How do we combine two treatment arm trials with multiple arms trials in IPD meta-analysis?

An Illustration with College Drinking Interventions

David Huh, PhD[1], Eun-Young Mun, PhD[2], & David C. Atkins, PhD[1]

[1] Department of Psychiatry & Behavioral Sciences, University of Washington

[2] Center for Alcohol Studies, Rutgers, The State University of New Jersey

Modern Modeling Methods Conference
May 20, 2014

Supported by NIAAA grants
R01 AA019511
T32 AA007455

IPD opens the door to new possibilities…

- Meta-analysis of individual participant-level data (IPD) opens the door to a greater variety of research hypotheses that can be tested, yet it's rarely done in the social sciences.

- Provides a means of combining information across studies more accurately.
  - Compared with traditional methods based on summary statistics, IPD-based meta-analysis can be more flexibility tailored to the characteristics of the data and study designs.

- A challenge in meta-analysis†, including with IPD:

  How to combine studies with varying numbers of treatments.
  - Most randomized trials (> 78%) are two arm studies‡, however, multiple arm trials are not uncommon.
  - Little discussion in the IPD meta-analysis literature about how to combine studies with varying numbers of arms.

†Gleser & Olkin, 2009; ‡ Hopewell, Dutton, Yu, Chan & Altman, 2010

IPD meta-analysis can accommodate varying arms and other data characteristics

- The appropriate combination of studies with varying numbers of arms was a key consideration in an IPD meta-analysis that our research group (Project INTEGRATE)† undertook of college drinking interventions.

- Other important analytic issues:
  - Differing number of assessments
  - Confounders and moderators of intervention outcome
  - Normally-distributed and zero-inflated count outcomes

- Ultimately settled a novel formulation of a Bayesian multilevel model that retained all the available data and accommodated differing numbers of treatment groups.

† Mun et al., 2014

A real-world application with drinking interventions

- For over two decades, brief motivational interventions (BMIs) have been implemented on college campuses to reduce heavy drinking and related negative consequences.
  - Recommended as a prevention strategy by the National Institute on Alcohol Abuse and Alcoholism (NIAAA).†
- Such interventions include:
  - In-person motivational interviews with personalized feedback (MI+PF)
  - Group motivational interviews (GMI)
  - Stand-alone PF interventions delivered via mail, computer, or the Web.
- Meta-analytic reviews using aggregate data from published studies suggest their short-term efficacy, but the effects vary.
  - Carey and colleagues‡ found that across 62 studies, 50% of tests of intervention outcomes were statistically significant.
  - Significant findings were associated with small effect sizes.

† NIAAA, 2002; ‡ Carey , Scott-Sheldon, Carey, DeMartini, 2007

Building on previous systematic reviews with IPD meta-analysis

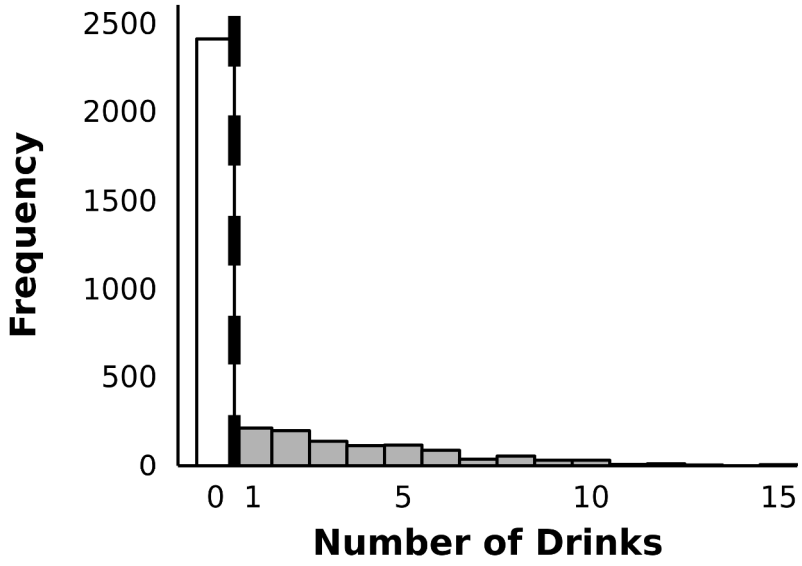
- Systematic reviews to-date have limitations
  - Effects at different time-points evaluated with different subsets of studies.
  - Moderators evaluated at the study-level (e.g., % female vs. male).
  - Alcohol outcomes are often highly skewed with many zeroes.
    - Both Gaussian and traditional count models under-represent the actual frequency of zeroes.†

