.

2013 Biopharmaceutical
Applied Statistics Symposium
(BASS) Program, Orlando FL

Tailored Medicines:
Approaches & Considerations
for Development, 11/6/13

Bayesian Hierarchical Models for Subgroup Analysis in Clinical Studies

Gene Pennello, PhD
Division of Biostatistics
FDA / CDRH
Silver Spring, MD

1

Abstract

Bayesian hierarchical models for subgroup analysis will be introduced and considered for early phase biomarker discovery studies, pre-market validation studies, and post-market signal detection. In Bayesian hierarchical modeling, the subgroups are considered jointly. That is, the Bayesian estimate for a subgroup considers not just the data from that subgroup, but the data from all the other subgroups as well. When appropriate, joint Bayesian estimates of subgroup-specific treatment effects are attractive: they tend to have fewer "random highs", greater precision, and an improved rank ordering relative to the point estimates for the subgroups when they are each considered alone.

The Subgroup Problem

- Patients enrolled into a clinical study can be heterogeneous.
- Treatment effects may therefore be heterogeneous among subgroups.
- If treatment effects are estimated separately within each subgroup, their variation will tend to be overstated.
- The chance of a falsely positive subgroup finding can be out of control.

The Subgroup Problem
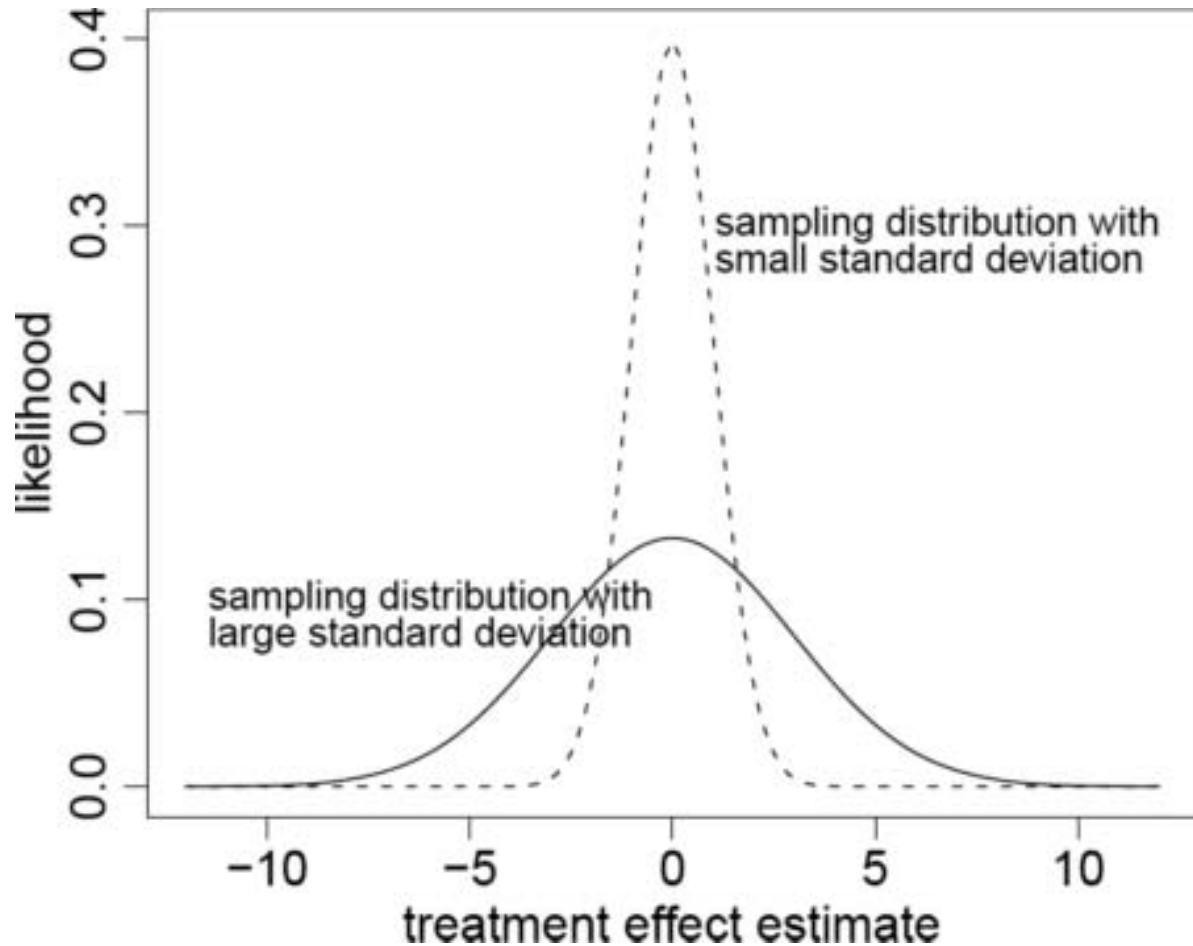
Gelman, Hill, Yajima, Why We (Usually) Don't Have to Worry About Multiple Comparisons. *J Res Educ Effectiveness* 2012; 5: 189–211.

