

Use of Bayesian Hierarchical Models in the Presentation of Subgroup Analyses*

Mark Rothmann

U.S. Food and Drug Administration

DIA Bayesian Scientific Working Group Webinar Series

November 19, 2021

Disclaimer

* The views expressed in this talk are those of the speaker and not necessarily those of the FDA

Outline

- Drug Trials Snapshots (DTS)
- Heterogeneous Treatment Effects
- Different Perspectives
- Story
- Shrinkage Estimation
- Bayesian Hierarchical Models
- Examples
- Concluding Remarks



Drug Trials Snapshots (DTS)

FDASIA 2012 Section 907

- SEC. 907. REPORTING OF INCLUSION OF DEMOGRAPHIC SUBGROUPS IN CLINICAL TRIALS AND DATA ANALYSIS IN APPLICATIONS FOR DRUGS, BIOLOGICS, AND DEVICES.
- (1) IN GENERAL.—Not later than 1 year after the date of enactment of this Act, the Secretary, acting through the Commissioner, shall publish on the Internet web site of the Food and Drug Administration a report, consistent with the regulations of the Food and Drug Administration pertaining to the protection of sponsors' confidential commercial information as of the date of enactment of this Act, addressing the extent to which clinical trial participation and the inclusion of safety and effectiveness data by demographic subgroups including sex, age, race, and ethnicity, is included in applications submitted to the Food and Drug Administration, and shall provide such publication to Congress



Message from Margaret A. Hamburg, M.D., Commissioner of Food and Drugs

One of the core tenets of rigorous biomedical research, as well as a guiding principle of the FDA's goal to meet the health needs of patients across the demographic spectrum, is the importance of encouraging diversity in clinical trials.

When a more diverse population participates in clinical trials, we increase the potential to know more about the extent to which different subgroups—males and females, young and old, people of various racial and ethnic backgrounds, and patients with differing comorbid diseases and conditions—might respond to a medical product. And when subgroup data are analyzed, we have available more information about the product that can be communicated to the public. The result is greater assurance in the safety and effectiveness of the medical products used by a diverse population.



Message from Margaret A. Hamburg, M.D., Commissioner of Food and Drugs (continue)

- ... Advances in science are also playing an increasingly important role in deepening our understanding of how patients within various subgroups respond to medical products. ... an action plan outlining “recommendations for improving the completeness and quality of analyses of data on demographic subgroups in summaries of product safety and effectiveness data”

Purpose of DTS



- Provide consumers and healthcare professionals with information about who participated in clinical trials that supported the FDA approval of new drugs (NMEs and original biologics).
- Provide information on study design.
- Highlight potential differences in efficacy and safety results among gender, race, and age subgroups.
- Increase transparency.

Snapshots are not a PI

Drug Trials Snapshot

- Intended for public
- Consumer-friendly Language
- Focus on subgroup data and analyses
- Links to PI and reviews in Drugs@FDA
- Published on fda.gov 30 days after drug approval

Package Insert

- Intended for Physicians
- Technical Language
- Comprehensive Resource for drug information
- No links to reviews
- Finalized with drug approval

Another difference between snapshots and package inserts

Drug Trials Snapshots

Subgroup analyses, interpretations and conclusions are those of FDA

Package Inserts

Negotiated between FDA and sponsor

Audience

- Should consider audience when constructing a DTS
 - What is the effect someone like me can expect?
- Increase usage of subgroup analysis in Safety
 - treatment-related adverse events can be more frequent as age increases

