq 0$

CONTINUED ...

- Pigeot et al. argued that one first must reject the AS null hypothesis, AS: $K_0 : \mu_R \leq \mu_P$ vs. $K_1 : \mu_R > \mu_P$
- However due to the hierarchical ordering of K_0 and H_0 no α adjustment is needed for NI testing
- Test for NI $\rightarrow \psi(\hat{\mu}) = \bar{X}_E - \theta \bar{X}_R - (1 - \theta) \bar{X}_P$, $\hat{\mu} = (\bar{X}_E, \bar{X}_R, \bar{X}_P)'$

$$\widehat{\text{var}}(\psi(\hat{\mu})) = \hat{\sigma}^2 \left(\frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P} \right), \quad \hat{\sigma}^2 = \frac{(n_E - 1)S_E^2 + (n_R - 1)S_R^2 + (n_P - 1)S_P^2}{n_E + n_R + n_P - 3}$$

- Large sample test,
- $$T = \frac{\bar{X}_E - \theta \bar{X}_R - (1 - \theta) \bar{X}_P}{\hat{\sigma} \sqrt{\left\{ \frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P} \right\}}}$$

where under H_0 , $\nu = n_E + n_R + n_P - 3$ & NI is claimed if $T > t_{1-\alpha, \nu}$

- Using the duality of testing and interval estimation the also provided Fieller's confidence interval for the ratio type estimate

POWER AND SAMPLE SIZE

Pigeot et al. also derived the power function based on t_ν distribution $\rightarrow K[\psi(\mu)] = \Pr[T < t_{\nu,1-\alpha} | \psi(\mu), \sigma^2], K[\psi(\mu) = 0] = \alpha$

- For example consider $-\Delta = -0.2(\mu_R - \mu_P) \rightarrow \theta = 1 + f = 0.8$, so the new treatment should retain 80% or more of the active control effect over placebo (in current trial)
- They noted, $\Theta = \frac{\mu_E - \theta\mu_R - (1 - \theta)\mu_P}{\sigma\sqrt{\left\{\frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P}\right\}}}$ (non-centrality parameter)

$$\Theta = \frac{\mu_E - \mu_P - \theta(\mu_R - \mu_P)}{\sigma\sqrt{\frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P}}} = \frac{\left(\frac{\mu_E - \mu_P}{\mu_R - \mu_P} - \theta \right) \times (\mu_R - \mu_P)}{\sigma\sqrt{\frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P}}}$$

assuming $\sigma = \varepsilon(\mu_R - \mu_P)$, i.e. fraction of the difference in means of the reference and the placebo

- They provided power as a function of ε . Note Δ is not specified, but % of effect E must retain is hypothesized

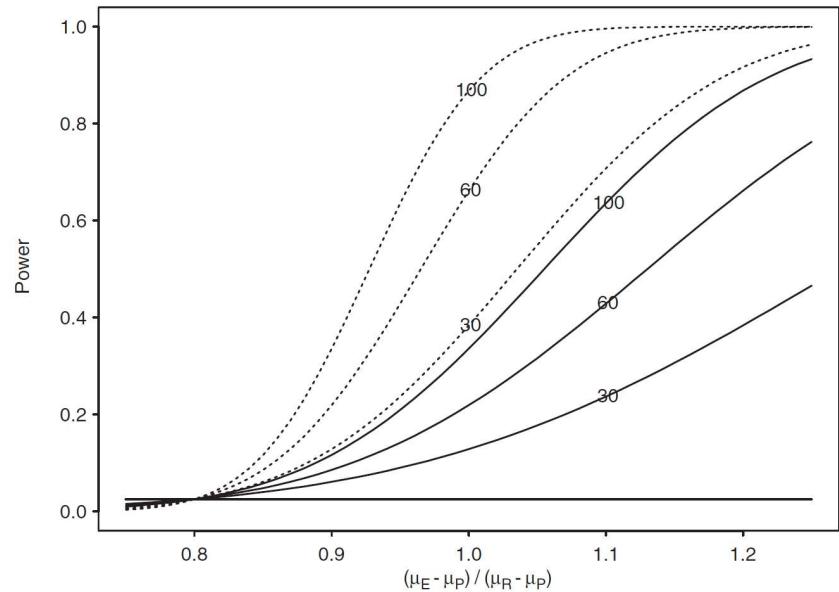


Figure 1. Power functions for $\theta = 0.8$, $\alpha = 0.025$, $n_E = n_R = n_P = n = 30, 60, 100$, and $\varepsilon = 1$ (solid line) and $\varepsilon = 0.5$ (dashed line).

- Power increases with, $n \uparrow$ & decreases as $\varepsilon \uparrow$
- Note, this is based on equal allocation on each arm

Allocation	$(\mu_E - \mu_P) / (\mu_R - \mu_P)$					
	0.85	0.90	0.95	1.00	1.05	1.10
1 : 1 : 1	5276 15828	1320 3960	587 1761	331 993	212 636	148 444
2 : 2 : 1	5401 13503	1351 3378	601 1503	339 848	217 543	151 378
3 : 2 : 1	6532 13065	1634 3269	727 1455	410 821	263 527	183 366
total sample sizes N						

- Equal allocation may not provide optimal power in NI
- Koch and Tangen (1999) suggested the difference between R and E is expected to be much smaller than the difference of both treatments compared to placebo
- It is also not very ethical to put large n on placebo
- Pigeot et al. also suggested strategy to find out optimal allocation ratio, i.e. minimum total N for fixed power

FRACTIONAL MARGIN APPROACH: BINARY OUTCOME

Consider a three-arm trial with $X_l \sim Bin(n_l, \pi_l)$ for $l = E, R, P$
Kieser and Friede (2007) defined NI testing in terms of RD

$$H_0 : \pi_E - \pi_R \leq -\Delta \text{ vs. } H_a : \pi_E - \pi_R > -\Delta, \text{ for } \Delta > 0$$

where $-\Delta = f(\pi_R - \pi_P)$ is a negative fraction in $(0,1)$. Hence

$$H_0 : \pi_E - \pi_R \leq f(\pi_R - \pi_P) \text{ vs. } H_a : \pi_E - \pi_R > f(\pi_R - \pi_P),$$

after similar algebra and putting $\theta=1+f$

$$H_0 : \pi_E - \theta\pi_R - (1-\theta)\pi_P \leq 0 \text{ vs. } H_1 : \pi_E - \theta\pi_R - (1-\theta)\pi_P > 0$$

Note this approach also makes the assumption $\pi_R - \pi_P > 0$.
This is very much parallel to Pigeot et al. (2003) approach

BINARY OUTCOME

- Test for NI $\rightarrow \hat{\psi} = \hat{\pi}_E - \theta \hat{\pi}_R - (1 - \theta) \hat{\pi}_P$ with $\hat{\pi}_i = X_i/n_i, i = P, R, E.$ and variance $\sigma^2(\hat{\psi}) = \pi_E(1 - \pi_E)/n_E + \theta^2 \pi_R(1 - \pi_R)/n_R + (1 - \theta)^2 \pi_P(1 - \pi_P)/n_P$
- They proposed Large sample ML test (normal $T = \hat{\psi}/\bar{\sigma}(\hat{\psi})$)
- Restricted ML, under $H_0: \pi_E - \theta \pi_R - (1 - \theta) \pi_P = 0$
- As well an exact test procedure based on nominal α

$$\alpha_T = \sum_{x_P=0}^{n_P} \sum_{x_R=0}^{n_R} \sum_{x_E=0}^{n_E} L_{H_0}(x_P, x_R, x_E) \times I_{\{T(x_P, x_R, x_E) \geq z_{1-\alpha}\}}$$

$$L_{H_0}(x_P, x_R, x_E) = P_{H_0}(X_P = x_P, X_R = x_R, X_E = x_E | \pi_P, \pi_R, \pi_E)$$

$$= \binom{n_P}{x_P} \binom{n_R}{x_R} \binom{n_E}{x_E} \pi_P^{x_P} (1 - \pi_P)^{n_P - x_P} \pi_R^{x_R} (1 - \pi_R)^{n_R - x_R} \pi_E^{x_E} (1 - \pi_E)^{n_E - x_E}$$

- They reported, Wald-type ML test requires larger sample size compared to RML based test. Exact test is best, but computationally intensive

Table I. Sample size and corresponding exact power for three sample size formulae when assessing the non-inferiority test problem (2) with the RML test (see text).

λ_P	λ_R	λ_E	θ	π_P	π_R	Method 1		Method 2		Method 3	
						n_{00}	Power	n_{01}	Power	n_{11}	Power
1	1	1	0.60	0.10	0.50	303	0.7816	309	0.7894	327	0.8114
				0.10	0.70	138	0.8209	135	0.8151	123	0.7799
				0.10	0.90	66	0.8996	54	0.8305	30	0.4802
				0.30	0.70	330	0.8262	318	0.8117	294	0.7803
				0.30	0.90	120	0.9073	99	0.8392	63	0.6289
				0.50	0.90	243	0.8954	213	0.8495	150	0.6959
	0.80	0.80	0.80	0.10	0.70	624	0.8287	606	0.8174	570	0.7938
				0.10	0.90	234	0.9070	201	0.8557	138	0.6900
				0.30	0.90	390	0.8952	345	0.8539	255	0.7268
				0.50	0.90	792	0.8774	726	0.8474	579	0.7552
1	2	2	0.60	0.10	0.50	265	0.7781	270	0.7859	285	0.8070
				0.10	0.70	120	0.8136	115	0.7996	105	0.7590
				0.10	0.90	60	0.9000	50	0.8471	30	0.5741
				0.30	0.70	300	0.8203	290	0.8073	270	0.7784
				0.30	0.90	110	0.8983	95	0.8425	65	0.6825
				0.50	0.90	235	0.8972	210	0.8606	155	0.7347
	0.80	0.80	0.80	0.10	0.70	525	0.8281	510	0.8169	480	0.7933
				0.10	0.90	195	0.9019	170	0.8542	120	0.7012
				0.30	0.90	335	0.8923	300	0.8551	225	0.7339
				0.50	0.90	690	0.8757	635	0.8469	515	0.7622
1	2	3	0.60	0.10	0.50	246	0.7726	252	0.7815	264	0.8000
				0.10	0.70	108	0.8054	108	0.8054	102	0.7837
				0.10	0.90	54	0.9045	42	0.8012	30	0.6187
				0.30	0.70	282	0.8178	276	0.8097	258	0.7831
				0.30	0.90	102	0.8966	90	0.8570	66	0.7240
				0.50	0.90	222	0.8985	204	0.8731	156	0.7694
	3	0.8	0.8	0.10	0.70	498	0.8341	486	0.8251	462	0.8056
				0.10	0.90	180	0.9167	156	0.8736	114	0.7439
				0.30	0.90	306	0.8973	282	0.8717	222	0.7797
				0.50	0.90	648	0.8821	606	0.8602	504	0.7910

Unequal allocation leads to smaller total sample size

ISSUE OF MULTIPLICITY IN TESTING NI (AND AS)

- They also studied the effect of pre AS testing on the loss of power as a two-step procedure (AS → NI)

λ_P	λ_R	λ_E	π_R	ε	n_p	Power		
						Test for superiority, R versus P		Test for non-inferiority, E versus R
						Multiple test procedure		
1	1	1	0.50	1.00	3791	1.0000	0.7948	0.7948
			0.49	1.05	2299	1.0000	0.8036	0.8036
			0.48	1.11	1540	1.0000	0.7896	0.7896
			0.47	1.18	1102	1.0000	0.7990	0.7990
			0.46	1.25	827	1.0000	0.7964	0.7964
			0.45	1.33	642	0.9994	0.8022	0.8016
			0.44	1.43	512	0.9958	0.7924	0.7884
			0.43	1.54	418	0.9761	0.7914	0.7680
			0.42	1.67	347	0.9096	0.7964	0.7114
			0.41	1.82	292	0.7985	0.7928	0.6076
			0.40	2.00	249	0.6478	0.7917	0.4729
1	2	2	0.50	1.00	1916	1.0000	0.8018	0.8018
			0.49	1.05	1162	1.0000	0.7905	0.7905
			0.48	1.11	779	1.0000	0.7921	0.7921
			0.47	1.18	557	1.0000	0.7952	0.7952
			0.46	1.25	418	0.9998	0.8076	0.8074
			0.45	1.33	324	0.9957	0.7978	0.7939
			0.44	1.43	259	0.9708	0.7935	0.7666
			0.43	1.54	211	0.8951	0.8005	0.7065
			0.42	1.67	175	0.7670	0.7934	0.5890
			0.41	1.82	147	0.6217	0.7877	0.4617
			0.40	2.00	126	0.4786	0.7835	0.3458
1	4	4	0.50	1.00	979	1.0000	0.7959	0.7959
			0.49	1.05	594	1.0000	0.7933	0.7933
			0.48	1.11	398	1.0000	0.7963	0.7963
			0.47	1.18	285	0.9997	0.8022	0.8019
			0.46	1.25	213	0.9905	0.7946	0.7862
			0.45	1.33	166	0.9492	0.8027	0.7584
			0.44	1.43	132	0.8466	0.7866	0.6581
			0.43	1.54	108	0.7060	0.7862	0.5480
			0.42	1.67	90	0.5614	0.7968	0.4353
			0.41	1.82	75	0.4136	0.7974	0.3204
			0.40	2.00	64	0.3061	0.7803	0.2288

The sample size was calculated with the formula using both the variance under the null and the alternative hypothesis to achieve a power of $1 - \beta = 0.80$. The power was estimated by simulations with 10 000 replications.