- Subgroup specific treatment effects can be falsely significant (statistically or clinically). [4]

Frequentist Subgroup Procedures

- Analyze subgroups separately, but
    - Control the *familywise* type 1 error rate, the probability of falsely rejecting at least one true null hypothesis among the family of subgroups being tested.
    - Lower significance level for subgroup hypothesis test.
    - Assess significance with adjusted p value (smallest familywise level at which subgroup is significant).
    - Widen confidence interval about subgroup estimate.

- Gatekeepers (possibly underpowered):
    - Significant treatment by subgroup interaction
    - Significant overall treatment effect

Bayesian Hierarchical Model

- Analyze subgroups not separately, but *jointly*.

- Assume subgroup treatment effects are *exchangeable*, which relates their estimation.

- Bayesian estimate (posterior mean) of a subgroup-specific treatment effect is a *weighted average* of
  - the sample estimate in the subgroup, and
  - the overall treatment effect estimate.

- Bayesian inference for a subgroup depends on
  - the data on the other subgroups, but
  - not on whether they are in the family of subgroups being formally tested (inference is *comparisonwise*)

Bayesian Hierarchical Model

- Relative to separate analyses of subgroups,
    - Bayesian subgroup estimates are *more* precise (borrows strength from overall estimate).
    - Bayesian credible intervals are *shorter* (due to borrowing strength).
    - Bayesian subgroup analysis exhibits **fewer "random highs"** (subgroup sample estimate is "shrunk" toward overall estimate).
    - Significant differences between subgroups are harder to declare (shrinkage toward overall estimate is faster than increase in precision).