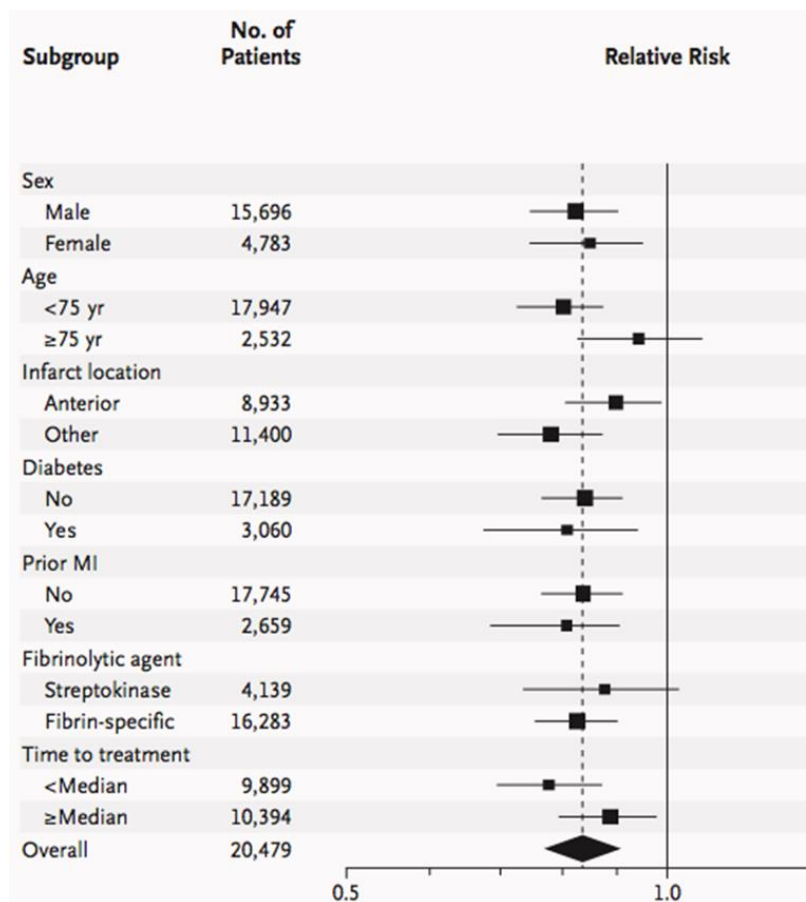


Heterogeneous Treatment Effects (HTE)

Treatment effects vary across subgroups of a factor

- Factor an effect modifier or
- Factor associated/correlated with one or more effect modifiers

Conventional subgroup analysis (1 of 3)



Conventional subgroup analysis (2 of 3)

- Subject to random highs and random lows
 - Estimated treatment effects vary more than underlying treatment effects
- Analysis are typically univariate or marginal - one subgrouping variable at a time
- Possibly confounded by correlation between subgrouping variables
 - If sex and age are correlated, the difference in treatment efficacy between men and women may be confounded by age
 - Groenwold (2009) - Aspirin's effect on stroke was larger in women, but women were older

Conventional subgroup analysis (3 of 3)

- Assume treatment effect equal across subgroups (unless compelling evidence that they differ). When estimating treatment effect in interested subgroup
 - Equal relevancy of outcomes for patient outside interested subgroup outcome and patient in interested subgroup
 - Under this assumption analysis (estimation of treatment effect) is simple, easy
- OR use only data from patients in interested subgroup to estimate the treatment effect for interested subgroup
 - Patient outside interested subgroup outcome provides no information (no relevancy)
 - Under this assumption analysis (estimation of treatment effect) is simple, easy

Recent Symposia and Workshops co-sponsored by FDA on Heterogeneous Treatment Effects

- Nov 28, 2018, Symposium of Assessing and Communicating Heterogeneity of Treatment Effects for Patient Subpopulations: Challenges and Opportunities
 - Agenda, Slides and Recording at <https://www.jhsph.edu/research/centers-and-institutes/center-of-excellence-in-regulatory-science-and-innovation/news-and-events/Critical-Issues-in-Heterogeneity-of-Treatment-Effect.html>
- Nov 30 - Dec 1, 2020, Workshop on Heterogeneity of Treatment Effects in Clinical Trials: Methods and Innovations
 - Agenda and Recording at <https://mrctcenter.org/news-events/heterogeneity-of-treatment-effects-in-clinical-trials-methods-and-innovations/#1602863324215-1289c9d5-a82a>