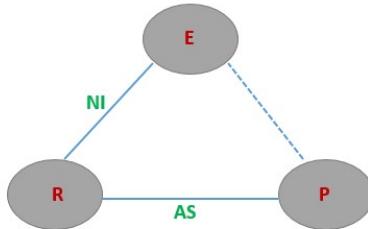
$$\varepsilon = (\pi_E - \pi_P) / (\pi_R - \pi_P)$$

OTHER DEVELOPMENTS

- Three arm trial when NI hypothesis is defined in the ratio scale (rather than difference)
- Three arm trial for continuous but non-normal distribution
- For binary data RD is not the only function of interest
- FDA's recently published guideline indicates RR, OR and NNT etc.
- Three-arm trial for *count* type outcome (Poisson and Negative Binomial)
- Multi arm trial for multiple experimental treatments
- Alternative to *fraction margin* approach ...a) Fixed Margin and b) Group-sequential 3-arm trial(!)

FIXED MARGIN APPROACH

The *fraction margin* approach of Pigeot et al. is not out of controversy. It essentially combines two tests in one,



$$\text{NI: } H_0 : \mu_E - \mu_R \leq -\Delta \text{ vs. } H_1 : \mu_E - \mu_R > -\Delta$$

$$\text{AS: } K_0 : \mu_R - \mu_P \leq \Delta \text{ vs. } K_1 : \mu_R - \mu_P > \Delta$$

- Under the assumption $\mu_R - \mu_P > 0$, it somewhat downplays AS testing, however use it to construct Δ ($-\Delta = f(\mu_R - \mu_P)$)
- The NI with AS is only established when H_0 and K_0 are jointly rejected. This approach of Hida and Tango, 2009 (modified by Kwong et al., 2011) is known as *fixed margin approach*
- Hida and Tango combined H_1 and K_1 to get,

$$\mu_P < \mu_R - \Delta < \mu_E \Rightarrow \mu_P + \Delta < \mu_R < \mu_E + \Delta,$$

CONTINUED

NI: $H_0 : \mu_E - \mu_R \leq -\Delta$ vs. $H_1 : \mu_E - \mu_R > -\Delta$

AS: $K_0 : \mu_R - \mu_P \leq \Delta$ vs. $K_1 : \mu_R - \mu_P > \Delta$

Caveat: Same Δ is used in both NI and AS testing, this is too liberal. NI margin (Δ) should be much lower than previously detected active treatment (R) effect. Kwong et al. (2011) *modified* it as,

NI: $H_0 : \mu_E - \mu_R \leq -r\Delta$ vs. $H_1 : \mu_E - \mu_R > -r\Delta$

AS: $K_0 : \mu_R - \mu_P \leq \Delta$ vs. $K_1 : \mu_R - \mu_P > \Delta$

$$\mu_P + \Delta < \mu_R < \mu_E + r\Delta, \text{ where } 0 < r \leq 1.$$

- Modified Hida and Tango (MHT) approach is extended for both continuous and binary endpoint (RD only)
- NI margin in MHT approach should be *fixed* apriori

JOINT TESTING

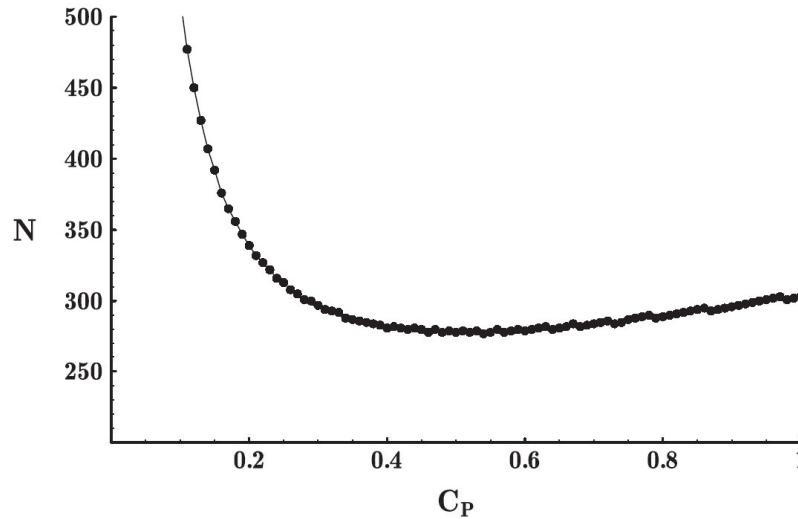
$$T_1 = \frac{\bar{X}_E - \bar{X}_R + \Delta}{\hat{\sigma}_1 \sqrt{\frac{1}{n_E} + \frac{1}{n_R}}}, \quad T_2 = \frac{\bar{X}_R - \bar{X}_P - \Delta}{\hat{\sigma}_2 \sqrt{\frac{1}{n_R} + \frac{1}{n_P}}}.$$

- Usual t-test is used and since we are looking for joint rejection of H_0 and K_0 by the Intersection-Union test (IU), no α adjustment is needed
- For MHT in T_1 , Δ needs to be replaced by $r \cdot \Delta$
- NI with assay sensitivity can be claimed if and only if

$$T_1 > t_{\alpha/2}(n_E + n_R - 2) \quad \text{and} \quad T_2 > t_{\alpha/2}(n_R + n_P - 2).$$

- Power = $\Pr\{T_1 > t_{\alpha/2}(n_E + n_R - 2) \cap T_2 > t_{\alpha/2}(n_R + n_P - 2) | H_1, K_1\}$
- Note T_1 and T_2 , are correlated, if $n_E : n_R : n_P = 1 : C_R : C_P$

$$1 - \beta = \Pr\{T_1 \geq z_{\alpha/2} \cap T_2 \geq z_{\alpha/2} \mid (T_1, T_2) \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})\}, \quad \rho = -\sqrt{\frac{C_P}{(1+C_R)(C_R+C_P)}}$$



Profile of the total sample size N for C_P
required for $1 - \beta = 80\%$ power at $\alpha/2 = 0.025$ level,
where $(\mu_E, \mu_R, \mu_P, \sigma) = (1, 1, 0, 1)$ and $\Delta = 0.4$.

- Fixed effect approach typically requires slightly larger N
- Unbalanced design leads to smaller sample size
- It asks for substantial superiority (Δ) for R over P
- IUT principle preserves α but may produce biased test
- The idea of changing the definition of type-I error (e.g. Average Testing Error) is interesting but regulatory agencies guideline on this is not clear
- Hida & Tango also published paper on Risk Difference

FIXED MARGIN

- The correlation is determined by the sampling ratio, but the two test statistics are always negatively correlated
- Can indirectly demonstrate the superiority of the experimental intervention relative to the placebo if H_0^{AS} and H_0^{NI} are rejected, without direct comparison of the experimental intervention to the placebo.
- Can reject H_0^{NI} when $\mu_C - \Delta < \mu_P < \mu_C$ is true

○ Ochiai et al. (2017) proposed a group sequential approach for both. Please recapitulate basic model,

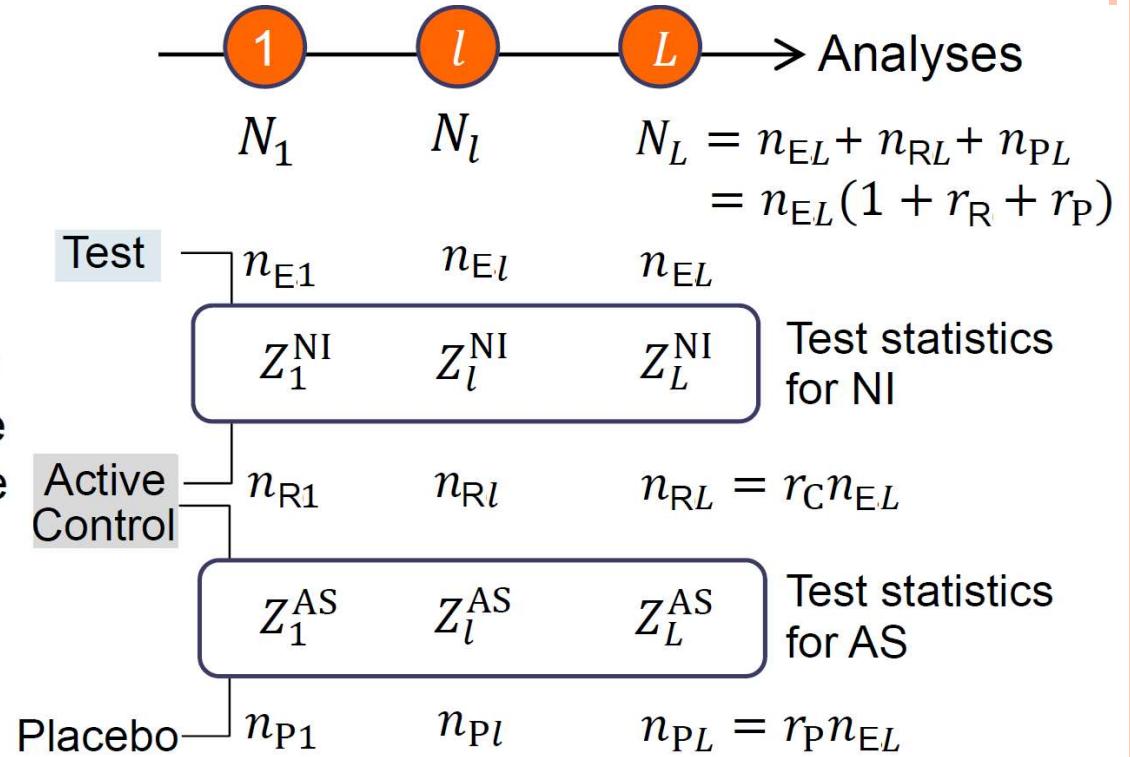
$$X_{li} = \mu_l + e_{li}, \quad l = E, R, P, \quad i = 1, \dots, n_L \quad \text{and} \quad V(X_{li}) = \sigma^2$$

FRACTION MARGIN

- The two test statistics are positively or negatively correlated depending on sampling ratio and fraction
- Can demonstrate $\mu_T > \mu_P$ irrespective of θ since $\mu_T - \mu_P > \theta(\mu_C - \mu_P) > 0$ if both null hypotheses H_0^{AS} and H_0^{NI} are rejected
- Cannot reject H_0^{NI} when $\mu_C - \Delta < \mu_P < \mu_C$ is true- whether the fraction approach can allow demonstration of noninferiority of the experimental intervention to the control intervention is questionable under $\mu_C - \Delta < \mu_P$.