7
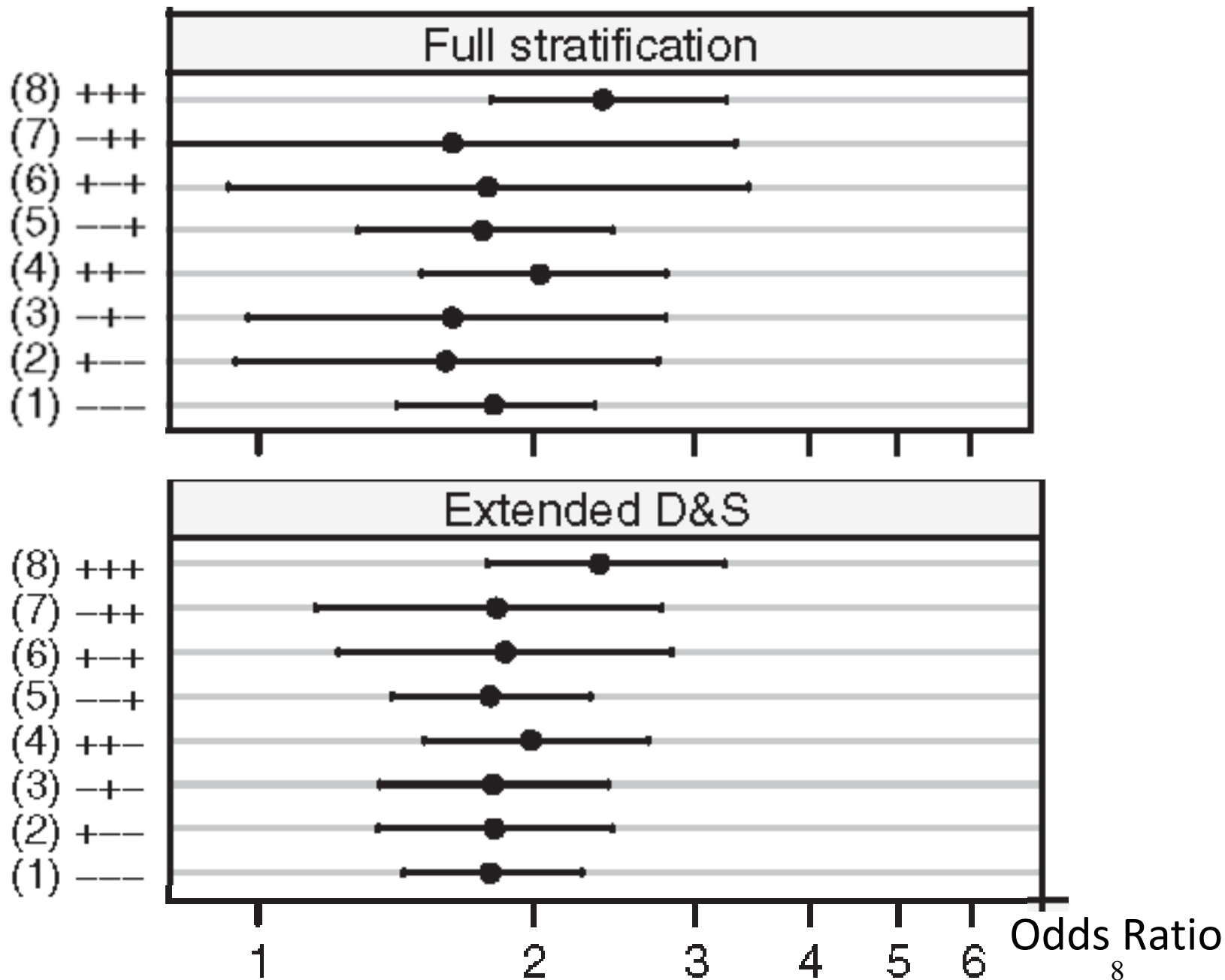
Extended D&S is a 3-way Bayesian hierarchical model
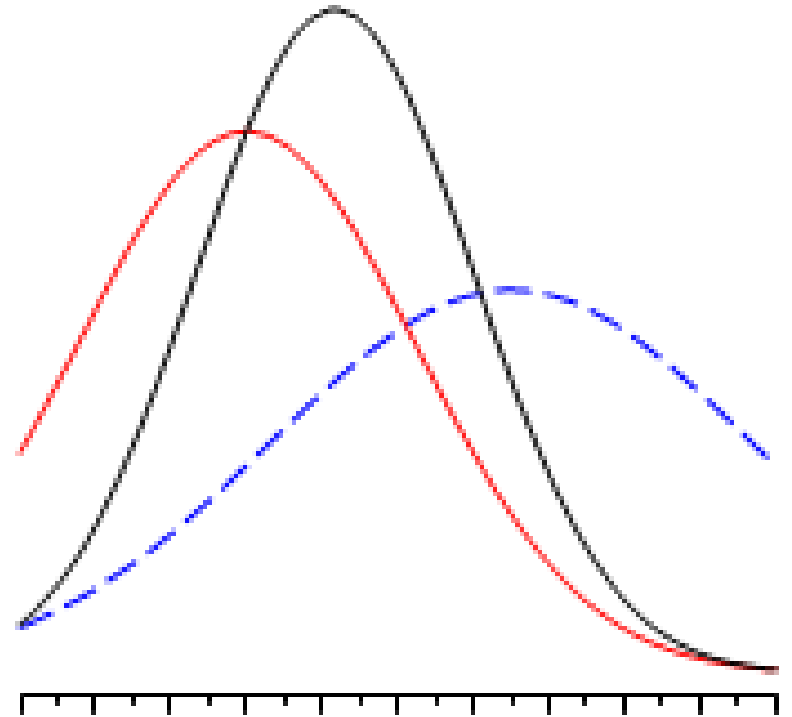
Early Clinical Studies

- In a feasibility, exploratory, or Phase II study, suppose
  - Sample Size is Small to Moderate.
  - Heterogeneous Treatment Effects are Plausible
  - Truly Zero Effects not likely.

- Bayesian HM Subgroup Analysis
  - increases precision of subgroup estimates,
  - obtains an 'honest' subgroup estimate (via shrinkage), useful for
    - deciding whether to validate the treatment in a pivotal or Phase III study of just one of the subgroups.
    - sizing that validation study.
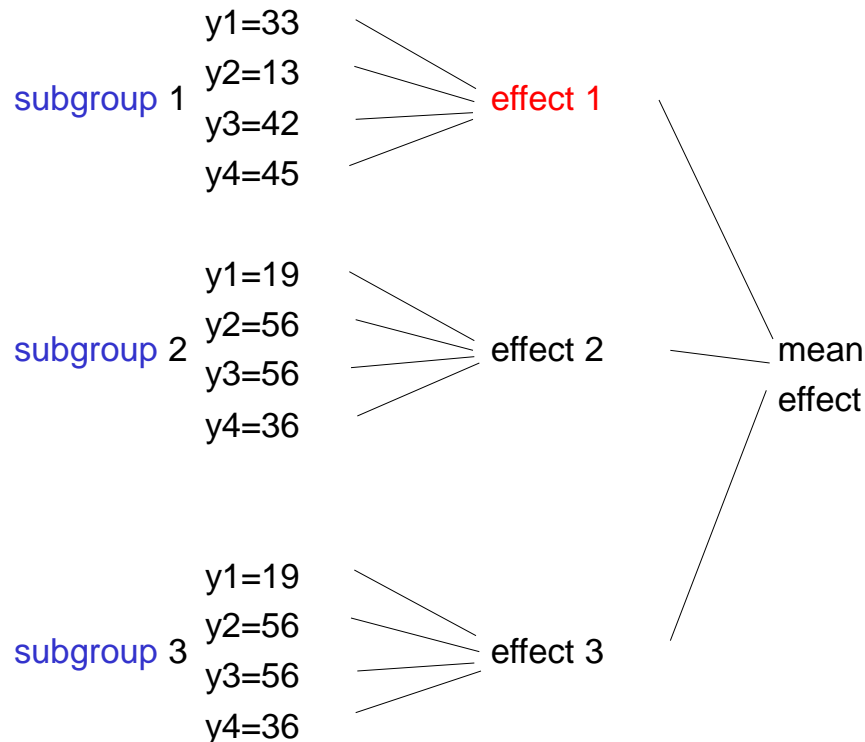
Pivotal Clinical Validation