Some messages from the session

- We can do more or better at understanding heterogeneous treatment effects
- Have used shrinkage estimation for some drug trial snapshots
- More complicated/better models could include factors known to affect the treatment effect
- Doable (somewhere) in some settings to have a repository of data to be used to provide individual patient advice which can account for patient preferences, patient demographics and medical history

Some Topics/Questions of interest

- Representation in clinical trials
- What can a patient like me expect?
- Is there consistency of treatment effect?
- If there is a benefit overall, where may there not be benefit?
- Difference in using subgroup analyses to make a claim vs. individual patient treatment decisions
- How do overall results affect how we view subgroup results?



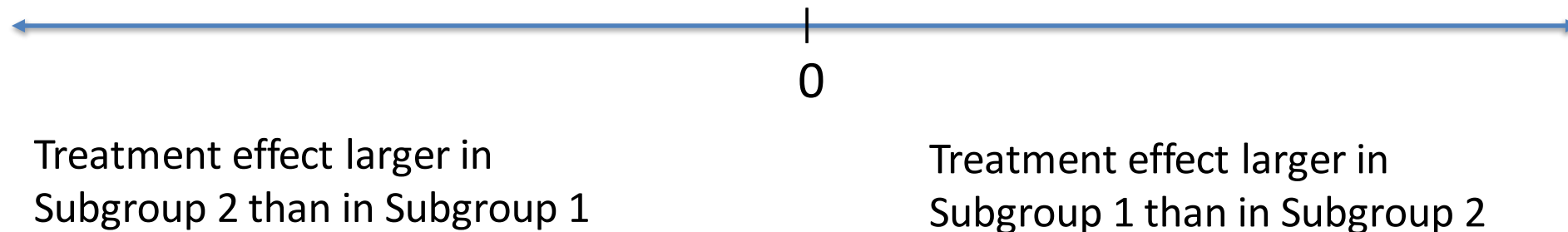
Different Perspectives/ Frameworks

Different subgroup analyses perspectives/frameworks

- L1: Actual (signed) differences in treatment effects vs
- L2: Variability in treatment effects

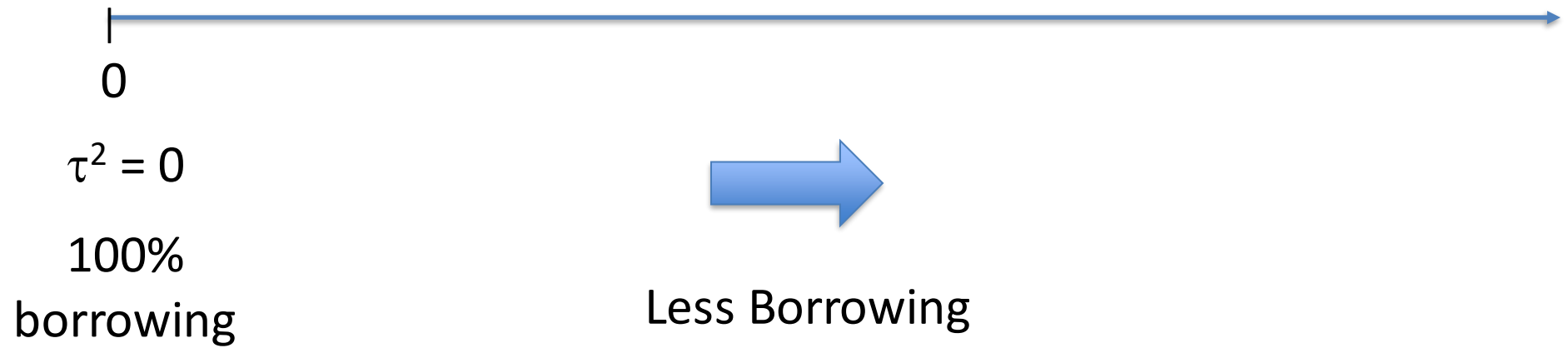
Testing for differences in effects

- Consider two subgroups (e.g., one factor with two levels)
- Interest in order/direction of effects and the difference in effects
- Zero difference in effects is at a key (central) location