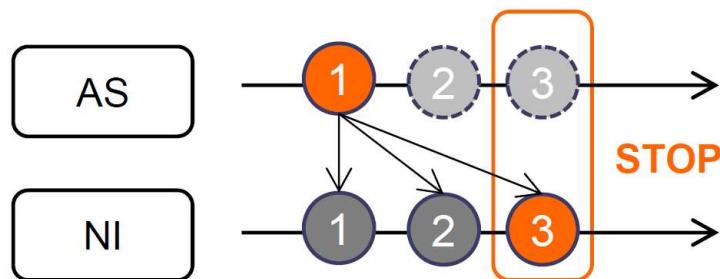
GST→

- Each test statistic normally distributed for large sample
- $2L$ test statistics $2L$ -variate normally distributed



- Note, assumes equal allocation in each arm to begin with
- Two different decision making frameworks are proposed
- Can be applied to both fixed and fraction margin approach
- Essentially it is a not direct competitor as a 3rd approach

GST DESIGNS: DF-A



Stopping rule for DF-A

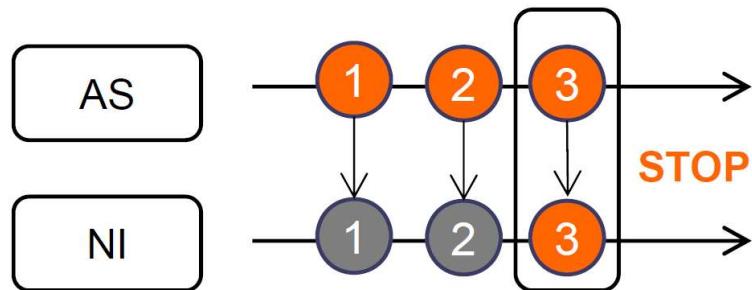
At the l th analysis ($l = l', \dots, L - 1$)
 if $Z_{l'}^{\text{AS}} > c_{l'}^{\text{AS}}(\alpha)$ for some l' ($1 \leq l' \leq l$) (H_0^{AS} has been rejected), and $Z_l^{\text{NI}} > c_l^{\text{NI}}(\alpha)$, then reject H_0^{NI} and stop the trial
 otherwise, continue the trial

At the L th analysis
 if $Z_{l'}^{\text{AS}} > c_{l'}^{\text{AS}}(\alpha)$ for some l' ($1 \leq l' \leq L$), and $Z_L^{\text{NI}} > c_L^{\text{NI}}(\alpha)$,
 then reject H_0^{NI}
 otherwise do not reject H_0^{NI}

- NI is evaluated only after the AS is demonstrated.
- A trial stops if the AS and the NI are achieved at any interim analysis, i.e., **not necessarily simultaneously**.
- If AS is demonstrated but NI is not, then the trial continues and subsequent hypothesis testing is repeatedly conducted only for NI until the NI is demonstrated.

$c_l^{\text{AS}}(\alpha)$ and $c_l^{\text{NI}}(\alpha)$ are critical values separately prespecified, using any GS methods

GST DESIGNS: DF-B



Stopping rule for DF-B

At the l th analysis ($l = 1, \dots, L - 1$)
 if $Z_l^{AS} > c_l^{AS}(\alpha)$ and $Z_l^{NI} > c_l^{NI}(\alpha)$, then reject H_0^{AS} and
 H_0^{NI} and stop the trial
 otherwise, continue the trial

At the L th analysis
 if $Z_L^{AS} > c_L^{AS}(\alpha)$ and $Z_L^{NI} > c_L^{NI}(\alpha)$, then reject H_0^{AS} and
 H_0^{NI} and stop the trial
 otherwise do not reject H_0^{AS} and H_0^{NI}

- A special case of DF-A, NI is evaluated only after the AS is demonstrated,
- A trial stops only if AS and NI are demonstrated at the same interim analysis **simultaneously**.
- Otherwise, the trial will continue and the subsequent hypothesis testing is repeatedly conducted for both AS and NI until simultaneous significance is reached.

$c_l^{AS}(\alpha)$ and $c_l^{NI}(\alpha)$ are critical values separately prespecified, using any GS methods

GROUP-SEQUENTIAL DESIGN

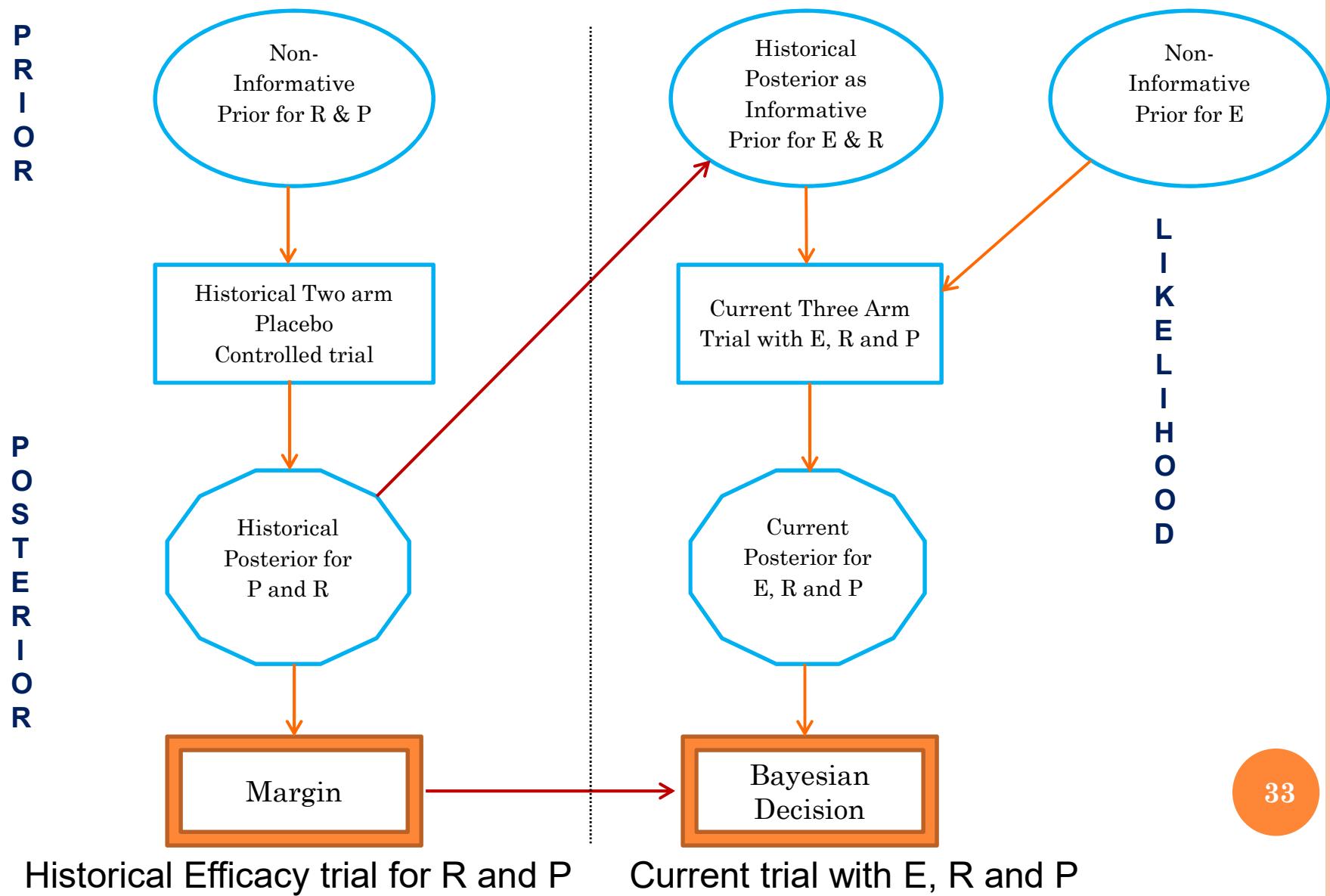
	Fixed Margin	Fraction
DF-A	<ul style="list-style-type: none">● Flexible, and slightly powerful than DF-B● Allows for dropping placebo group if AS is demonstrated at the interim	<ul style="list-style-type: none">● Flexible, and slightly powerful than DF-B● Not allow for dropping placebo group as the test statistics for the NI includes the amount of \bar{Y}_{Pm}.
DF-B	<ul style="list-style-type: none">● Simple, but less powerful than DF-A● Not allow for dropping of the placebo group even if AS is demonstrated at the interim	<ul style="list-style-type: none">● Simple, but less powerful than DF-A
○ GSDs offer the possibility to stop a trial early when evidence is overwhelming and thus offers efficiency		
○ There are no major differences in both MSS and ASN between DF-A and DF-B for the fixed margin and the fraction approaches, though DF-A is little more powerful than DF-B		
○ Also proposed Delayed analysis of NI, adaptive allocation of a		

Bayesian Analysis of 3-Arm Non-Inferiority Trial

BAYESIAN PHILOSOPHY: A BRIEF PRIMER

- In Statistics there are two schools of thought
 - 1. Classical/Frequentists
 - 2. Bayesian
- In classical paradigm we work within the framework of likelihood to make Inference on parameter (e.g. mean= μ)
- In Bayesian regime we add an extra element called Prior
- A Bayesian way of drawing Inferences are following
 - 1. Choose a parameter of interest (e.g. mean= μ , proportion= θ)
 - 2. Put a prior distribution on parameter, which may come from *expert opinion* or from a *similar historical study*
 - 3. In the absence of above prior we put Vague or Flat or Non-Informative prior
 - 4. Construct Likelihood based on current data
 - 5. Multiply Prior and Likelihood to get the Posterior distribution
 - 6. Make inference based on Posterior distribution
- In the CER context prior information is guaranteed

INFORMATION FLOW IN BAYESIAN DESIGN



BAYESIAN FORMULATION (FRACTION MARGIN)

- Ghosh et al. (2011) proposed the first Bayesian approach for 3-arm trial. For two-arm NI Gamalo et al. published few interesting papers. Testing problem in hand

$$H_0 : \mu_E - \theta\mu_R - (1-\theta)\mu_P \leq 0 \text{ vs. } H_1 : \mu_E - \theta\mu_R - (1-\theta)\mu_P > 0$$

- A Bayesian setup involves some “prior” specification

$$\begin{aligned} \mu_l | \sigma_l^2 &\sim N(\mu_{0l}, \sigma_l^2 / \kappa_{0l}), & p(\mu_l | \sigma_l^2, \mu_{0l}, \kappa_{0l}) &\propto \exp\left(-\frac{\kappa_{0l}}{2\sigma_l^2}(\mu_l - \mu_{0l})^2\right) \\ \sigma_l^2 &\sim \text{Inv-gamma}(v_{0l}/2, \sigma_{0l}^2 v_{0l}/2), \quad l \in \{E, R, P\}, \text{ i.e.} & p(\sigma_l^2 | v_{0l}, \sigma_{0l}^2) &\propto \exp\left(-\frac{\sigma_{0l}^2 v_{0l}}{2\sigma_l^2}\right) \left(\frac{1}{\sigma_l^2}\right)^{\frac{v_{0l}+1}{2}} \end{aligned}$$

- Posterior: $p(\mu_E, \mu_P, \mu_R | X) \propto I\{\mu_R > \mu_P\} \prod_{l \in \{E, P, R\}} t_{v_{nl}}(\mu_l | \mu_{nl}, \sigma_{nl}),$

- Test: $P\left(H_1 : \frac{\mu_E - \mu_P}{\mu_R - \mu_P} > \theta | \text{Data}\right) > R_{\text{NI}} \rightarrow \widehat{P}\left(H_1 : \frac{\mu_E - \mu_P}{\mu_R - \mu_P} > \theta | \text{Data}\right) = \frac{1}{T} \sum_{l=1}^T I\left(\frac{\mu_E^l - \mu_P^l}{\mu_R^l - \mu_P^l} > \theta\right)$

Choice of R_{NI} is user driven (e.g. 0.5) (higher posterior prob)

BAYESIAN FORMULATION (FIXED MARGIN)