- A subgroup finding can be assessed for statistical and clinical significance using Bayesian hierarchical model subgroup analysis.

- Clinical significance of the subgroup finding can be assessed by

  - using the posterior mean shrinkage estimate and associated posterior standard deviation,

  - not by taking the subgroup sample estimate at face value.

Bayesian Analysis

- Prior Distribution
  - A priori distribution of parameter

- Likelihood
  - Distribution of data given parameter

- Posterior Distribution
  - Update of prior distribution of parameter given likelihood

Hierarchical Model: Subgroups Have Different But Related Effects

subgroup 1
y1=33
y2=13
y3=42
y4=45
effect 1

subgroup 2
y1=19
y2=56
y3=56
y4=36
effect 2

subgroup 3
y1=19
y2=56
y3=56
y4=36
effect 3

mean effect

All patients are related: more so within subgroups than between subgroups.

Level 1: Patients *exchangeable* within subgroups
Level 2: Subgroups *exchangeable*.

12

Are Subgroups Exchangeable?

- Consider estimating a treatment effect.

- <u>Subgroups</u> are *exchangeable* in their treatment effects if any ordering of them is considered equally likely a priori (i.e., before seeing the data).

- Put another way, if the subgroup-specific treatment effects were revealed, but their subgroup labels were not, then the effects would not be helpful in predicting their labels.

Beta-blocker for Hypertension

Bayesian subgroup analysis

Beta-blocker for Hypertension

Losartan versus atenolol randomized trial

Endpoint: composite of Stroke/ MI/ CV Death

N=9193   losartan (4605),
                atenolol (4588)

 # Events   losartan   (508),
                atenolol (588)

80% European Caucasians 55-80 years old.

http://www.accessdata.fda.gov/drugsatfda_docs/label/2013/020386s058lbl.pdf

Beta-blocker for Hypertension

Cox Analysis

	N	logHR	SE	HR	..95%CI..	p val
Overall	9193			0.87	0.77,0.98	0.021
Race Subgroups						
Non-Black	8660	−.19	.06	0.83	0.73,0.94	0.003
Black	533	.51	.24	1.67	1.04,2.66	0.033

Is the Finding Among Blacks Real or a Directional Error due to Multiplicity?

Bayesian Hierarchical Model

For subgroups $i = 1, 2$ with n_i patients, assume observed log hazard ratio

$$y_i \sim N(\mu_i, \sigma_i^2 / n_i),$$

$$(\text{se of } y_i)^2 = s_i^2 \sim \Gamma(f_i / 2, \ n_i f_i / 2\sigma_i^2)$$

Null Hypothesis: $\quad H : \mu_i > 0$

Bayesian Hierarchical Model

Data:
$$y_i \sim N(\mu_i, \sigma_i^2 / n_i)$$

$$s_i^2 \sim \Gamma(f_i / 2, \ n_i f_i / 2\sigma_i^2)$$

Prior:
$$\mu_i \sim N(\mu_0, \sigma_\mu^2), \ i = 1, 2$$

$$\mu_0 \sim N(0, 1000)$$

$$\sigma_i^{-2}, \sigma_\mu^{-2} \sim \Gamma(.001, .001)$$

Beta-blocker for Hypertension

Bayesian HM Analysis

	logHR	se/sd	HR	..95%CI..		p val	Pr HR>1
non-black							
frequentist	-.19	.06	0.83	0.73	0.94	0.003	
Bayesian	-.18	.06	0.84	0.74,	0.95		0.002
black							
frequentist	.51	.24	1.67	1.04,	2.67	0.033	
Bayesian	.38	.27	1.47	0.87,	2.44		0.914

Bayesian analysis casts some doubt on the surprising finding among blacks. Shrinkage of effect size in blacks is predicated on exchangeability, that is, not expecting a priori a worse effect in blacks than in non-blacks.