Estimating treatment effects across subgroups

- No difference in effects is at an extreme location (i.e., $\tau^2 = 0$)





Story

Introductory Paragraph from the Impact Story

- When evaluating drug treatments, determining how and to what extent a drug works in different patient subgroups can be addressed by statistical approaches that make use of results from every subgroup when understanding the treatment effect for a given subgroup.

[using-innovative-statistical-approaches-provide-most-reliable-treatment-outcomes](#)

Treatment Decisions

- Physicians make treatment decisions on past experience with patients

New Drug

- Sex may or may not be an effect modifier
- No other factor is considered as a possible effect modifier
- Physician has experience on the use of the new drug and outcomes in 2 males and 2 females
- The next patient, a female, is prescribed the new drug from that physician. What can she expect?
 - Probably use information from all four previous patients.
 - Outcomes from females may be more relevant than the outcomes from males.

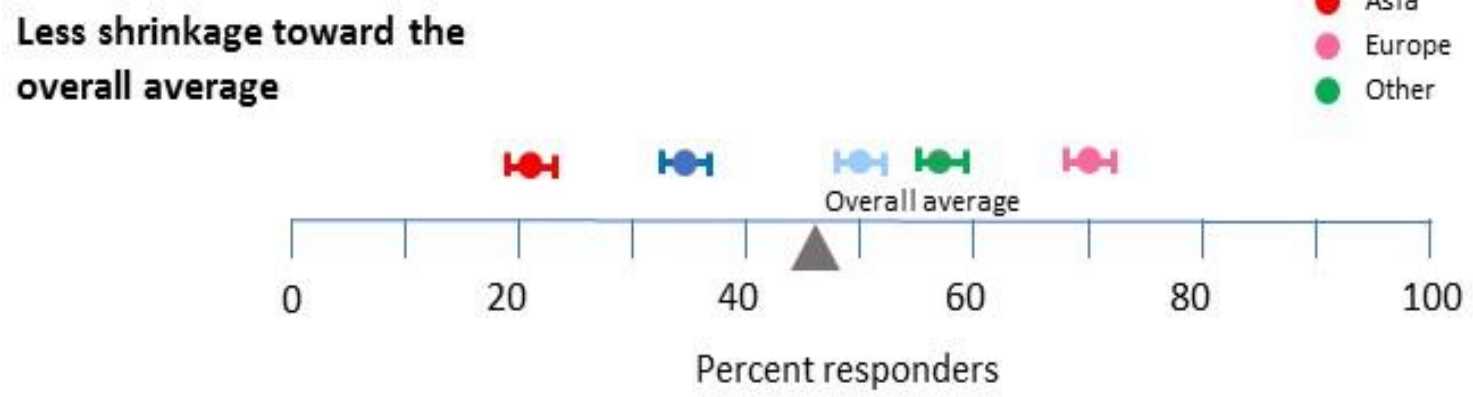
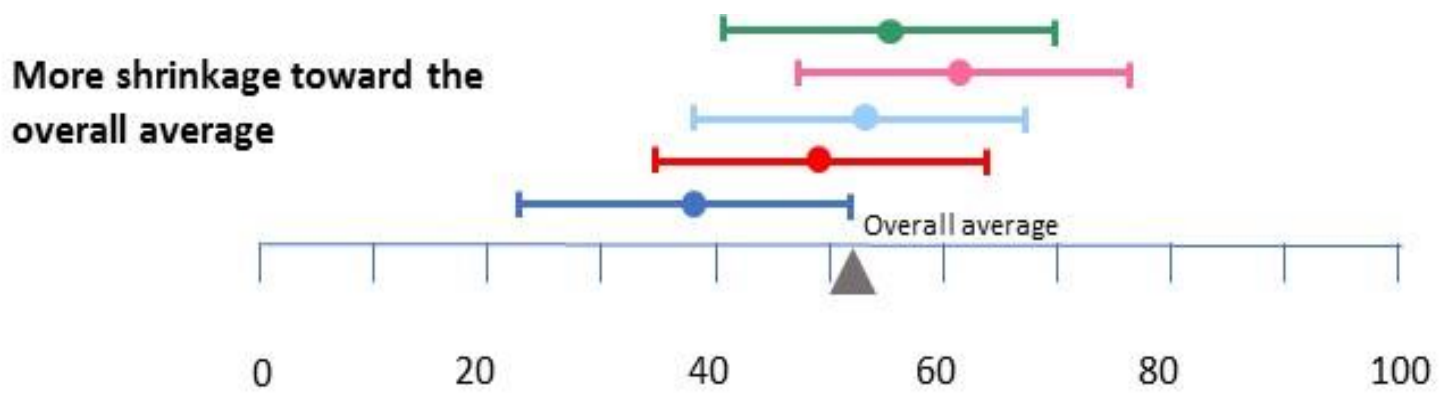
As Previous Experience Grows