- More analytic options with IPD compared with classical meta-analysis using aggregate data.
  - Ability to control for participant-level covariates.
  - Model can be easily extended to evaluate individual-level moderators.
  - Distribution-appropriate analysis

† Atkins, Baldwin, Zheng, Gallop, & Neighbors, 2013

Important to attend to excess zeroes…

- Distribution of the data is another important consideration.
  - Behavioral outcomes assessing short intervals will often contain a lot of zeroes.
    - Substance use
    - Sexual behavior



- Zeroes may be a key feature of the outcome and not just a nuisance of the data.
  - An intervention may have an effect on either:
    - **The decision to drink (zero drinks vs. 1 or more drinks)**
    - **The number of drinks once started (1, 2, 3, …)**

What IPD meta-analysis options are available?

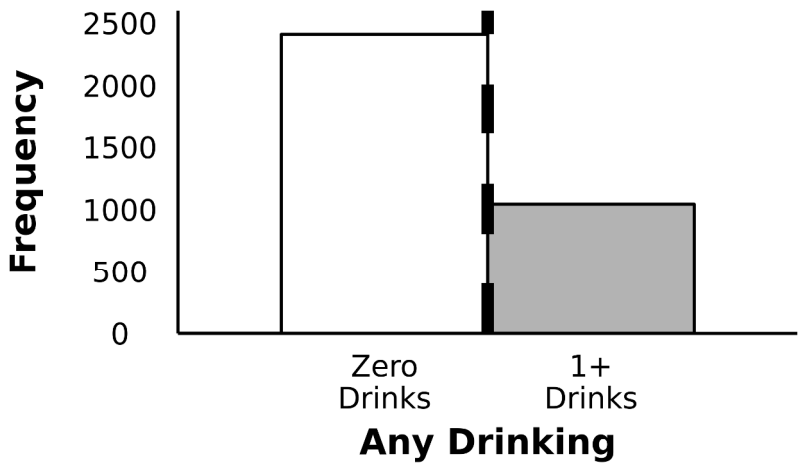
- Two-stage IPD meta-analysis
  - The most common†
  - Raw data are converted into standardized effect sizes.
    - For continuous data (*d* and *g*)
    - For dichotomous and count data (*OR*, *RR*)
  - Standardized effect sizes are pooled.

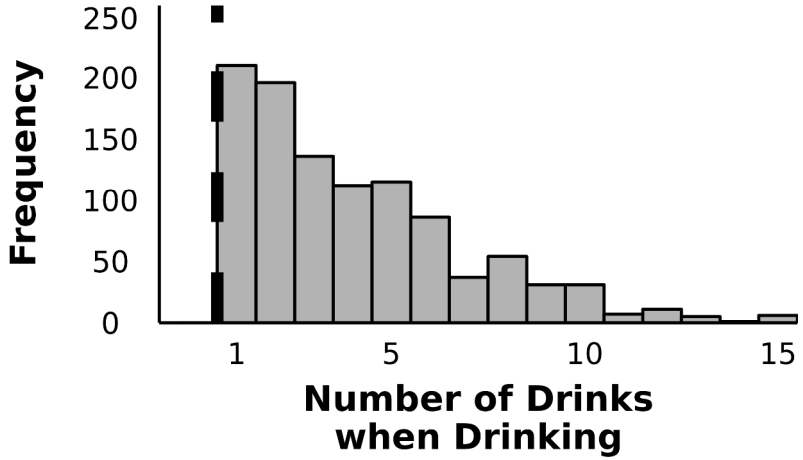
- Single-stage IPD meta-analysis
  - We have the raw data, why not use it?
    - Less variation in IPD-generated estimates, thus greater power.
    - Participant-level covariates can be incorporated.
  - Greater variety of statistical models at our disposal.

† Cooper & Patall, 2009

Accounting for zero-inflated outcomes using a hurdle model

- Hurdle models, a type of two-part model are appropriate for zero-inflated count data, such as drinking.†
- A threshold must be crossed from zero into positive counts.
- The outcome is effectively divided into two parts.
  - No drinking vs. any drinking:
    - Logistic regression
  - Amount of drinking when drinking:
    - Zero-truncated Poisson or Negative binomial regression





† Huh, Kaysen, & Atkins, 2014

An example with longitudinal IPD

- Project INTEGRATE
  - One of the largest IPD meta-analysis projects to-date evaluating brief motivational interventions for college drinking.[†,‡]
  - Focused on randomized controlled studies evaluating one or more BMIs:
    - Individual Motivational Interview with Personalized Feedback
    - Standalone Personalized Feedback
    - Group Motivational Interview
  - IPD sample included 17 studies of 8,275 individuals
    - 14 two-arm studies
    - 2 three-arm studies
    - 1 four-arm study
  - 2 – 5 repeated measures up to 12 months post-baseline