Other Clinical Situations

- Biomarker Discovery
  - A huge number of biomarker defined subgroups may be examined (e.g., GWAS SNP based subgroups).

- Post-Market Signal Detection
  - Assess disproportionality in the reporting of an adverse event with the use of a medical product (a drug, biologic, or device).
  - Number of combinations of type of adverse event and medical product is huge.
  - For most combinations, disproportionality is not expected.

- Variants on hierarchical model for these cases:
  - Model subgroup effects as being drawn from a mixture of two distributions, e.g., a distribution of non-zero effects and a degenerate distribution of zero effects.

Bayesian Hierarchical Model, Normal Data, Normal Prior

Subgroups $\quad j = 1, 2, ..., J$

Treatment effect data

$$y_j \sim N(\mu_j, \tau_y^{-1}), \tau_y^{-1} = \sigma_y^2$$

Prior $\quad \mu_j \sim N(\mu_0, \tau_\mu^{-1}), \tau_\mu^{-1} = \sigma_\mu^2$

(true treatment effects $\{\mu_j\}$ are exchangeable)

For simplicity, assume hypermean μ_0, variances σ_μ^2 , σ_y^2 are known.

Bayesian Hierarchical Model, Normal Data, Normal Prior

- **Posterior distribution of treatment effect in subgroup *j*:**

$$\mu_j \mid \{y_j\} \sim N\big((1-S)\mu_0 + Sy_j, S\sigma_y^2\big)$$

$S = 1 - 1/\Phi$ is the shrinkage factor

$\Phi = (\sigma_y^2 + \sigma_\mu^2)/\sigma_y^2$ is the 'true' F ratio

Bayesian Hierarchical Model, Normal Data, Normal Prior

Posterior distribution of treatment effect difference between two subgroups:

$$\delta = \mu_1 - \mu_2 \mid \{y_j\} \sim N\left(Sd, S\sigma_d^2\right)$$

$$d = y_1 - y_2$$

$$\sigma_d^2 = 2\sigma_y^2$$

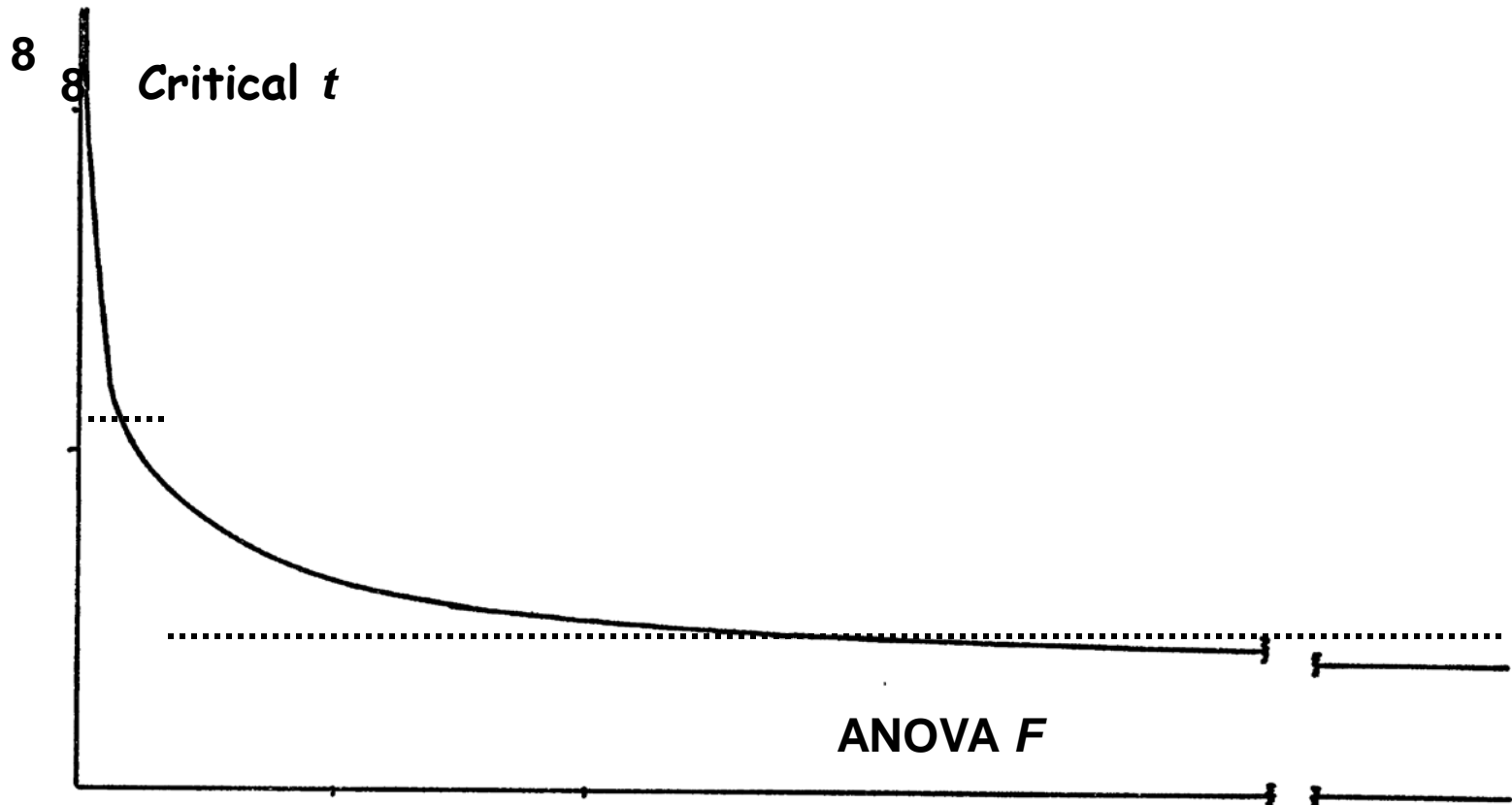# Bayesian Hierarchical Model, Normal Data, Normal Prior