- Observe the outcomes from more and more males and females
- The relevancy of the results from males in what the next female can expect. Depends on
 - How similar the results from males are to the results from females
 - How much data on females

Shrinkage Estimation from Bayesian Hierarchical Models Works this Way

- Data from a subject in the given subgroup are more relevant than data from a subject outside that subgroup
 - The relevancy goes to zero as the number of subjects in the subgroup of interest goes to infinity
- The data decide how much borrowing is done
 - Depends on ratio of variability within subgroup to variability between subgroup
 - Less borrowing when within subgroup variability decreases (its sample size increases)

How much shrinkage/borrowing



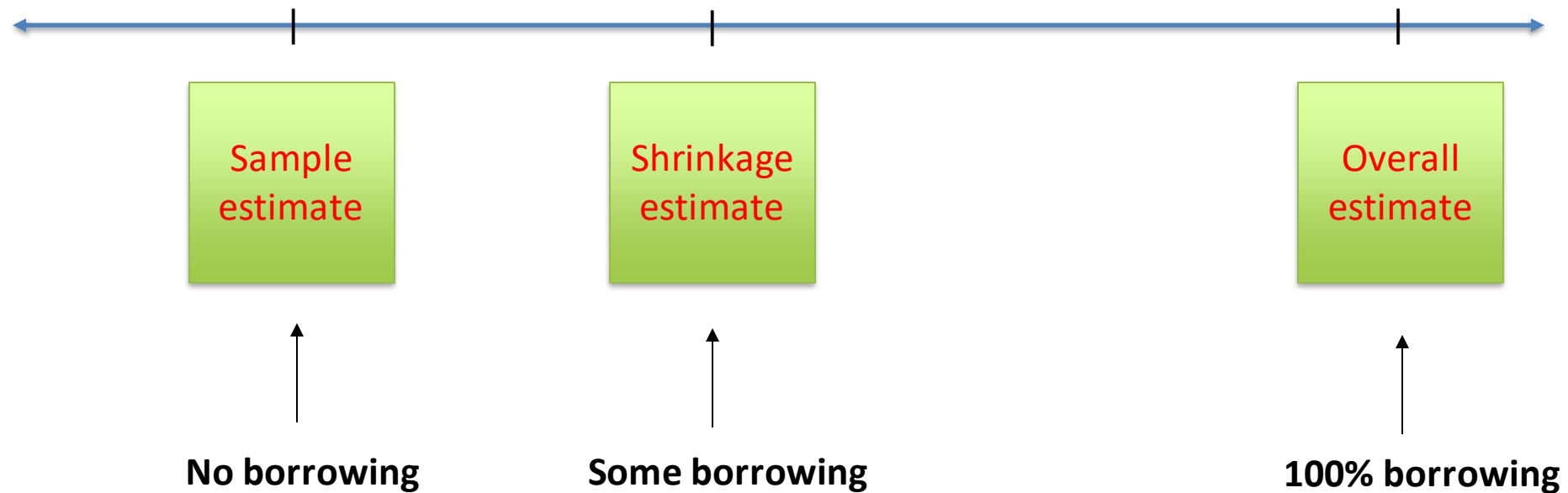


Shrinkage Estimation

What is Shrinkage Estimation?