† Mun et al., 2014; ‡ Huh et al., 2014

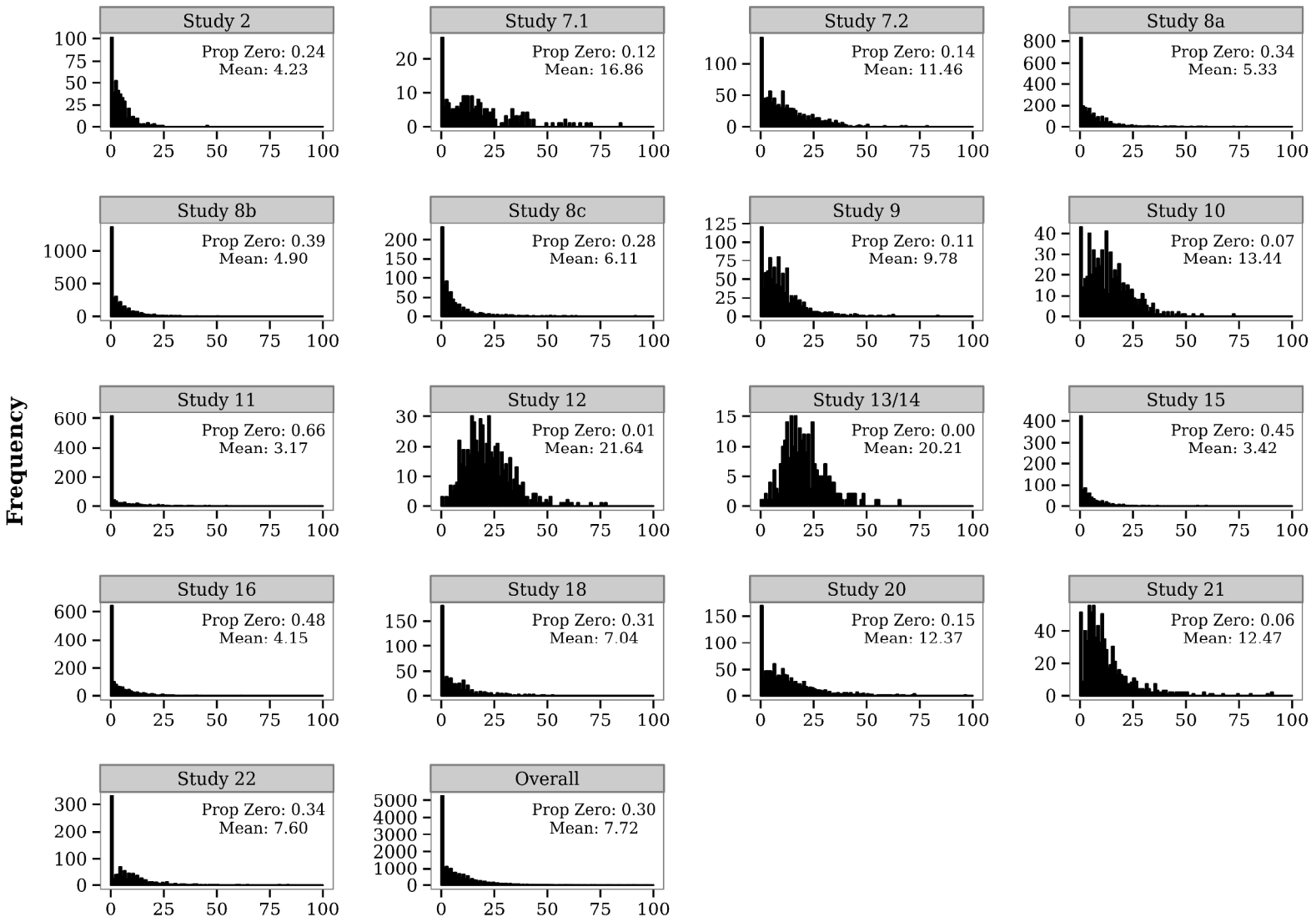
The longitudinal drinking outcomes

- **Total drinks in a typical week**
  - Daily Drinking Questionnaire (DDQ)†
  - Zero-inflated count variable.