Critical Region:

$$\Pr(\delta > 0 \,|\, \{\, y_j \}) > 1 - \alpha$$

iff
$$z_d = \frac{d}{\sigma_d} > \frac{z_{1-\alpha}}{\sqrt{S}}$$

Critical z_d value $\dfrac{z_{1-\alpha}}{\sqrt{S}}$ $\uparrow$ as $\Phi, \dfrac{\sigma_\mu^2}{\sigma_y^2}$ $\downarrow$

24

Fully Bayesian Critical *t* Values

Bayesian Two-Way Hierarchical Model

$$\bar{y}_{ij} \sim N(\mu_{ij}, \sigma^2 / n), i = 1, \ldots a, j = 1, \ldots b$$

$$fs^2 / \sigma^2 \sim \chi^2(f), \ f = ab(r-1)$$

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

$$\alpha_i \sim N(0, \sigma_\alpha^2), \beta_j \sim N(0, \sigma_\beta^2), \gamma_{ij} \sim N(0, \sigma_\gamma^2)$$

Jeffreys prior on $(\mu, \sigma^2, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2)$

Known Variances Inference

Subgroup Problem: $\quad \delta_{12,j} = \mu_{1j} - \mu_{2j}$

Posterior $\quad \delta_{12,j} \mid \underline{y}, \underline{\sigma^2} \sim$

$$N(S_A d_{12\bullet} + S_C d_C, (S_A + (b-1)S_C)\sigma_d^2 / b)),$$

$$d_{12\bullet} = \bar{y}_{1\bullet} - \bar{y}_{2\bullet}, \quad d_C = d_{12,j} - d_{12\bullet}$$

$$S_A = 1 - 1/\Phi_A, \quad \Phi_A = \sigma_A^2/\sigma^2, \quad \sigma_A^2 = \sigma_C^2 + br\sigma_\alpha^2$$

$$S_C = 1 - 1/\Phi_C, \quad \Phi_C = \sigma_C^2/\sigma^2, \quad \sigma_C^2 = r\sigma_\gamma^2 + \sigma^2$$

Bayes Decision Rule

Let $\quad z_{12,j} = d_{12,j} / \sigma\sqrt{2/r}, \quad d_{12,j} = \bar{y}_{1j} - \bar{y}_{2j}$