- A shrinkage estimate of a parameter for a subgroup
 - a weighted average of sample estimate and overall estimate (stratified by subgroup).
 - Could be a posterior mean in a Bayesian setting
 - Sample estimate is “shrunk” towards the overall estimate

Amount of Borrowing



Two components of variability in sample estimates across subgroups

- The total variability in the sample estimates is sum of
 - the within subgroup variability of the sample estimator and
 - the across subgroups variability in the underlying/true parameter values

Purpose of Shrinkage Estimation

- To address within study/subgroup variability in the estimation
 - Obtain estimates where unaccounted variability is across studies/subgroups variability
 - Collection of shrinkage estimates tends to be closer to the collection of true subgroup effects than the sample estimates

Importance of shrinkage in estimation

- Greater precision
 - narrower 95% CIs
- Quantitatively addresses random highs and random lows

Lipsky, A. M., Gausche-Hill, M., Vienna, M., Lewis, R. J. (2010). The importance of "shrinkage" in subgroup analyses. *Annals of Emergency Medicine*. Jun;55(6):544-552

Pennello G., Rothmann M., Bayesian Subgroup Analysis with Hierarchical Models, in *Biopharmaceutical Applied Statistics Symposium Volume 2: Biostatistical Analysis of Clinical Trials*, Eds. Karl E. Peace, Ding-Geng Chen, Sandeep Menon, Springer

Model for Shrinkage Estimation

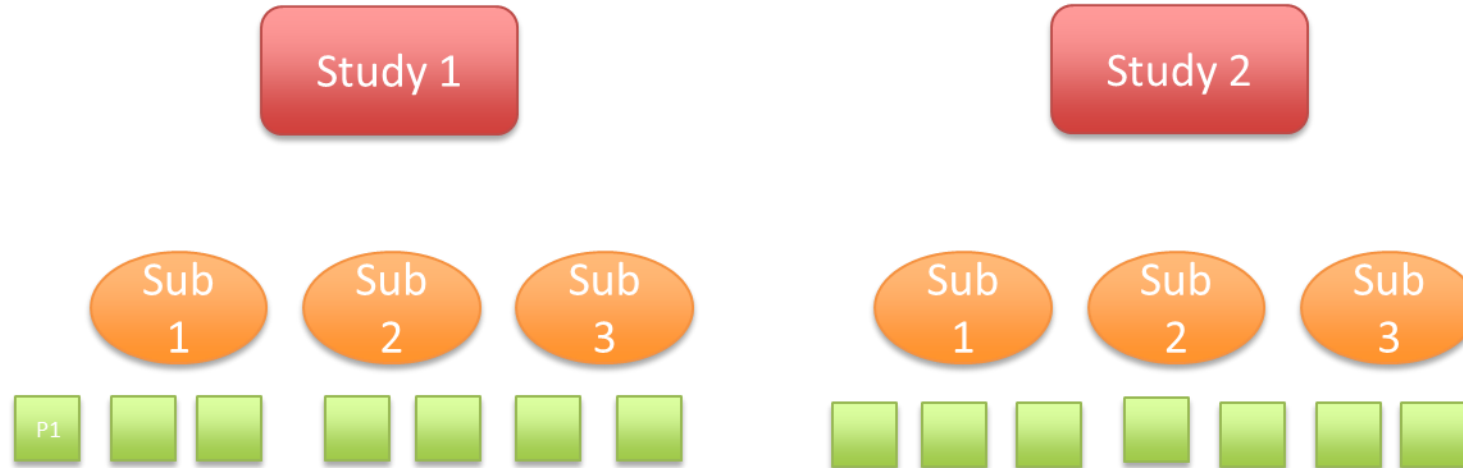
- Do Shrinkage Estimation through Bayesian Hierarchical Models



Bayesian Hierarchical Models

Hierarchical Models

- Statistical model written in multiple levels (hierarchical form).