- **Alcohol Problems**
  - Six questionnaires used to derive latent trait scores.
    - E.g., Rutgers Alcohol Problem Index (RAPI), Alcohol Use Disorders Identification Test (AUDIT)
  - Relatively normally-distributed outcome

† Collins et al., 1985

Frequencies of Drinks per Week by Study



The Analytic Approach Used…

- Bayesian Multilevel Modeling (MLM)†
  - Markov-chain Monte Carlo estimation
    - MCMCglmm package in R‡,*
  - Permitted distribution-appropriate analysis
    - Hurdle Poisson model for zero-inflated drinking outcome
      - Logistic regression
        - No drinking vs. any drinking
      - Truncated Poisson regression
        - Number of drinks when drinking
    - Gaussian Model for alcohol problems outcome
      - Relatively normally-distributed

† Gelman & Hill, 2006; ‡ Hadfield, 2010; * R Core Team, 2013

Why Bayesian and not maximum likelihood estimation?

- MCMC sampling yields a complete distribution of the regression coefficients and random effects, rather than a single point estimate for each parameter in an ML (frequentist) model.
  - Why this is important:

    Random effects for each treatment group can be estimated with uncertainty (i.e., confidence intervals).

The first model attempted: A 3-level model (3: Study → 2: Participant → 1: Observation)

$$\begin{aligned}\text{OUTCOME}_{t>0,ig} = \\ b_0 + b_1\text{OUTCOME}_{t=0,ig} + b_2\text{COVARIATE}_{ig} + b_3\text{MI\_PFP}_g \\ + b_4\text{PFP}_g + b_5\text{GMI}_g + u_{0g} + u_{1g}\text{MI\_PFP}_g + u_{2g}\text{PFP}_g \\ + u_{3g}\text{GMI}_g + r_{0ig} + e_{tig}\end{aligned}$$

- Study is the highest level of the model.
- Study-specific treatment effects (random slopes) are included for each distinct intervention type.
- This model has intuitive appeal.

Illustrating with Project INTEGRATE

Studies (17)

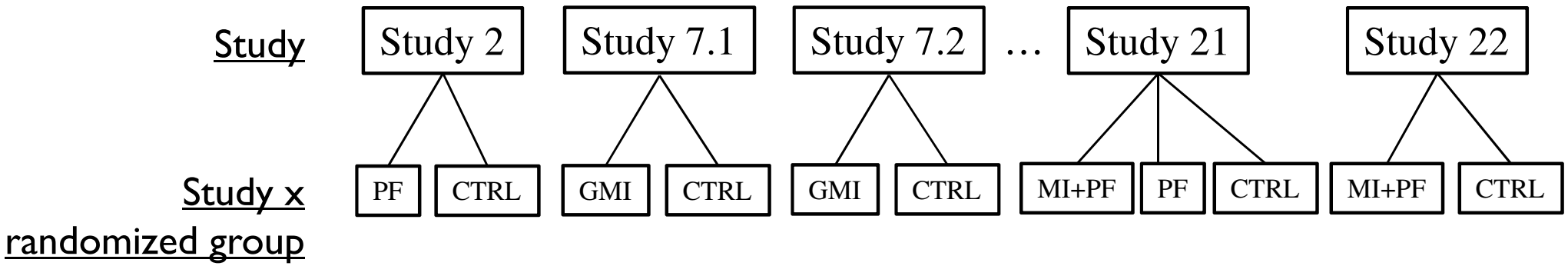
Treatment Types (3)	2	7.1	7.2	8a	8b	8c	9	10	11	12	13/14	15	16	18	20	21	22
MI+PF	×	×	×	×	×	×	✓	✓	×	✓	✓	×	×	×	✓	✓	✓
PF	✓	×	×	✓	✓	✓	✓	×	✓	×	✓	×	×	✓	×	✓	×
GMI	×	✓	✓	×	×	×	✓	×	×	×	×	✓	✓	×	×	×	×

- Problem: Not all treatments evaluated in each study, so the resulting model is rank deficient.
  - 51 possible treatment by study combinations
    - 30 combinations (59%) don't exist.
- Model does not converge using diffuse default priors.

The model with study at the highest level doesn't work, what are our options?

- Keep the model as-is, but use a more informative prior for the random effects.
  - Is it worth that much effort to get the model to work?
  - Informative priors have their critiques and drawbacks.
- Pool active intervention conditions within a study or remove one or more conditions.
  - Reduces each study to a 2-arm RCT design.
  - Potential loss of information
- **Exclude the non-existent study by treatment combinations that are making the model rank deficient.**

Defining study × randomized group at the highest level



- The highest level of the model is study by randomized group rather than study.
  - Preserves the randomization within studies in the model.
- There is no fixed effect for treatment.
  - Intervention effect sizes are calculated from the posterior distribution of the randomization group random effects.