$$z_{12,\bullet} = d_{12,\bullet} / \sigma\sqrt{2/br}, \quad d_{12,\bullet} = \bar{y}_{1\bullet} - \bar{y}_{2\bullet}$$
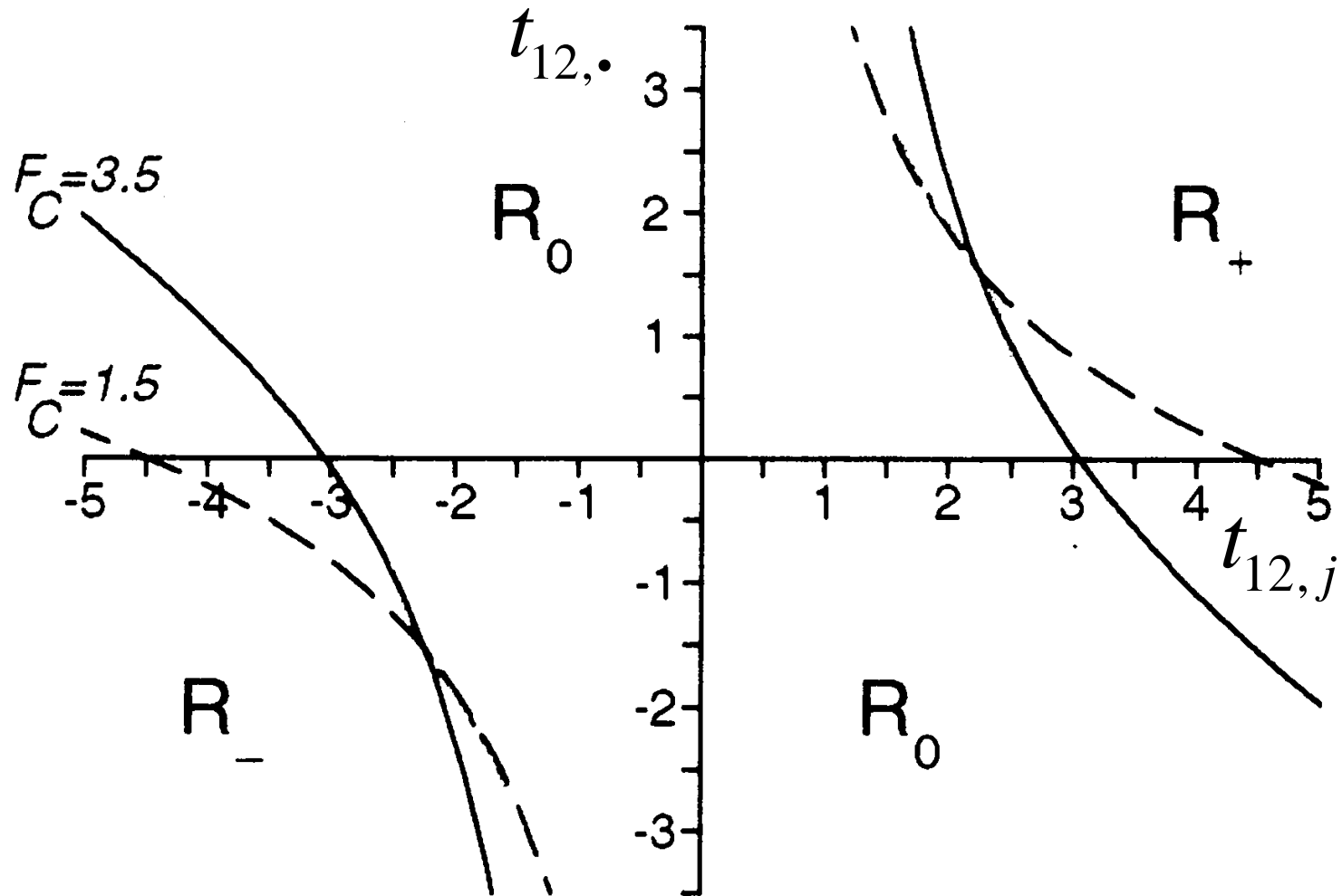
Bayes Critical z Value

$$\text{Decide } \delta_{12,j} > 0 \qquad \text{if}$$

$$z_{12,j} > \frac{z_k}{\sqrt{S_C}}\left\{\frac{S_A}{bS_C} + \frac{b-1}{b}\right\}^{1/2} - \frac{z_{12,\bullet}}{\sqrt{b}}\left\{\frac{S_A}{S_C} - 1\right\}$$

Linear dependence on standardized marginal treatment effect $z_{12,\bullet}$

$\uparrow$ with $\downarrow$ interaction $(\uparrow S_A / S_C)$

$\downarrow$ with $\uparrow$ # subgroups *b*.

Fully Bayes Critical t Boundaries

Concluding Remarks

- Bayesian subgroup analysis offers efficiency in estimation.

- Using a Bayesian hierarchical model,

  - a sample estimate of a subgroup treatment effect that is large and clinically impressive in magnitude could be shrunk to a much smaller, less compelling value.

  - Such dramatic shrinkage could suggest that the large point estimate may have been a random high due to multiple testing and large standard errors within subgroups.

Concluding Remarks

- Bayesian subgroup analysis may be most useful for clinical studies with small samples sizes (feasibility, exploratory, Phase II) or for rare event endpoints (safety).

- In clinical validation studies (pivotal, Phase III), regulatory authorities may still prefer the frequentist approach to formally control operating characteristics (studywise type 1 error).

- When the direction of subgroup effects can be anticipated (e.g., biomarker measurement of the target of a drug), a priori exchangeability of subgroups may not be warranted.

Concluding Remarks

- Tools are available for assessing Bayesian model fit
  - Exchangeability of subgroups can be assessed a posteriori by computing the Bayesian p value of a sensitive statistic based on its predictive distribution (Pennello, Thompson, 2008).