- Ghosh et al. (2016) proposed Bayesian fixed margin version for joint testing of NI and AS
- We will consider first the known variance case. So the parameter space is $(\mu_l, \sigma_l^2), l \in \{E, R, P\}$
- To construct NI margin consider 1st historical placebo controlled trial (R and P),

$$X_{Pj}^H | \mu_{PH}, \sigma_{PH}^2 \sim N(\mu_{PH}, \sigma_{PH}^2), j = 1, \dots, n_P^H.$$

$$X_{Rj}^H | \mu_{RH}, \sigma_{RH}^2 \sim N(\mu_{RH}, \sigma_{RH}^2), j = 1, \dots, n_R^H$$

- Since historical trial is not superseded by another trial we assume non-informative prior $\pi(\mu_{iH}) \propto 1$, for $i = R, P$.
- Historical posteriors are,

$$\mu_{PH} | \bar{X}_P^H, \sigma_{PH}^2 \sim N \left(\bar{X}_P^H, \frac{\sigma_{PH}^2}{n_P^H} \right); \mu_{RH} | \bar{X}_R^H, \sigma_{RH}^2 \sim N \left(\bar{X}_R^H, \frac{\sigma_{RH}^2}{n_R^H} \right)$$

MARGIN COMPUTATION

- It is easy to show,

$$\mu_{RH} - \mu_{PH} | \bar{X}_P^H, \bar{X}_R^H, \sigma_{PH}^2, \sigma_{RH}^2 \sim N \left(\bar{X}_R^H - \bar{X}_P^H, \frac{\sigma_{RH}^2}{n_R^H} + \frac{\sigma_{PH}^2}{n_P^H} \right)$$

- NI margin is obtained by computing,

$$\begin{aligned} & P(\mu_{RH} - \mu_{PH} \geq \Delta^B | \bar{X}_P^H, \bar{X}_R^H, \sigma_{PH}^2, \sigma_{RH}^2) \\ = & P \left(Z \geq \frac{\Delta^B - (\bar{X}_R^H - \bar{X}_P^H)}{\sqrt{\frac{\sigma_{RH}^2}{n_R^H} + \frac{\sigma_{PH}^2}{n_P^H}}} | \bar{X}_P^H, \bar{X}_R^H, \sigma_{PH}^2, \sigma_{RH}^2 \right) = 1 - \alpha/2. \end{aligned}$$

- With a specified preservation level ($0 < \lambda < 1$), Frequentist NI margin, $\delta_{NI}^F = (1-\lambda)\Delta^F = (1-\lambda) \left(\bar{X}_R^H - \bar{X}_P^H - z_{1-\alpha/2} \sqrt{\frac{\sigma_{RH}^2}{n_R^H} + \frac{\sigma_{PH}^2}{n_P^H}} \right)$
- The AS margin is given by, $\delta_{AS}^B = \delta_{NI}^B / r = \frac{(1-\lambda)\Delta^B}{r}$
- With $r = 1 - \lambda \rightarrow$ NI Margin = AS Margin (HT) and $r > 1 - \lambda$ is Modified HT with the restriction $(1 - \lambda) < r \leq 1$
- The choice λ and r should be guided by clinical practice

POSTERIOR DISTRIBUTION (CURRENT TRIAL)

- Current trial with three arms E , R and P . We choose non-informative prior for E , and historical posterior for R and P as prior (informative).
- Note the implicit assumption of Constancy
- Posterior is given by,

$$\mu_E \propto N\left(\bar{X}_E, \frac{\sigma_E^2}{n_E}\right); \mu_l \propto N\left(\tilde{\mu}_l, \tilde{\sigma}_l^2\right) \text{ for } l \in \{R, P\},$$

$$\tilde{\mu}_l = \tilde{\sigma}_l^2 \left(\frac{n_l \bar{X}_l}{\sigma_l^2} + \frac{n_l^H \bar{X}_l^H}{\sigma_{lH}^2} \right) \text{ and } \tilde{\sigma}_l^2 = \left(\frac{n_l}{\sigma_l^2} + \frac{n_l^H}{\sigma_{lH}^2} \right) \text{ for } l \in \{R, P\}.$$

- And the distribution of test statistic,

$$(\mu_E - \mu_R) | \bar{X}_E, \bar{X}_R, \sigma_E^2, \sigma_R^2 \propto N\left(\bar{X}_E - \tilde{\mu}_R, \frac{\sigma_E^2}{n_E} + \tilde{\sigma}_R^2\right)$$

$$(\mu_R - \mu_P) | \bar{X}_R, \bar{X}_P, \sigma_R^2, \sigma_P^2 \propto N\left(\tilde{\mu}_R - \tilde{\mu}_P, \tilde{\sigma}_R^2 + \tilde{\sigma}_P^2\right)$$

- Clearly they are not independent

CONTINUED...

- And the variance-covariance matrix is $\Sigma = \begin{bmatrix} \sigma_E^2 + \tilde{\sigma}_R^2 & -\tilde{\sigma}_R^2 \\ -\tilde{\sigma}_R^2 & \tilde{\sigma}_R^2 + \tilde{\sigma}_P^2 \end{bmatrix}$
- Decision criterion for deciding that the experiential treatment is non-inferior to the AC with assay sensitivity if,

$$\begin{aligned} p(\delta_{NI}^B, \lambda, r, 1 - \alpha/2) &= P(\mu_E - \mu_R \geq -\delta_{NI}^B \cap \mu_R - \mu_P \geq \delta_{NI}^B/r | X_E, X_R, X_P) \\ &= P\left[Z_1 \geq \frac{-\delta_{NI}^B - (\bar{X}_E - \tilde{\mu}_R)}{\frac{\sigma_E^2}{n_E} + \tilde{\sigma}_R^2} \cap Z_2 \geq \frac{\delta_{NI}^B/r - (\tilde{\mu}_R - \tilde{\mu}_P)}{\tilde{\sigma}_R^2 + \tilde{\sigma}_P^2}\right] \\ &\geq p^* \end{aligned}$$

- Where p^* is pre-specified from the level (α) condition (such as 0.975 or 0.95.) and/or via Bayesian Calibration
- Note we are in the realm of multiple testing (NI and AS) with negatively correlated test statistic(s)
- However under intersection–union (IU) testing setup no multiplicity adjustment is needed (have other issues)

BAYESIAN FORMULATION (UNKNOWN VARIANCE)

- Data from historic trial, $X_{Pj}^H | \mu_{PH}, \sigma_{PH}^2 \sim N(\mu_{PH}, \sigma_{PH}^2), j = 1, \dots, n_P^H$.

$$X_{Rj}^H | \mu_{RH}, \sigma_{RH}^2 \sim N(\mu_{RH}, \sigma_{RH}^2), j = 1, \dots, n_R^H$$

- Non informative Priors: $\pi(\mu_{iH}) \propto 1, \pi(\sigma_{iH}^2) \propto \sigma_{iH}^{-2}$ for $i = R, P$.
- With some algebra marginal historical posteriors are,

$$\mu_{lH} | \bar{X}_l^H \sim t_{n_l^H - 1} \left(\bar{X}_l^H, \frac{s_{lH}^2}{n_l^H} \right); s_{lH}^2 = \frac{1}{n_l^H - 1} \sum_{j=1}^{n_l^H} (X_{jl}^H - \bar{X}_l^H)^2 \text{ where } l \in \{R, P\}$$

- Unfortunately the difference $(\mu_{RH} - \mu_{PH} | Data)$ does not have a closed form solution and we resort to a sampling based approach to calculate lower limit of the $100(1-\alpha)\%$ credible interval using the equation,

$$P(\mu_{RH} - \mu_{PH} \geq \Delta^B | \bar{X}_P^H, \bar{X}_R^H) = 1 - \alpha/2.$$

- Δ^B is the Bayesian NI margin. Frequentists margin Δ^F 39 can be obtained using t-distribution.

POSTERIOR DISTRIBUTION (CURRENT TRIAL)

- Like the known variance case, we choose non-informative prior for E , and historical posterior as prior for R and P in the current trial (informative).

$$\pi(\mu_E) \propto 1, \pi(\sigma_E^2) \propto \sigma_E^{-2}$$

- For the E-arm prior s are

- Following standard algebra, conditional posteriors

$$\begin{cases} \mu_E | \sigma_E^2, \bar{X}_E \propto N\left(\bar{X}_E, \frac{\sigma_E^2}{n_E}\right) \\ \sigma_E^2 | \bar{X}_E \propto IG\left(\frac{n_E - 1}{2}, \frac{n_E - 1}{2} s_E^2\right) \end{cases}$$

- Marginal posterior,

$$\mu_E | \bar{X}_E \sim t_{n_E - 1} \left(\bar{X}_E, \frac{s_E^2}{n_E} \right); s_E^2 = \frac{1}{n_E - 1} \sum_{j=1}^{n_E} (X_{jE} - \bar{X}_E)^2$$

- Historical posteriors for R and P is taken as prior,

$$\pi(\mu_l | \sigma_l^2) \propto N\left(\bar{X}_l^H, \frac{\sigma_l^2}{n_l}\right); \pi(\sigma_l^2) \propto IG\left(\frac{n_l^H - 1}{2}, \frac{n_l^H - 1}{2} s_{lH}^2\right) \text{ where } l \in \{R, P\}$$

CONTINUED...

- With some algebra marginal posterior for R and P in the current trial,

$$\mu_l | \bar{X}_l \propto t_{n_l^*} \left(\mu_l^*, \frac{\psi_l^2}{K_l} \right) \text{ where } l \in \{R, P\}$$

$$\psi_l^2 = \frac{1}{n_l^*} \left[\frac{n_l^H - 1}{2} s_{lH}^2 + (n_l - 1) s_l^2 + \frac{n_l n_l^H}{n_l + n_l^H} (\bar{X}_l - \bar{X}_l^H)^2 \right]$$

$$\mu_l^* = \frac{1}{K_l} (n_l^H \bar{X}_l^H + n_l \bar{X}_l)$$

$$K_l = n_l^H + n_l$$

$$n_l^* = \frac{n_l^H - 1}{2} + n_l$$

- However the difference of t-distribution is not available in closed form, as a result our test statistic $(\mu_E - \mu_R | Data)$ for NI and $(\mu_R - \mu_P | Data)$ for AS cannot be expressed analytically. We use simulation as a solution.

BAYESIAN DECISION RULE

- The general form for the NI with AS decision rule,

$$p(\delta_{NI}^B, \lambda, r, 1 - \alpha/2) = P(\mu_E - \mu_R \geq -\delta_{NI}^B \cap \mu_R - \mu_P \geq \delta_{NI}^B / r | X_E, X_R, X_P) \geq p^*$$

- We sample from marginal posterior of μ_E, μ_R and μ_P (all t). Let's say we draw C many such samples
- We declare NI with AS if,

$$\frac{1}{C} \sum_{i=1}^C I\{\mu_E^i - \mu_R^i \geq -\delta_{NI}^B, \mu_R^i - \mu_P^i \geq \delta_{NI}^B / r\} \geq p^*$$

For a pre-specified p^* , which needs to be chosen from the level condition (we choose 0.975)

- Like the known variance case this procedure also does not require any multiple testing adjustment
- We have used the working correlation estimated using simulated samples (above) to apply appropriate Bayesian calibration

SIMULATION STEPS

1. Specify $n_P^H, n_R^H, \mu_P^H, \mu_R^H, \sigma_{PH}^2, \sigma_{RH}^2, n_E, n_R, n_P, \mu_R$ where $\mu_R^H > \mu_P^H$ in order to generate historical data
2. Generate data $X_P^H \sim N(\mu_P^H, \sigma_{PH}^2)$ $X_R^H \sim N(\mu_R^H, \sigma_{RH}^2)$ and calculate sample mean and variance
3. Use Gibbs sampling to generate posterior samples from $\mu_P^H, \mu_R^H, \sigma_{PH}^2, \sigma_{RH}^2$ for $l = 1, 2, \dots, N$
4. Using this samples calculate posterior mean and standard deviation
5. Calculate $\Delta_l = \bar{X}_{R,l}^H - \bar{X}_{P,l}^H - z_{1-\alpha/2} \sqrt{\frac{\sigma_{RH,l}^2}{n_R^H} + \frac{\sigma_{PH,l}^2}{n_P^H}}$ for l -th iteration and $\bar{\Delta} = \sum_{l=1}^N \Delta_l$ is the MCMC approximation of Δ^B
6. Choose a preservation level λ , and define $\bar{\delta}_{NI}^B$ and $\bar{\delta}_{NI}^F$
7. Also choose a “ r ” to define AS margin
8. Setup $\mu_E = \mu_R - NI + \xi_1$ and $\mu_P = \mu_R - AS - \xi_2$, where ξ_1, ξ_2 are positive constants