The Basic Model: Similar to an ANCOVA

Post-baseline values of the outcome

Unique randomization group random effect

Residual error

$$\mathrm{OUTCOME}_{t>0,ig} = b_0 + b_1\mathrm{OUTCOME}_{t=0,ig} + b_2\mathrm{COVARIATE}_{ig} + u_{0g} + r_{0ig} + e_{tig}$$

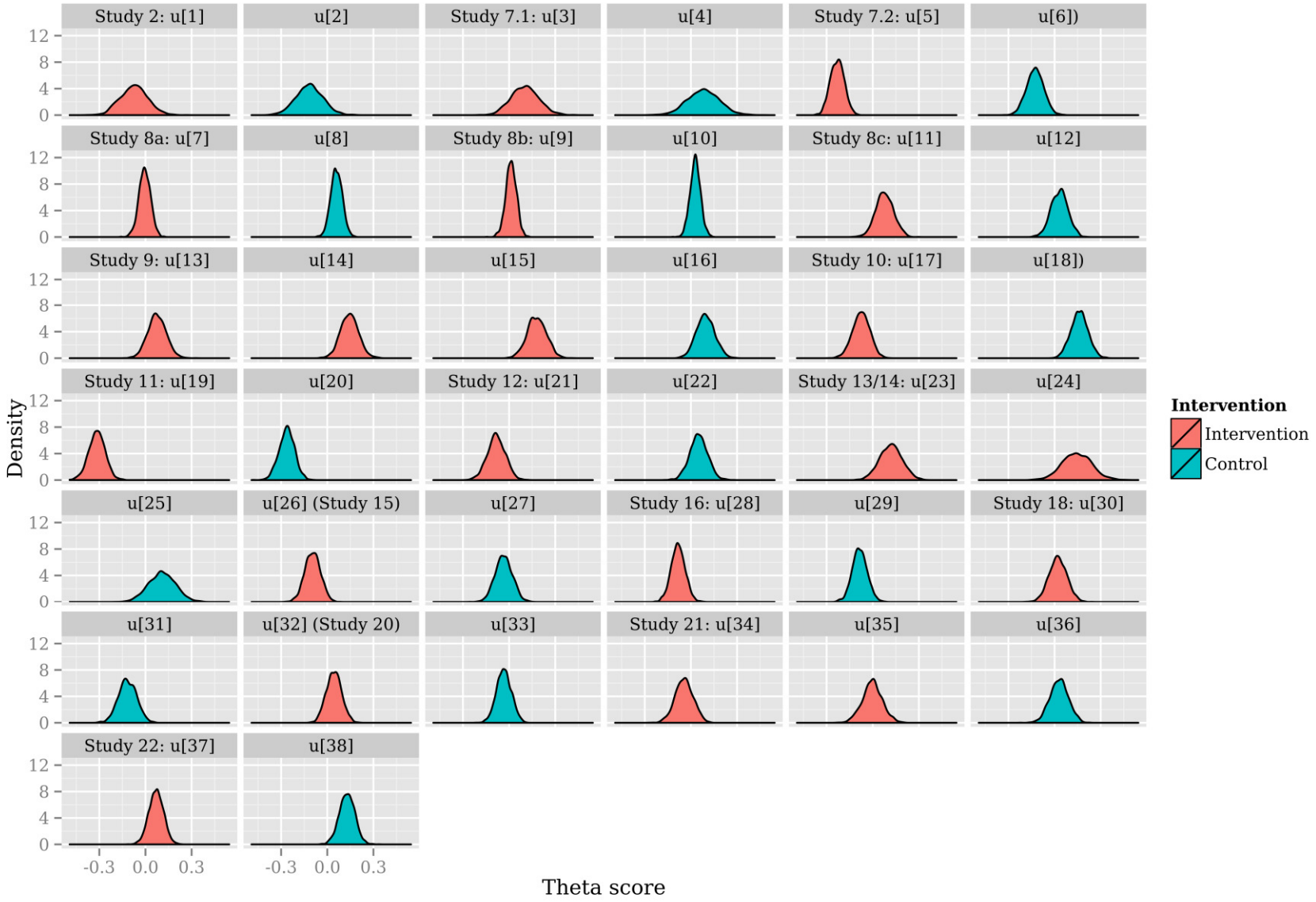
Baseline value of the outcome

Additional covariate(s)

Participant random effect

t = repeated measure
i = individual
g = unique randomization group

Unique randomization group random effects includes intervention and control groups



Calculating the intervention effect

- The key estimates of interest are the samples from the posterior distributions of the random effects for randomization group.
  - Each random effect has its' own distribution of samples.