Linking

- Treatment effects (e.g., across studies, subgroups, products in the same class) are linked. This linking of the parameters makes their estimation linked.

Exchangeability – a starting point

- Subgroups treatment effects are exchangeable if possible orderings of treatment effects are considered equally likely a priori (i.e., before seeing data)
 - This is fair, not favoring any subgroup treatment effects over any other
 - Joint prior distribution of subgroup treatment effects is exchangeable
 - Very unlikely that the joint posterior distribution will be exchangeable

Common choice for a model

- Treatment effects drawn randomly from the same distribution implies that treatment effects are exchangeable

Exchangeability

- Exchangeability of subgroup treatment effects not always a correct assumption.
 - Drug expected to be more effective in subgroup of cancer patients who exhibit molecular target than subgroup of patients without target.
- Shrinkage analysis incorrectly assuming exchangeability still accounts for a component of variability

Smallest Expected Mean Square Error

- Estimation error: $\hat{\theta}_i - \theta_i$
- Squared error: $(\hat{\theta}_i - \theta_i)^2$
- Mean squared error at particular true effect: $E((\hat{\theta}_i - \theta_i)^2 \mid \theta_i)$
- Expected Sum of squared error (expectation over the parameter space):
 $E\left(E(\sum_{l=1}^k (\hat{\theta}_l - \theta_l)^2 \mid \theta_1, \dots, \theta_k)\right)$
- Relative to joint prior distribution over parameter space, shrinkage estimation gives smallest expected sum of square error for collection of subgroup treatment effect across all joint estimators
 - And within each subgroup, the optimal estimator



Example

LEADER trial

- Cardiovascular outcome trial
- Liraglutide vs. placebo
- Time to first major adverse cardiac event
- Rule out a hazard ratio greater than 1.3

- Overall Result: HR =0.87 95% CI (0.78, 0.97)

Subgroup Analyses

Region	Results
	HR (95% CI)
Asia	0.62 (0.37, 1.04)
Europe	0.82 (0.68, 0.98)
North America	1.01 (0.83, 1.22)
The Rest of The World	0.83 (0.68, 1.03)

Modeling Assumptions

- $1/\tau^2$ is distributed Gamma (.001, .001)
- μ is distributed Normal mean 0, variance 16
- μ_i is distribution Normal mean μ , variance τ^2 $i = 1, 2, 3, 4$
- For $i = 1, 2, 3, 4$ Y_i represents the observed subgroup log hazard ratio
- Y_i is distributed Normal mean μ_i variance σ_i^2 where
 - $\sigma_1^2 = 0.0688$, $\sigma_2^2 = 0.0088$, $\sigma_3^2 = 0.0094$ and $\sigma_4^2 = 0.114$

Defines Joint Prior Distribution

- $1/\tau^2$ is distributed Gamma (.001, .001)
- μ is distributed Normal mean 0, variance 16
- μ_i is distribution Normal mean μ , variance τ^2 $i = 1, 2, 3, 4$
- For $i = 1, 2, 3, 4$ Y_i represents the observed subgroup log hazard ratio
- Y_i is distributed Normal mean μ_i variance σ_i^2 where
 - $\sigma_1^2 = 0.0688$, $\sigma_2^2 = 0.0088$, $\sigma_3^2 = 0.0094$ and $\sigma_4^2 = 0.114$