SIMULATION STEPS CONTINUED

9. Generate M posterior samples from the current posterior distribution of mean (E, R and P)
10. Calculate the posterior probability,

$$P \left[Z_1 \geq \frac{-\bar{\delta}_{NI}^B - (\bar{X}_E - \tilde{\mu}_{R,m})}{\frac{\sigma_{E,m}^2}{n_E} + \tilde{\sigma}_{R,m}^2} \cap Z_2 \geq \frac{\bar{\delta}_{NI}^B/r - (\tilde{\mu}_{R,m} - \tilde{\mu}_{P,m})}{\tilde{\sigma}_{R,m}^2 + \tilde{\sigma}_{P,m}^2} \right] ; m = 1, \dots, M.$$

The average of these over M iterations give MCMC approximation of the expected value

11. Bayesian Criteria, if

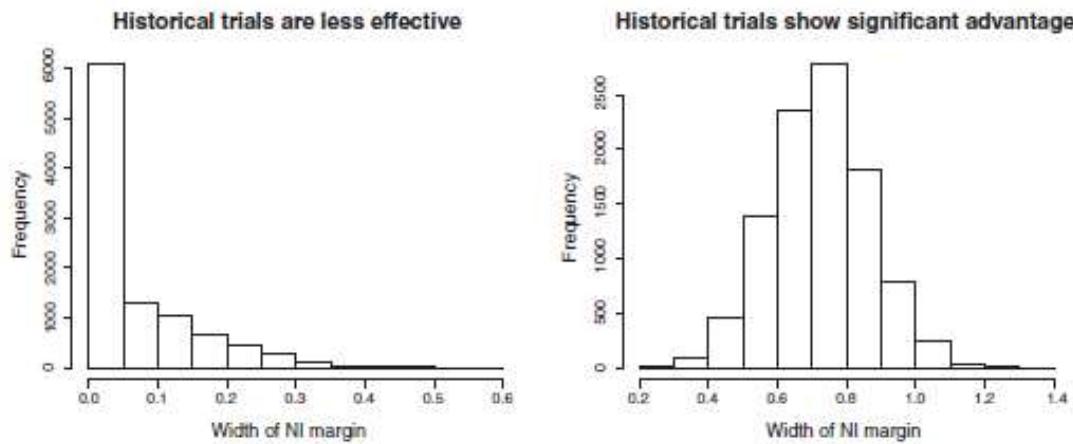
$$E \left[P \left(Z_1 \geq \frac{-\bar{\delta}_{NI}^B - (\bar{X}_E - \tilde{\mu}_{R,m})}{\frac{\sigma_{E,m}^2}{n_E} + \tilde{\sigma}_{R,m}^2} \cap Z_2 \geq \frac{\bar{\delta}_{NI}^B/r - (\tilde{\mu}_{R,m} - \tilde{\mu}_{P,m})}{\tilde{\sigma}_{R,m}^2 + \tilde{\sigma}_{P,m}^2} \right) \right] \geq p^*$$

Then increase the Count by 1; 0 otherwise.

12. Similarly perform the Counts by using frequentist NI and AS margin

SIMULATION SETUP

- A major criticism of selecting NI margin based on R vs. P historical trial is Selection Bias. As very limited failed trials are reported leading to over-estimated effect size
- Hence NI margin may be biased
- To guard against this we consider two situations,
 1. Historical reference is not so strongly established, $p \in [0.01, 0.05]$
 2. Historical reference is strongly established, $p < 0.001$



Density plots of non-inferiority (NI) margin over 10,000 simulations for two different situations, evaluating the effect of a historical trial over the width of NI margin.

SIMULATION RESULTS (UNKNOWN VARIANCE)

We compared Bayesian method with the frequentist methods with respect to the power. Throughout this simulation we set $\mu_E = \mu_R - \delta_{NI}^B + \xi_1$ and $\mu_P = \mu_R - AS - \xi_2$. To reduce the computational cost we set $\xi_1 = \xi_2/2$. We consider $n_{P,H} = n_{R,H} = 100$ and $n_R = n_P = n_E = 150, 300$. The criteria for Bayesian method to reject H_0 and K_0 if the joint posterior probability of H_1 and K_1 is greater than 0.975. For the power comparison, let ξ vary from 0.1 to 0.7 by increments of 0.2. For the power comparison, we consider $\mu_P^H = 0$, $\mu_R^H = 1$, $\sigma_{P,H}^2 = 1$, $\sigma_{R,H}^2 = 1$, $\sigma_P^2 = 2$, $\sigma_R^2 = 1$ and $\sigma_E^2 = 1.5$, where $r = 1 - \lambda$.

- ❑ λ is the preservation level of the effect size obtained from the historical trial used in NI hypothesis. Note that this specific choice of r ($=1 - \lambda$) give the AS hypothesis the full effect size and makes it independent of λ as,

$$\delta_{AS}^B = \delta_{NI}^B / r = \frac{(1-\lambda)\Delta^B}{r}$$

- ❑ No Bayesian calibration is done which will reduce p^* , thus improving power

HISTORICAL TRIALS STRONGLY ESTABLISHING R

$n_R = n_p = n_E = 150$			$n_R = n_p = n_E = 300$		
ξ_2	Bayesian method	Frequentist method	ξ_2	Bayesian method	Frequentist method
$\lambda = 0$					
0.2	0.08	0.01	0.2	0.11	0.06
0.4	0.35	0.23	0.4	0.64	0.57
0.6	0.72	0.63	0.6	0.93	0.91
0.8	0.92	0.87	0.8	1.00	0.99
$\lambda = 0.25$					
0.2	0.08	0.01	0.2	0.11	0.05
0.4	0.34	0.23	0.4	0.64	0.57
0.6	0.71	0.62	0.6	0.94	0.90
0.8	0.92	0.87	0.8	1.00	0.99
$\lambda = 0.50$					
0.2	0.07	0.01	0.2	0.11	0.06
0.4	0.35	0.23	0.4	0.63	0.57
0.6	0.71	0.62	0.6	0.93	0.94
0.8	0.92	0.87	0.8	1.00	0.99
$\lambda = 0.75$					
0.2	0.08	0.01	0.2	0.11	0.05
0.4	0.35	0.24	0.4	0.62	0.57
0.6	0.71	0.62	0.6	0.93	0.91
0.8	0.92	0.87	0.8	1.00	0.99

For the power comparison, we set $\mu_E = \mu_R - \delta_{NI}^H + \xi_1$ and $\mu_p = \mu_R - \delta_{AS}^H - \xi_2$, where $\mu_p^H = 0$, $\mu_R^H = \mu_R = 1$, $\sigma_{p,H}^2 = 1$, $\sigma_{R,H}^2 = 1$, $\sigma_p^2 = 2$, $\sigma_R^2 = 1$ and $\sigma_E^2 = 1.5$, where $r = 1 - \lambda$, $p^* = 0.975$ in 50,000 replications and 5000 samplers. To reduce the computational cost $\xi_1 = \xi_2/2$ with $n_{p,H} = n_{R,H} = 100$; and $n_R = n_p = n_E = 150, 300$.

HISTORICAL TRIALS NOT STRONGLY ESTABLISHING R

$n_R = n_P = n_E = 150$			$n_R = n_P = n_E = 300$		
ξ_2	Bayesian method	Frequentist method	ξ_2	Bayesian method	Frequentist method
$\lambda = 0$					
0.2	0.04	0.00	0.2	0.08	0.03
0.4	0.24	0.09	0.4	0.54	0.23
0.6	0.67	0.51	0.6	0.90	0.79
0.8	0.87	0.83	0.8	1.00	0.98
$\lambda = 0.25$					
0.2	0.05	0.00	0.2	0.07	0.01
0.4	0.24	0.08	0.4	0.55	0.22
0.6	0.66	0.51	0.6	0.91	0.78
0.8	0.88	0.83	0.8	1.00	0.98
$\lambda = 0.50$					
0.2	0.04	0.00	0.2	0.07	0.01
0.4	0.25	0.07	0.4	0.57	0.23
0.6	0.66	0.50	0.6	0.91	0.78
0.8	0.88	0.83	0.8	1.00	0.97
$\lambda = 0.75$					
0.2	0.04	0.00	0.2	0.09	0.02
0.4	0.25	0.07	0.4	0.58	0.22
0.6	0.67	0.51	0.6	0.92	0.77
0.8	0.86	0.83	0.8	1.00	0.97

For the power comparison, we set $\mu_E = \mu_R + \xi_1$ and $\mu_P = \mu_R - \xi_2$, where $\mu_P^H = 0$, $\mu_R^H = \mu_R = 0.3$, $\sigma_{P,H}^2 = 1$, $\sigma_{R,H}^2 = 1$, $\sigma_P^2 = 2$, $\sigma_R^2 = 1$ and $\sigma_E^2 = 1.5$, where $r = 1 - \lambda$, $p^* = 0.975$ in 50,000 replications and 5,000 samplers. To reduce the computational cost $\xi_1 = \xi_2/2$ with $n_{P,H} = n_{R,H} = 100$; and $n_R = n_P = n_E = 150, 300$.

BAYESIAN APPROACH: RISK DIFFERENCE

- Ghosh et al. (2017) extended it for RD
- We will first consider the conjugate prior situation
- We will consider first the conjugate prior for $\pi_l, l \in \{E, R, P\}$
- To construct NI margin consider 1st historical placebo controlled trial (R and P),

$$X_l^H \sim Bin(n_l^H, \pi_l^H) \text{ (for } l = R, P)$$

- Priors are $\pi_l^H \sim Beta(\alpha_l^H, \beta_l^H)$ (for $l = R, P$). A flexibility of beta distribution is it can be made very flat in [0,1]
- Historical posteriors are,

$$\pi_l^H | X_l^H \sim beta(X_l^H + \alpha_l^H, n_l^H - X_l^H + \beta_l^H), \text{ for } l = R, P$$

- For Δ computation we need the distbⁿ $(\pi_R^H - \pi_P^H | X_R^H, X_P^H)$

MARGIN COMPUTATION

- Unfortunately the difference $(\pi_R^H - \pi_P^H | X_R^H, X_P^H)$ does not have a closed form solution and we resort to a sampling based approach to calculate lower limit of the $100(1-\alpha)\%$ credible interval using the equation

$$P(\pi_R^H - \pi_P^H \geq \Delta^B | X_R^H, X_P^H) \geq 1 - \alpha/2$$

- The Frequentist NI margin is given by,

$$\delta_{NI}^F = (1 - \lambda)\Delta^F = (1 - \lambda) \left(\frac{X_R^H}{n_R^H} - \frac{X_P^H}{n_P^H} - z_{1-\alpha/2} \sqrt{\frac{\sigma_{RH}^2}{n_R^H} + \frac{\sigma_{PH}^2}{n_P^H}} \right)$$

λ is the preservation level ($0 < \lambda < 1$) and $\sigma_{lH}^2 = n_l^H \pi_l^H (1 - \pi_l^H)$

- The AS margin is given by, $\delta_{AS}^B = \delta_{NI}^B / r = \frac{(1-\lambda)\Delta^B}{r}$
- With $r = 1 - \lambda$, NI Margin = AS Margin (HT) and $r > 1 - \lambda$ is Modified HT with the restriction $(1 - \lambda) \leq r \leq 1$.
- The choice λ and r should be guided by clinical practice

POSTERIOR DISTRIBUTION (CURRENT TRIAL)

- For current trial with three arms E , R and P , we choose beta prior for E , and historical posterior for R and P as prior (informative).
- Note the implicit assumption of Constancy
- Posterior is given by,

$$\pi_E | X_E \sim \text{beta}(X_E + \alpha_E, n_E - X_E + \beta_E),$$

$$\pi_l | X_l, X_l^H \sim \text{beta}(X_l + X_l^H + \alpha_l^H, n_l + n_l^H - X_l - X_l^H + \beta_l^H) \text{ for } l \in \{R, P\}$$

- Joint posterior distribution of (π_E, π_R, π_P) is simple product. However to test NI with AS we need the distribution of $(\pi_E - \pi_R, \pi_R - \pi_P | Data)$.
- This is not available in closed form and the test statistic(s) are not independent

CONTINUED...