	Fixed effects		Randomization group effects					
			Study 2		…	Study 21		
(Sample)	b_0	b_1	u_1	u_2	…	u_{34}	u_{35}	u_{36}
1	-0.037	0.670	-0.036	-0.037		-0.070	-0.137	-0.088
2	-0.008	0.675	-0.167	-0.191		-0.009	-0.047	-0.055
3	-0.072	0.680	-0.001	0.100		-0.050	-0.020	0.012
⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮
2000	-0.039	0.660	-0.145	-0.023		-0.019	-0.032	0.062
			Intervention (PF)	Control		Intervention (MI+PF)	Intervention (PF)	Control

Calculating the intervention effect (cont.)

- Example: The intervention effect in Study 2.

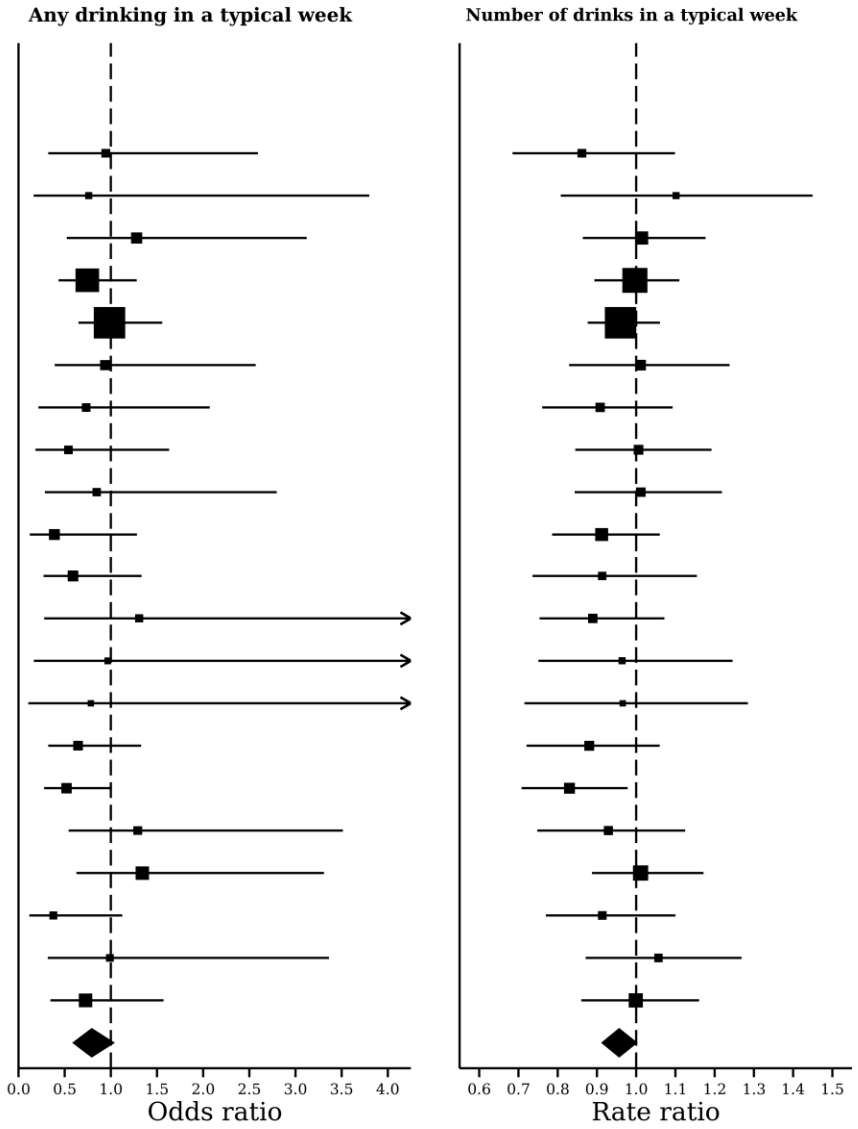
- Three steps to calculating the effect size for a treatment group
  1. Identify the posterior draws from the random effect for an intervention group and its' corresponding control group.
  2. Take the difference ($u_{intervention} - u_{control}$).
  3. Calculate the mean and 95% confidence interval of that difference.

- Repeat for all other intervention groups.