- Bayesian hierarchical models and false discovery rate:
  - For some loss functions, the directional FDR is controlled
    - Lewis C and Thayer DT. A loss function related to the FDR for random effects multiple comparisons. *J Stat Plan Inference* 2004; 125, 49-58.
  - The posterior probability of a null hypothesis is like a *local* false discovery rate
    - Efron B, Tibshirani R, Storey JD, Tusher V . Empirical Bayes Analysis of a Microarray Experiment, *J Amer Statist Assoc* 2001; 96, 1151-60.

References

- Berry DA. Subgroup analyses. *Biometrics* 1990; 46: 1227–30.
- Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: A three-level hierarchical mixture model. *Biometrics* 2004; 60: 418-426.
- Benjamini Y., Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 1995; 57, 289-300.
- Dixon DO, Simon R. Bayesian Subset Analysis, *Biometrics* 1991; 47(3), 871-81.
- Dixon DO, Simon R . Bayesian subset analysis in a colorectal cancer clinical trial. *Stat Med* 1992; 11: 13-22.
- DuMouchel, W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Amer. Statist.* 1999; 53, 170-190.
- Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes Analysis of a Microarray Experiment, *J Amer Statist Assoc* 2001*;* 96: 1151-60.
- FDA. *Draft Guidance: Design Considerations for Pivotal Clinical Investigations for Medical Devices.* http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Guidance Documents/ucm265553.htm, 2011.
- FDA. *Guidance for Use of Bayesian Statistics in Medical Device Clinical Trials.* http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Guidance Documents/ucm071072.htm, 2010.

References

- Gelman A, Hill J, Yajima M. Why We (Usually) Don't Have to Worry About Multiple Comparisons, *J Research on Educational Effectiveness* 2012; 5: 189-211.

- Houwelingen HCV, Thorogood J. Construction, validating and updating of a prognostic model for kidney graft survival, *Stat Med*, 1995; 14: 1999-2008.

- Jones HE, Ohlssen DI, Neuenschwander B, Racine A, Branson M. Bayesian models for subgroup analysis in clinical trials. *Clinical Trials* 2011; 8: 129–143

- Freidlin B, Korn EL. Borrowing Information across subgroups in Phase II Trials: Is It useful? *Clin Cancer Res* 2013; 19(6); 1–9.

- Lewis C, Thayer DT. A loss function related to the FDR for random effects multiple comparisons. *J Stat Plan Inference* 2004; 125, 49-58.

- Lipsky AM, Gausche-Hill M, Vienna M, Lewis RJ. The Importance of "Shrinkage" in Subgroup Analyses, *Annals of Emergency Medicine* 2010; 55(6): 544-552.

- Madigan D, Ryan P, Simpson S, Zorych I. Bayesian Methods in Pharmacovigilance, in *Bayesian Statistics 9*, Bernardo, Bayarri, Berger, Dawid, Heckerman, Smith and West, eds. Oxford University Press, 2010.

- Millen BA, Dmitrienko A, Ruberg S, and Shen L. A Statistical Framework for Decision Making in Confirmatory Multipopulation Tailoring Clinical Trials. *Drug Inf J* 2012 ; 46(6): 647-656..

References

- Pennello, GA. The k-ratio multiple comparison Bayes rule for the balanced two-way design, *J Amer Statist Assoc* 1997; 93: 675-684.

- Pennello G, Thompson L. Experience with reviewing Bayesian medical device trials. *J Biopharm Stat* 2008; 18: 81–115.

- Pocock, Assmann, Enos, Kasten. Subgroup analysis covariate adjustment and baseline comparisons in clinical trial reporting Practice and problems. *Stat Med* 2002; 21:2917–2930.

- Sivaganesan S, Laud PW, Muller P. A Bayesian subgroup analysis with a zero-enriched Polya Urn scheme, *Stat Med* 2010; 30: 312-23.

- Segal JB, Weiss C, Varadhan R for the Center for Medical Technology and Policy. Understanding Heterogeneity of Treatment Effects in Pragmatic Trials With an Example of a Large, Simple Trial of a Drug Treatment for Osteoporosis. 2012, http://cmtpnet.org/wp-content/uploads/downloads/2012/03/Segal-Heterogeneity-in-Pragmatic-Trials.pdf.

- Simon R. Bayesian subset analysis: Application to studying treatment-by-gender interactions. *Stat Med* 2002; 21: 2909–16.

- Thall PF, Wathen JK. Bayesian designs to account for patient heterogeneity in phase II clinical trials. *Cur Opin Oncol* 2008; 20: 407–411.

- Weizemann TR. *Sex-specific Reporting of Scientific Research: a Workshop Summary*. Washington, DC: Institute of Medicine, National Academies Press; 2012.