- We can sample from marginal posterior of π_E, π_R and π_P (all *beta*). Let's say we draw M many such samples
- We declare NI with AS if,

$$\frac{1}{M} \sum_{i=1}^M I\{\pi_E^i - \pi_R^i \geq -\delta_{NI}^B, \pi_R^i - \pi_P^i \geq \delta_{NI}^B / r\} \geq p^*$$

- Where p^* is pre-specified from the level (α) condition (such as 0.975 or 0.95.) and Bayesian Calibration
- Bayesian calibration asks for p^* such that type-I error is not too conservative or too liberal
- Note we are in the realm of multiple testing (NI and AS) with correlated test statistic
- However under intersection–union (IU) testing setup no multiplicity adjustment is needed

AN APPROXIMATE BAYESIAN SOLUTION

- We approximate beta by Normal via moment matching, this eliminates need for Monte Carlo approximation of Δ
- Consider posterior distribution for Historical trial

$$\pi_l^H | X_l^H \sim \text{beta}(X_l^H + \alpha_l^H, n_l^H - X_l^H + \beta_l^H), \text{ for } l = R, P$$

approximate it by $\pi_l^H | X_l^H \sim AN(\mu_{lH}, \sigma_{lH}^2)$

where $\mu_{lH} = \frac{X_l^H + \alpha_l^H}{n_l^H + \alpha_l^H + \beta_l^H}$, $\sigma_{lH}^2 = \frac{(X_l^H + \alpha_l^H)(n_l^H - X_l^H + \beta_l^H)}{(n_l^H + \alpha_l^H + \beta_l^H)^2(n_l^H + \alpha_l^H + \beta_l^H + 1)}$

- This helps directly expressing Δ analytically as

$$(\pi_R^H - \pi_P^H | X_R^H, X_P^H) \sim AN(\mu_{RH} - \mu_{PH}, \sigma_{RH}^2 + \sigma_{PH}^2)$$

- Hence Δ^B is obtained by computing

$$P(\pi_R^H - \pi_P^H > \Delta^B | X_R^H, X_P^H) = P\left(Z > \frac{\Delta^B - (\mu_{RH} - \mu_{PH})}{\sqrt{\sigma_{RH}^2 + \sigma_{PH}^2}}\right) = 1 - \alpha/2$$

- Frequentist Δ^F is even more straight forward

CONTINUED ... CURRENT TRIAL

- We again approximate beta by normal via moment matching

$$\pi_E | X_E \sim \text{beta}(X_E + \alpha_E, n_E - X_E + \beta_E),$$

$$\pi_l | X_l, X_l^H \sim \text{beta}(X_l + X_l^H + \alpha_l^H, n_l + n_l^H - X_l - X_l^H + \beta_l^H) \text{ for } l \in \{R, P\}$$

$$\pi_l | Data \sim AN(\mu_l, \sigma_l^2) \text{ for } l = \{E, R, P\}$$

$$\mu_E = \frac{X_E + \alpha_E}{n_E + \alpha_E + \beta_E}, \sigma_E^2 = \frac{(X_E + \alpha_E)(n_E - X_E + \beta_E)}{(n_E + \alpha_E + \beta_E)^2(n_E + \alpha_E + \beta_E + 1)}$$

$$\mu_l = \frac{X_l^* + \alpha_l^H}{n_l^* + \alpha_l^H + \beta_l^H}, \sigma_l^2 = \frac{(X_l^* + \alpha_l^H)(n_l^* - X_l^* + \beta_l^H)}{(n_l^* + \alpha_l^H + \beta_l^H)^2(n_l^* + \alpha_l^H + \beta_l^H + 1)} \text{ for } l = \{R, P\}, X_l^* = X_l + X_l^H, n_l^* = n_l + n_l^H$$

For joint testing of NI and AS $(\pi_E - \pi_R, \pi_R - \pi_P | Data) \sim AN(\eta, \Psi)$

where $\eta = (\mu_E - \mu_R, \mu_R - \mu_P)'$ and $\Psi = \begin{bmatrix} \sigma_E^2 + \sigma_R^2 & -\sigma_R^2 \\ -\sigma_R^2 & \sigma_R^2 + \sigma_P^2 \end{bmatrix}$

- Decision criterion for deciding NI with AS is based bivariate normal if

$$P(\pi_E - \pi_R > -\delta_{NI}^B, \pi_R - \pi_P > \delta_{NI}^B / r | Data) \geq p^*$$

BAYESIAN FORMULATION (BY DE-CONSTRAINING)

- Sampling from a constrained parameter space ($0 \leq \pi \leq 1$) is problematic, especially when putting non-conjugate prior due to lack of log-concavity
- We formulated a de-constraining transformation on π
- Consider again $X_l^H \sim \text{Bin}(n_l^H, \pi_l^H)$ (for $l = R, P$)

$$\theta_l = \text{logit}(\pi_l) \Rightarrow \pi_l = \frac{e^{\theta_l}}{1 + e^{\theta_l}}; \theta_l \in \mathfrak{R}$$

- This allow us to put normal prior on θ . We follow hierarchical Bayesian setup,
 $f(\theta_l) \sim N(\mu_l, \sigma_l^2),$
 $\mu_l | \sigma_l^2 \sim N(\mu_{0l}, \sigma_l^2 / \kappa_{0l}),$
 $\sigma_l^2 \sim \text{Inv-gamma}(v_{0l}/2, \sigma_{0l}^2 v_{0l}/2),$
where $\mu_{0l}, \kappa_{0l}, v_{0l}, \sigma_{0l}^2$ are fixed hyperparameters.

- Easy to show,

$$f(\mu_l, \sigma_l^2 | \theta_l) \sim N - \text{Inv} - \chi^2(\mu_n, \sigma_n^2 / k_n; \nu_n, \sigma_n^2)$$

DE-CONSTRAINED SOLUTION

- Posterior distribution of $\theta_l, l \in \{R, P\}$

$$(\theta_l | X_l, \mu_l, \sigma_l^2) \propto \text{Binomial-Likelihood } (\theta_l | X_l) \times N(\theta_l | \mu_l, \sigma_l^2) = \frac{e^{X_l \theta_l}}{(1+e^{\theta_l})^{n_l}} \times N(\theta_l | \mu_l, \sigma_l^2)$$

- We apply Grouped Gibbs Sampling to obtain sample from the above posterior distribution of θ_l for each R & P
- Once such sample is obtained we reverse transform it to obtain π_l
- Then take pairwise difference to obtain sample from the difference distribution $(\pi_R^H - \pi_P^H | X_R^H, X_P^H)$
- This are next sorted to calculate lower limit of the $100(1-\alpha)\%$ credible interval using the equation as NI margin

$$P(\pi_R^H - \pi_P^H \geq \Delta^B | X_R^H, X_P^H) \geq 1 - \alpha/2$$

POSTERIOR DISTRIBUTION (CURRENT TRIAL)

- Current trial with three arms E , R and P . We choose normal prior for E , and historical posterior for R and P as prior (informative) for $\theta_l, l \in \{E, R, P\}$
- For E arm,

$$(\theta_E | X_E, \mu_E, \sigma_E^2) \sim \frac{e^{X_E \theta_E}}{(1 + e^{\theta_E})^{n_E}} \times N(\theta_E | \mu_E, \sigma_E^2),$$

- For R and P arm,

$$(\theta_l | X_l, X_{lH}, \mu_{lH}, \sigma_{lH}^2) \sim \frac{e^{X_l \theta_l + X_{lH} \theta_l}}{(1 + e^{\theta_E})^{n_l + n_{lH}}} \times N(\theta_l | \mu_{lH}, \sigma_{lH}^2), \text{ for } l \in \{R, P\}$$

- Note though posteriors are complicated the posterior propriety holds as all priors are proper distributions
- We implemented grouped Gibbs sampler to draw sample from each marginal posterior of θ_l to π_l
- The decision rule for NI and AS remains same

SIMULATION RESULTS

- We used a beta(1,1) prior for historical data for both R and P arm. First we determine the value of p^* such that Type-I error ≈ 0.025

Bayesian Type1 Error

λ	p^*	CBM	CPA	CBD	p^*	UCBM	UCPA	UCBD
0.0	0.757	0.0230	0.0090	0.016	0.975	0.0008	0.0001	0.0003
0.25	0.763	0.0228	0.0156	0.018	0.975	0.0008	0.0001	0.0005
0.50	0.765	0.0212	0.0130	0.020	0.975	0.0006	0.0002	0.0003
0.75	0.768	0.0226	0.0182	0.019	0.975	0.0006	0.0003	0.0004

Table 1: The above type-1 errors are obtained using $n_P^H = n_R^H = n_R = n_P = n_E = 300$, $\pi_P^H = 0.2$, $\pi_R^H = \pi_R = 0.75$ and $\pi_E = \pi_P - NI + \xi$. CBM denotes calibrated Bayesian Method, and UCBM denotes Uncalibrated Bayesian Method.

Empirical Power						
λ	ξ	CBM	UCBM	Frequentist	Posterior	Approximation
$n_E = n_R = np = 100$						
0.0	0.05	0.618	0.026	0.001	0.021	
0.0	0.10	0.882	0.277	0.067	0.258	
0.0	0.15	0.984	0.728	0.463	0.731	
0.0	0.20	0.999	0.952	0.842	0.948	
0.25	0.05	0.596	0.027	0.002	0.026	
0.25	0.10	0.878	0.224	0.072	0.227	
0.25	0.15	0.980	0.642	0.429	0.656	
0.25	0.20	0.999	0.937	0.842	0.936	
0.50	0.05	0.588	0.021	0.002	0.025	
0.50	0.10	0.884	0.234	0.071	0.231	
0.50	0.15	0.985	0.677	0.451	0.670	
0.50	0.20	1.000	0.940	0.871	0.944	
0.75	0.05	0.576	0.028	0.004	0.025	
0.75	0.10	0.898	0.235	0.081	0.256	
0.75	0.15	0.993	0.738	0.510	0.745	
0.75	0.20	1.000	0.975	0.912	0.976	

- λ is the preservation level of the effect size obtained from the historical trial used in NI hypothesis. Note that this specific choice of r ($=1-\lambda$) give the AS hypothesis the full effect size and make it independent of λ as,

$$\delta_{AS}^B = \delta_{NI}^B / r = \frac{(1-\lambda)\Delta^B}{r}$$

Empirical Power

λ	ξ	CBM	UCBM	Frequentist	Posterior Approximation
$n_E = n_R = n_P = 300$					
0.0	0.05	0.805	0.108	0.031	0.091
0.0	0.10	0.987	0.796	0.661	0.790
0.0	0.15	1.000	0.991	0.984	0.994
0.0	0.20	1.000	1.000	1.000	1.000
0.25	0.05	0.784	0.085	0.033	0.081
0.25	0.10	0.987	0.761	0.637	0.748
0.25	0.15	1.000	0.989	0.974	0.991
0.25	0.20	1.000	1.000	1.000	1.000
0.50	0.05	0.775	0.081	0.033	0.086
0.50	0.10	0.986	0.734	0.633	0.737
0.50	0.15	1.000	0.992	0.980	0.991
0.50	0.20	1.000	1.000	1.000	1.000
0.75	0.05	0.791	0.086	0.031	0.088
0.75	0.10	0.996	0.770	0.672	0.785
0.75	0.15	1.000	0.998	0.989	0.997
0.75	0.20	1.000	1.000	1.000	1.000

Bayesian Type1 Error

λ	p^*	CBM	p^*	UCB
0.00	0.630	0.023	0.975	0.0001
0.25	0.667	0.024	0.975	0.0002
0.50	0.658	0.021	0.975	0.0001
0.75	0.660	0.020	0.975	0.0001

Table 2: The above type1 errors are obtained using $n_P^H = n_R^H = n_R = n_P = n_E = 300$, $\pi_P^H = 0.2$, $\pi_R^H = 0.75$, $\pi_R = 0.70$ and $\pi_E = \pi_P - NI + \xi$.