	Study 2		
	Intervention	Control	Effect Size
(Sample)	u_1	u_2	$u_1 - u_2$
1	-0.036	-0.037	0.010
2	-0.167	-0.191	0.240
3	-0.001	0.100	-0.999
⋮	⋮	⋮	⋮
2000	-0.145	-0.023	-0.122

Forest Plot for Drinks per Week (Hurdle)

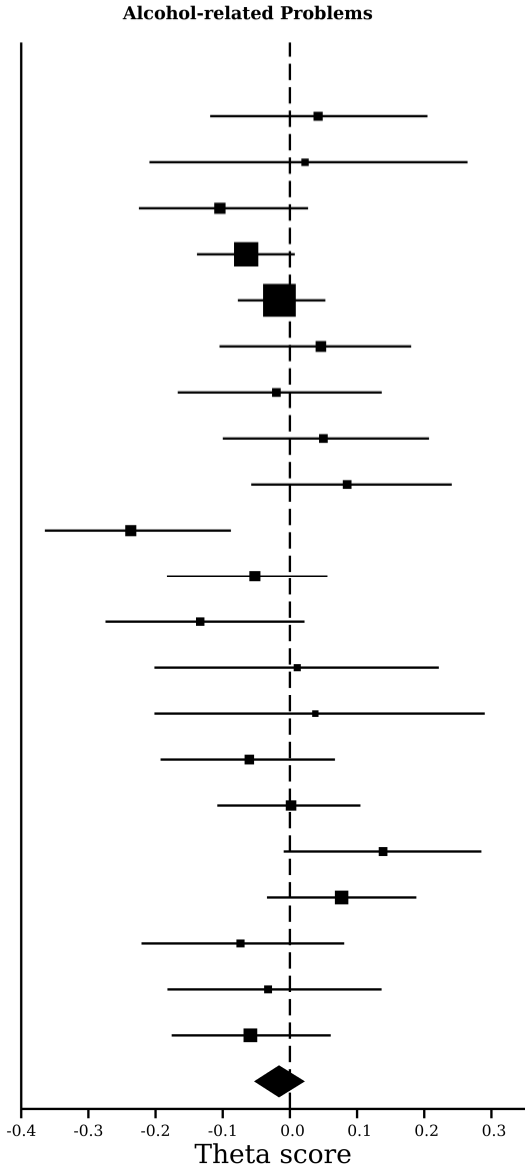
Study	Intervention	Any drinking OR [95% CI]	No. of drinks RR [95% CI]
2	PF	0.95 [0.32, 2.59]	0.86 [0.68, 1.10]
7.1	GMI	0.76 [0.16, 3.80]	1.10 [0.81, 1.45]
7.2	GMI	1.28 [0.52, 3.12]	1.01 [0.86, 1.18]
8a	PF	0.74 [0.43, 1.28]	1.00 [0.89, 1.11]
8b	PF	0.99 [0.65, 1.56]	0.96 [0.88, 1.06]
8c	PF	0.94 [0.39, 2.57]	1.01 [0.83, 1.24]
9	MI+PF	0.73 [0.22, 2.07]	0.91 [0.76, 1.09]
	PF	0.54 [0.18, 1.63]	1.01 [0.84, 1.19]
	GMI	0.85 [0.29, 2.80]	1.01 [0.84, 1.22]
10	MI+PF	0.39 [0.12, 1.28]	0.91 [0.79, 1.06]
11	PF	0.59 [0.27, 1.33]	0.91 [0.74, 1.15]
12	MI+PF	1.31 [0.28, 6.51]	0.89 [0.75, 1.07]
13/14	MI+PF	0.97 [0.17, 5.54]	0.96 [0.75, 1.25]
	PF	0.78 [0.11, 5.47]	0.97 [0.72, 1.28]
15	GMI	0.65 [0.32, 1.33]	0.88 [0.72, 1.06]
16	GMI	0.52 [0.28, 0.99]	0.83 [0.71, 0.98]
18	PF	1.29 [0.54, 3.51]	0.93 [0.75, 1.12]
20	MI+PF	1.34 [0.63, 3.31]	1.01 [0.89, 1.17]
21	MI+PF	0.38 [0.12, 1.12]	0.91 [0.77, 1.10]
	PF	0.99 [0.32, 3.36]	1.06 [0.87, 1.27]
22	MI+PF	0.73 [0.35, 1.57]	1.00 [0.86, 1.16]
Overall		0.79 [0.58, 1.04]	0.96 [0.91, 1.00]



- MI = Individual Motivational Interview, PF = Standalone Personalized Feedback, MI + PF = MI with Personalized Feedback, GMI = Group Motivational Interview

Forest Plot for Alcohol Problems (Gaussian)