FDA Shrinkage Analysis

Region	Sample estimate		Bayes Shrinkage estimate	
	HR	95% CI	HR	95% CI
Asia	0.622	(0.372, 1.040)	0.803	(0.591, 1.089)
Europe	0.815	(0.678, 0.979)	0.836	(0.715, 0.978)
North America	1.010	(0.835, 1.220)	0.936	(0.786, 1.115)
The Rest of the World	0.833	(0.676, 1.027)	0.847	(0.716, 1.003)

FDA Shrinkage Analysis

Region	Sample estimate		Bayes Shrinkage estimate	
	HR	95% CI	HR	95% CI
Asia	0.622	(0.372, 1.040)	0.803	(0.591, 1.089)
Europe	0.815	(0.678, 0.979)	0.836	(0.715, 0.978)
North America	1.010	(0.835, 1.220)	0.936	(0.786, 1.115)
The Rest of the World	0.833	(0.676, 1.027)	0.847	(0.716, 1.003)

FDA Shrinkage Analysis

	Sample estimate		Bayes Shrinkage estimate	
Region	HR	95% CI	HR	95% CI
Asia	0.622	(0.372, 1.040)	0.803	(0.591, 1.089)
Europe	0.815	(0.678, 0.979)	0.836	(0.715, 0.978)
North America	1.010	(0.835, 1.220)	0.936	(0.786, 1.115)
The Rest of the World	0.833	(0.676, 1.027)	0.847	(0.716, 1.003)



Example - ACR20 response

Bayesian hierarchical model assumptions

- For $i = 1, \dots, k$, Y_i represents the estimated difference in ACR20 response in a subgroup level i , assume Y_i approximately $N(\mu_i, \sigma_i^2)$ where
- σ_i^2 are the estimated CMH weighted variance of the difference in ACR20 in subgroup level i
- $\mu_i \sim N(\mu, \tau^2)$
- $\mu \sim N(0, \omega^2)$, $1/\tau^2 \sim \text{Gamma}(0.001, 0.001)$
- ω was chosen to be 4

Table 1. Effects of Upadacitinib on Percent ACR20 Responders by Subgroups, Methotrexate-Controlled Trial

Demographic Parameter	ACR20 % (n/N)		Treatment Difference ^a
	Methotrexate	Upadacitinib	(95% CI)
Sex			
Male	43% (16/37)	60% (26/43)	21% (2%, 40%)
Female	41% (73/179)	70% (121/174)	28% (18%, 38%)
Age in Years			
Younger than 40	52% (12/23)	82% (23/28)	27% (10%, 45%)
40 to 64	39% (58/148)	68% (100/147)	28% (18%, 38%)
65 or Older	42% (19/45)	57% (24/42)	21% (4%, 38%)
Race			
White	43% (76/176)	68% (118/173)	26% (16%, 35%)
Black / African American	45% (5/11)	33% (5/15)	6% (-32%, 43%)
Asian	29% (7/24)	88% (21/24)	51% (27%, 75%)

Shrinkage Estimates for Treatment Difference

^a Treatment differences and credibility intervals may not match value of (treatment - control) since estimates include relevance of outcomes from other subgroups



Concluding Remarks

Other models

- A one-way hierarchical model that also accounts for an effect modifier
 - Rothmann, M., [*Applying Hierarchical Models When Evaluating Treatment Effects Across Regions*](#) presented at 2018 ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop
- A multi-way hierarchical model
 - Example from the SOLVD trial in “Bayesian analysis of heterogeneous treatment effects for patient-centered outcomes research” by Henderson, Louis, Wang and Varadhan

Let's Do Things Better

- Greater precision
 - narrower 95% CIs
- Addresses random highs and random lows
 - 95% CIs have 95% coverage after the data are known
- When exchangeability assumptions do not hold, can possibly do even better

Chapter from a new book

Pennello G., Rothmann M., Bayesian Subgroup Analysis with Hierarchical Models, in Biopharmaceutical Applied Statistics Symposium Volume 2: Biostatistical Analysis of Clinical Trials, Eds. Karl E. Peace, Ding-Geng Chen, Sandeep Menon, Springer

Thank You!
Questions!