Empirical Power

λ	ξ	CBM	UCBM	Frequentist	Posterior Approximation
$n_E = n_R = n_P = 100$					
0.0	0.05	0.259	0.004	0.0002	0.002
0.0	0.10	0.713	0.096	0.021	0.072
0.0	0.15	0.948	0.463	0.208	0.445
0.0	0.20	0.996	0.796	0.589	0.773
0.25	0.05	0.266	0.003	0.0004	0.003
0.25	0.10	0.649	0.075	0.025	0.073
0.25	0.15	0.918	0.369	0.207	0.389
0.25	0.20	0.990	0.732	0.585	0.719
0.50	0.05	0.257	0.004	0.0004	0.003
0.50	0.10	0.642	0.070	0.022	0.070
0.50	0.15	0.920	0.372	0.193	0.374
0.50	0.20	0.993	0.752	0.598	0.731
0.75	0.05	0.263	0.006	0.0002	0.004
0.75	0.10	0.644	0.071	0.027	0.078
0.75	0.15	0.934	0.384	0.219	0.414
0.75	0.20	0.997	0.791	0.636	0.795

Empirical Power

λ	ξ	CBM	UCBM	Frequentist	Posterior Approximation
$n_E = n_R = n_P = 300$					
0.0	0.05	0.470	0.016	0.007	0.010
0.0	0.10	0.957	0.509	0.344	0.502
0.0	0.15	0.999	0.958	0.861	0.958
0.0	0.20	1.000	1.000	0.989	1.000
0.25	0.05	0.468	0.016	0.009	0.015
0.25	0.10	0.946	0.474	0.332	0.456
0.25	0.15	0.999	0.933	0.854	0.939
0.25	0.20	1.000	1.000	0.988	1.000
0.50	0.05	0.451	0.009	0.008	0.011
0.50	0.10	0.925	0.435	0.318	0.438
0.50	0.15	0.999	0.925	0.854	0.930
0.50	0.20	1.000	1.000	0.989	0.999
0.75	0.05	0.433	0.009	0.007	0.017
0.75	0.10	0.937	0.435	0.340	0.460
0.75	0.15	1.000	0.951	0.862	0.946
0.75	0.20	1.000	1.000	0.989	1.000

A CONDITIONAL TESTING APPROACH (FRACTION MARGIN)

- In fraction margin since simultaneous NI and AS testing is not done, so when NI testing is carried out it is either assumed (or pretested) that AS condition holds
- Our proposal was to use this information in NI testing

$$H_0 : \pi_E - \theta\pi_R - (1 - \theta)\pi_P \leq 0 \text{ vs } H_1 : \pi_E - \theta\pi_R - (1 - \theta)\pi_P > 0.$$

Usual test statistic: $T = \hat{\pi}_E - \theta\hat{\pi}_R - (1 - \theta)\hat{\pi}_P$

Our proposal: $W = (\hat{\pi}_E - \theta\hat{\pi}_R - (1 - \theta)\hat{\pi}_P | \hat{\pi}_R - \hat{\pi}_P > 0) \equiv (U - \theta V | V > 0)$

Lemma 2.2.1 Under conditional normal approximation, the mean μ_w and variance σ_w^2 of $W = (\hat{\pi}_E - \theta\hat{\pi}_R - (1 - \theta)\hat{\pi}_P | \hat{\pi}_R - \hat{\pi}_P > 0)$ are given by

$$\begin{aligned} \mu_w &= \mu_U + \sigma_U \frac{\rho}{c} \phi(d) - \theta \left(\mu_V + \sigma_V \frac{1}{c} \phi(d) \right) \\ \sigma_w^2 &= \sigma_U^2 \left[1 + \frac{\rho^2}{c} d\phi(d) - \left(\frac{\rho}{c} \phi(d) \right)^2 \right] + \sigma_V^2 \left[1 - \frac{\phi(d)}{c} \left(\frac{\phi(d)}{c} - d \right) \right] \\ &\quad - 2\theta \left[\sigma_U \sigma_V \frac{\rho}{c} (c + d\phi(d)) + \sigma_U \mu_V \frac{\rho}{c} \phi(d) + \sigma_V \mu_U \frac{1}{c} \phi(d) + \mu_U \mu_V - \left(\mu_U + \sigma_U \frac{\rho}{c} \phi(d) \right) \left(\mu_V + \sigma_V \frac{1}{c} \phi(d) \right) \right] \end{aligned}$$


SAMPLE SIZE COMPARISON

- Power of the test for a point alternative $\pi_E = \pi_E^{alt}$:

$$P_{H_1}(W > k^*) = 1 - \Phi\left(\frac{k^* - \mu_w^{alt}}{\sigma_w^{alt}}\right) \quad \text{where} \quad k^* = \mu_w^{null} + z_{1-\alpha}\sigma_w^{null}$$

To obtain the power function of the test we fix π_R, π_P and θ and vary π_E such that $\frac{\pi_E - \pi_P}{\pi_R - \pi_P} \in [0.5, 1.4]$.

θ	π_E	Existing		Conditional (0.7,0.1)		Existing		Conditional (0.6,0.55)	
		n_P	N	n_P	N	n_P	N	n_P	N
0.8	0.9	26	78	26	78	30	90	28	84
	0.85	38	114	38	114	43	129	41	123
	0.8	58	174	58	174	68	204	64	192
	0.75	99	297	99	297	120	360	114	342
	0.7	203	609	203	609	257	771	248	744
	0.65	604	1812	604	1812	875	2625	866	2598
0.7	0.9	17	51	17	51	27	81	26	78
	0.85	24	72	24	72	39	117	38	114
	0.8	34	102	34	102	61	183	60	180
	0.75	51	153	51	153	106	318	104	312
	0.7	85	255	85	255	222	666	218	654
	0.65	165	495	165	495	703	2109	698	2094

Sample Size for Marginal vs. Conditional Frequentist Approach

When π_R and π_P is not far away the N is quite smaller, however, for $\pi_R \gg \pi_P$ both methods produces equal N

BAYESIAN SETUP

- Conjugate Beta Prior (CBP): $\pi_l \sim Beta(\alpha_l, \beta_l), l \in \{E, R, P\}$
- Posterior: $0 < \pi_l < 1$

$$f(\pi_E, \pi_R, \pi_P | X_E, X_R, X_P, \alpha_l, \beta_l) \propto I(\pi_R > \pi_P) \prod_{l \in \{E, R, P\}} \pi_l^{\alpha_l + x_l - 1} (1 - \pi_l)^{\beta_l + n_l - x_l - 1}$$

- Non-conjugate Prior: Dirichlet prior (π_R, π_P) s.t. $\pi_R > \pi_P$
- Posterior:

$$f(\pi_E, \pi_R, \pi_P | \mathbf{X}) \propto \pi_P^{\alpha_1 + X_P - 1} (1 - \pi_P)^{n_P - X_P} \pi_R^{X_R} (1 - \pi_R)^{n_R + \alpha_3 - X_R - 1} (\pi_R - \pi_P)^{\alpha_2 - 1} \times \pi_E^{\alpha_E + X_E - 1} (1 - \pi_E)^{\beta_E + n_E - X_E}, \quad 0 < \pi_P < \pi_R < 1, \quad 0 < \pi_E < 1.$$

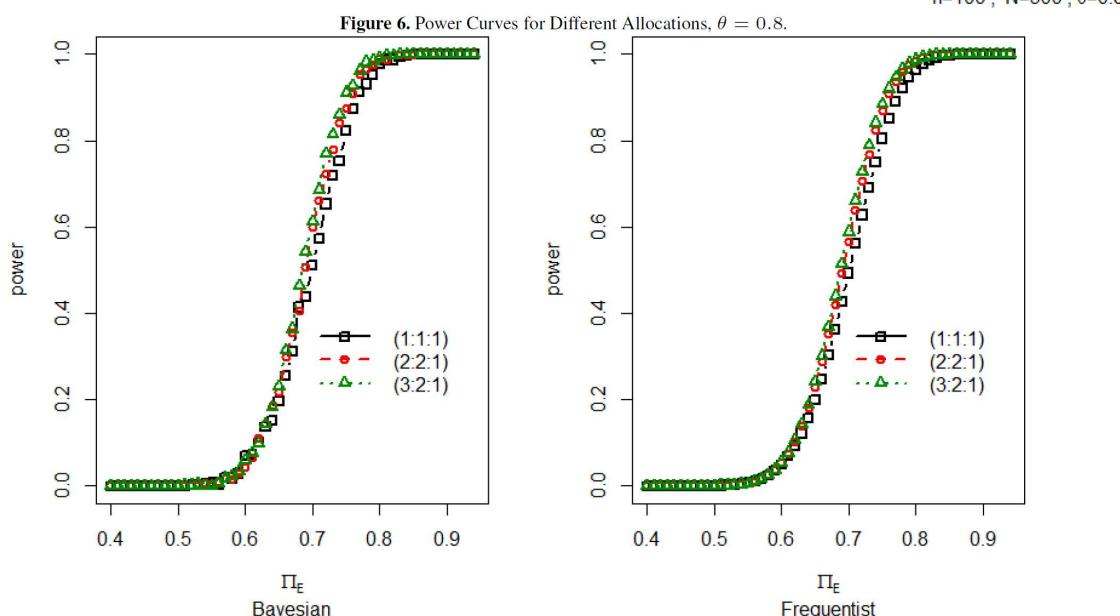
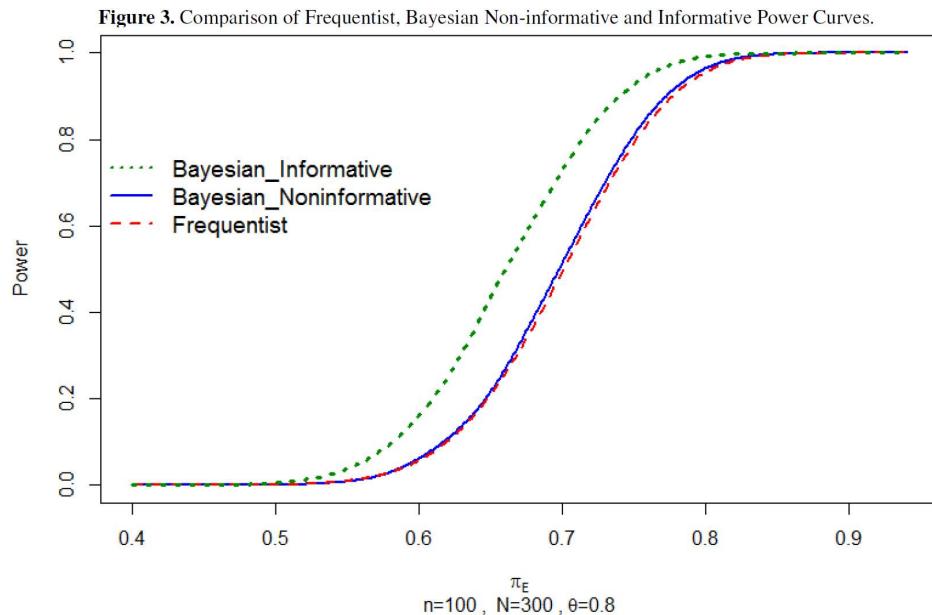
- Test Procedure: $P\left(H_1 : \frac{\pi_E - \pi_P}{\pi_R - \pi_P} > \theta | (\pi_R - \pi_P > 0), \mathbf{X}\right) > R_{NI}$

$$\hat{P}\left(H_1 : \frac{\pi_E - \pi_P}{\pi_R - \pi_P} > \theta | (\pi_R - \pi_P) > 0, \mathbf{X}\right) \approx \frac{1}{T} \sum_{t=1}^T I\left(\frac{\pi_E^t - \pi_P^t}{\pi_R^t - \pi_P^t} > \theta | (\pi_R^t - \pi_P^t) > 0\right)$$