Study	Intervention	B [95% CI]
2	PF	0.04 [-0.12, 0.20]
7.1	GMI	0.02 [-0.21, 0.26]
7.2	GMI	-0.10 [-0.22, 0.03]
8a	PF	-0.07 [-0.14, 0.01]
8b	PF	-0.02 [-0.08, 0.05]
8c	PF	0.05 [-0.10, 0.18]
9	MI+PF	-0.02 [-0.17, 0.14]
	PF	0.05 [-0.10, 0.21]
	GMI	0.09 [-0.06, 0.24]
10	MI+PF	-0.24 [-0.36, -0.09]
11	PF	-0.05 [-0.18, 0.06]
12	MI+PF	-0.13 [-0.27, 0.02]
13/14	MI+PF	0.01 [-0.20, 0.22]
	PF	0.04 [-0.20, 0.29]
15	GMI	-0.06 [-0.19, 0.07]
16	GMI	0.00 [-0.11, 0.11]
18	PF	0.14 [-0.01, 0.29]
20	MI+PF	0.08 [-0.03, 0.19]
21	MI+PF	-0.07 [-0.22, 0.08]
	PF	-0.03 [-0.18, 0.14]
22	MI+PF	-0.06 [-0.18, 0.06]
Overall		-0.02 [-0.05, 0.02]



- MI = Individual Motivational Interview, PF = Standalone Personalized Feedback, MI + PF = MI with Personalized Feedback, GMI = Group Motivational Interview

Discussion

- Wide variation of intervention effects on alcohol outcomes is generally consistent with results from meta-analyses based on summary statistics.
  - When alcohol outcomes are modeled in a distribution-appropriate analysis, intervention effects in most studies are non-significant.
  - Across studies, there are small, statistically non-significant reductions in alcohol consumption and negative consequences.

- Bayesian MLM using study by randomization group as the highest level of the model was a practical approach to combining studies with varying numbers of treatment arms.
  - Avoids the need to collapse intervention conditions or discard data.

Discussion (cont.)

- Allowed the calculation of effect sizes for:
  - Individual intervention groups
  - Across all interventions
  - For specific intervention types (not shown)

- Weighting of the intervention estimates was handled within the multilevel model.
  - The IPD is weighted within the likelihood distribution.
  - The precision of the estimates is proportional to the amount of contributing data.

- The detailed approach is generalizable to outcomes beyond alcohol use.

Analysis of non-normal outcomes not trivial…

- Bayesian MCMC estimation required a good deal of computing time, especially for the non-Gaussian model.
  - Gaussian model of alcohol problems: **< 1 hour**
  - Hurdle model of drinks per week: **36 hours**

Next steps…

- Conduct a simulation study comparing results of the Bayesian MLM approach used in the present study with summary-statistic based meta-analysis.
  - How biased are estimates using summary statistic based methods that assume normal distribution?

Questions?

- For post-conference questions, contact:
  - David Huh (dhuh@uw.edu).

References

- Atkins, D. C., Baldwin, S. A., Zheng, C., Gallop, R. J., & Neighbors, C. (2013). A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of Addictive Behaviors, 27,* 166–177. doi:10.1037/a0029508
- Carey, K. B., Scott-Sheldon, L. A. J., Carey, M. P., & DeMartini, K. S. (2007). Individual-level interventions to reduce college student drinking: A meta-analytic review. *Addictive Behaviors, 32,* 2469–2494. doi:10.1016/j.addbeh.2007.05.004
- Collins, R. L., Parks, G. A., & Marlatt, G. A. (1985). Social determinants of alcohol consumption: the effects of social interaction and model status on the self-administration of alcohol. *Journal of Consulting and Clinical Psychology, 53,* 189–200.
- Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods, 14,* 165–176. doi:10.1037/a0015565
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University.
- Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 357–376). New York: Russell Sage Foundation.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software, 33,* 1–22.
- Hopewell, S., Dutton, S., Yu, L.-M., Chan, A.-W., & Altman, D. G. (2010). The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* (Clinical Research Ed.), 340, c723.
- Huh, D., Kaysen, D., & Atkins, D. C. (2014). *Modeling cyclical patterns in daily college drinking data with many zeroes.* Manuscript submitted for publication.
- Huh, D., Mun, E.-Y., Larimer, M. E., White, H. R., Ray, A. E., Rhew, I. C., . . . Atkins, D. C. (2014). *Brief motivational interventions for college student drinking may not be as powerful as we think: An individual participant-level data meta-analysis.* Manuscript submitted for publication.
- Mun, E.-Y., de la Torre, J., Atkins, D. C., White, H. R., Ray, A. E., Kim, S.-Y., … the Project INTEGRATE Team. (2014). *Project INTEGRATE: An integrative study of brief alcohol intervention trials for college students.* Manuscript under review.
- National Institute on Alcohol Abuse and Alcoholism (2002) *A call to action: Changing the culture of drinking at U.S. colleges* (No. 02-5010). Department of Health and Human Services, National Institutes of Health, Rockville, MD: USA.
- R Core Team (2014) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.