- Also possible to perform Approx. Bayesian procedure

SAMPLE SIZE AND POWER

- Bayesian no-informative and Frequentist is similar
- Unbalanced allocation is slightly better for RD
- Dirichlet prior is more flexible



SOME APPLICATIONS & CASE STUDIES

SOME APPLICATIONS TO REAL TRIAL DATA

Depression Trial (Fixed Margin)

- From a three-arm comparative study on major depressive disorder patients reported by Higuchi et al. The objective of this trial was to compare the efficacy and safety of duloxetine (SNRI) with those of paroxetine (SSRI) and the placebo
- Objective:- Non-inferiority of duloxetine over paroxetine
- The study design was a double-blind, randomized, active-controlled, parallel-group, multicenter study of a 6-week treatment with duloxetine ($n=147$) or paroxetine ($n=148$) or placebo ($n=145$)
- The primary endpoint was the change from baseline in HAMD-17 total score at 6 weeks. The mean decreases in HAMD-17 total score were 10.2 ± 6.1 (mean \pm SD) in the duloxetine group, 9.4 ± 6.9 in the paroxetine group and 8.3 ± 5.8 in the placebo group.
- We resort to Non-informative prior to begin with

We assume for historical trial placebo arm has mean = 8.5 (sd=6.0) and control arm has mean = 10 (sd=6.0). Also, we assume each has 130 samples. ($r = 1 - \lambda$)

λ	Post. Probab.	δ^B	δ^F	Bayesian Decision	NI^B Decision	AS^B Decision	Frequentist Decision	NI^F Decision	AS^F Decision
		Margin	Margin		Decision	Decision		Decision	Decision
0.00	0.95	0.716	0.732	1	1	1	0	1	0
0.25	0.93	0.537	0.549	1	1	1	0	1	0
0.50	0.90	0.358	0.366	1	1	1	0	1	0
0.75	0.86	0.179	0.183	1	1	1	0	1	0
0.90	0.83	0.071	0.073	1	1	1	0	1	0
1.00	0.80	0.000	0.000	1	1	1	0	1	0

Posterior Estimates and 95% credible interval.

Parameter	Median	Interval
μ_P	8.12	(7.39, 8.85)
μ_R	9.87	(9.13, 10.61)
μ_E	10.48	(9.45, 11.51)
$\mu_E - \mu_R$	0.61	(-0.65, 1.88)
$\mu_R - \mu_P$	1.76	(0.72, 2.79)

NI: $H_0: \mu_E - \mu_R \leq -r\Delta$ vs. $H_1: \mu_E - \mu_R > -r\Delta$

AS: $K_0: \mu_R - \mu_P \leq \Delta$ vs. $H_1: \mu_R - \mu_P > \Delta$

Home Based Blood Pressure Intervention Trial

- From a three-arm comparative study on the effectiveness of organizational interventions at improving blood pressure (BP) from Pezzin et al. The primary goal is to see if the basic intervention is at least as good as the augmented one, relative to the usual care
- Usual care is the placebo group, augmented care is the reference group and basic care is the experimental group
- A total of 845 newly admitted patients with uncontrolled hypertension (HTN). A complete follow-up sample size of 525 patients, divided roughly equally across groups. Study end point is at 3 months after the beginning
- The primary endpoint was the decrease from baseline in Systolic Blood Pressure (SBP)
- Unadjusted SBP Outcome at 3-Month Follow-up are Usual care 160.5 ± 25.3 , Basic care 158.7 ± 25.4 and Augmented care 152.5 ± 26.2

We assume for historical trial placebo arm has mean = 159 (sd=26) and control arm has mean = 154 (sd=27). Also, we assume each has 200 samples. ($r = 1 - \lambda$)

λ	Post.	δ^B	δ^F	Bayesian	NI^B	AS^B	Frequentist	NI^F	AS^F
	Probab.	Margin	Margin	Decision	Decision	Decision	Decision	Decision	Decision
0	0.11	1.792	1.591	0	0	1	0	0	1
0.25	0.08	1.344	1.193	0	0	1	0	0	1
0.5	0.05	0.896	0.795	0	0	1	0	0	1
0.75	0.04	0.448	0.397	0	0	1	0	0	1
0.9	0.03	0.179	0.159	0	0	1	0	0	1
1	0.02	0	0	0	0	1	0	0	1

Table 2: Posterior Estimates and 95% credible interval.

Parameter	Median	Interval
μ_P	160.98	(158.59, 163.39)
μ_R	154.21	(151.91, 156.61)
μ_E	158.77	(155.10, 162.48)
$\mu_E - \mu_R$	4.55	(0.15, 9.00)
$\mu_P - \mu_R$	6.77	(3.40, 10.12)

NI: $H_0: \mu_E - \mu_R \leq -r\Delta$ vs. $H_1: \mu_E - \mu_R > -r\Delta$

AS: $K_0: \mu_R - \mu_P \leq \Delta$ vs. $H_1: \mu_R - \mu_P > \Delta$

Depression Trial (fixed margin)

- Same Trial (Higuchi et al.) in case study 1 but the outcomes are now binary in nature
- Objective: Non-inferiority of duloxetine over paroxetine
- The study design was a double-blind, randomized, active-controlled, parallel-group, multicenter study of a 6-week treatment with duloxetine ($n=147$) or paroxetine ($n=148$) or placebo ($n=145$)
- The primary endpoint was the change from baseline in HAMD-17 total score at 6 weeks. Hida and Tango analyzed the data proportion of patients who achieved "remission", HAM-D17 total score at < 7 at 6 weeks
- A secondary outcome is "Response" : $< 50\%$ reduction in HAM-D17

- Hida & Tango reports NI but lack of AS in terms of 95%CI,

$$-0.0961 \leq \pi_E - \pi_R \leq 0.1253, \quad 0.0264 \leq \pi_R - \pi_P \leq 0.2486,$$

λ	CBM	CBD	CPA	δ_{AS}^F	δ_{NI}^F	CBM	CBD	CPA	Frequentist
	Post. Prob., $\delta_{AS}^B, \delta_{NI}^B$	Post. Prob., $\delta_{AS}^B, \delta_{NI}^B$	Post. Prob.			Decision	Decision	Decision	Decision
0.00	0.791 , 0.098, 0.098	0.845 ,0.101, 0.101	0.787	0.097	0.097	1	1	1	0
0.25	0.718 , 0.098, 0.073	0.843 ,0.101, 0.076	0.717	0.097	0.073	0	1	0	0
0.50	0.591 , 0.098, 0.049	0.832 ,0.101, 0.051	0.592	0.097	0.049	0	1	0	0
0.75	0.425 , 0.098, 0.025	0.800 ,0.101, 0.025	0.426	0.097	0.024	0	1	0	0
1.00	0.255 , 0.098, 0	0.773 ,0.101, 0	0.256	0.097	0	0	1	0	0

Table 1: Bayesian and frequentist decision in the Depression trial where “1” stands for rejection and “0” stands for acceptance. These values are obtained using $n_P^H = 145$, $n_R^H = 148$, $\pi_P^H = 0.20$, $\pi_R^H = 0.331$.

We choose ($r = 1 - \lambda$)

NI: $H_0 : \mu_E - \mu_R \leq -r\Delta$ vs. $H_1 : \mu_E - \mu_R > -r\Delta$

AS: $K_0 : \mu_R - \mu_P \leq \Delta$ vs. $H_1 : \mu_R - \mu_P > \Delta$

Depression Trial (fraction margin)

- The primary endpoint was the change from baseline in HAMD-17 total score at 6 weeks. Primary outcome, "remission", HAM-D17 total score at < 7 at 6 weeks. A secondary outcome is "Response" :< 50% reduction in HAM-D17
- NI can be claimed for $\theta \leq 0.55$ for Response data and $\theta = 0.5$ for the Remission data.

θ	Response Data					
	Freq-Marg- p	Freq-Cond- p	CBP Non-info	CBP Info	DP Non-info	DP Info
0.8	0.1975	0.195	0.81	0.845	0.37	0.91
0.75	0.1589	0.157	0.836	0.879	0.397	0.928
0.70	0.1251	0.1238	0.871	0.911	0.437	0.947
0.65	0.0966	0.0958	0.908	0.942	0.468	0.956
0.60	0.0732	0.0728	0.933	0.962	0.506	0.97
0.55	0.0547	0.0545	0.944	0.976*	0.546	0.981*
0.50	0.0404*	0.0404*	0.955	0.985*	0.592	0.985*

θ	Remission Data					
	Freq-Marg- p	Freq-Cond- p	CBP Non-info	CBP Info	DP Non-info	DP Info
0.8	0.265	0.2586	0.723	0.778	0.158	0.809
0.75	0.225	0.2198	0.776	0.832	0.168	0.861
0.70	0.1877	0.1837	0.811	0.872	0.183	0.896
0.65	0.1538	0.1509	0.832	0.905	0.198	0.924
0.60	0.1239	0.1221	0.872	0.935	0.212	0.95
0.55	0.0983	0.0973	0.899	0.953	0.228	0.971
0.50	0.0769	0.0765	0.916	0.976*	0.249	0.985*

Frequentist p -values and Bayesian Posterior Probabilities under Different Informative (Info) and Non-informative (Non-Info) Priors

TOPICS FOR FURTHER READING

- NI trial based on Non-parametric Rank based approach
- NI trial design for Binary outcome with RR, OR and NNT type functional
- NI trial for Count data (with or without over-dispersion)
- When more than one historical R vs. P, RCT exists, how to combine them in a meta-analytic framework
- Methods for Longitudinal/Clustered NI trial design
- Non-Inferiority testing based on Pragmatic trial
- Simultaneous testing of NI and Superiority
- Several variations of group-sequential design applicable for NI trial
- NI trial with multiple experimental drugs (or one drug⁷⁵ with several dosage)

THANK YOU

ANY QUESTION OR SUGGESTION ?

SOME SELECTED REFERENCES

1. DAgostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues—the encounters of academic consultants in statistics. *Statistics in Medicine* 2003; **22**:169--186.
2. Pigeot I, Schafer J, Rohmel J, Hauschke D. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine* 2003; **22**:883--899.
3. Ghosh P., Nathoo F., Gönen M., Tiwari R. C. Assessing noninferiority in a three-arm trial using the Bayesian approach. *Statistics in Medicine* 2010; **30**:1795–1808.
4. Hida T. and Tango T. On the three-arm non-inferiority trial including a placebo with pre-specified margin. *Statistics in Medicine* 2009; **30**: 224–231.
5. Kwong K.S., Cheung S.H., Hayter A.J. and Wen MJ. Extension of three-arm non-inferiority studies to trials with multiple new treatments. *Statistics in Medicine* 2012; **31**: 2833–2843.
6. Kieser M and Friede T. Planning and analysis of three-arm non-inferiority trials with binary endpoints. *Statistics in medicine*, 26(2):253{273, 2007.
7. Ghosh, S., Ghosh, S., & Tiwari, R. C. (2016). Bayesian approach for assessing non-inferiority in a three-arm trial with pre-specified margin. *Statistics in medicine*, 35(5), 695-708.

SELECTED REFERENCES

8. Ochiai, T., Hamasaki, T., Evans, S. R., Asakura, K., & Ohno, Y. (2017). Group-sequential three-arm noninferiority clinical trial designs. *Journal of biopharmaceutical statistics*, 27(1), 1-24.
9. Pigeot I, Schafer J, Rohmel J, Hauschke D. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine* 2003; **22**:883--899.
10. FDA. Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry, 2016.
11. EMA. Guideline on the choice of the Non-inferiority margin (Doc. Ref. EMEA/CPMP/EWP/215.), 2005.
12. Committee for medical products for human use (CHMP) (2005). Guideline on the choice of the non-inferiority margin.
13. Koch A and Rohmel J. Hypothesis testing in the gold standard design for proving the efficacy of an experimental treatment relative to placebo and a reference. *J. Biopharm. Stat.*, 14(2):315–325, 2004.
14. Freidlin, B., Korn, E. L., George, S. L., & Gray, R. (2007). Randomized clinical trial design for assessing noninferiority when superiority is expected. *Journal of Clinical Oncology*, 25(31), 5019-5023.