A re-evaluation of fixed-effect(s) meta-analysis

Ken Rice, Thomas Lumley & Julian Higgins

with extensions by Clara Dominguez, Kendrick Li,
Spencer Hansen & Albert Osom

June 24, 2025

What's your talk about? (briefly)

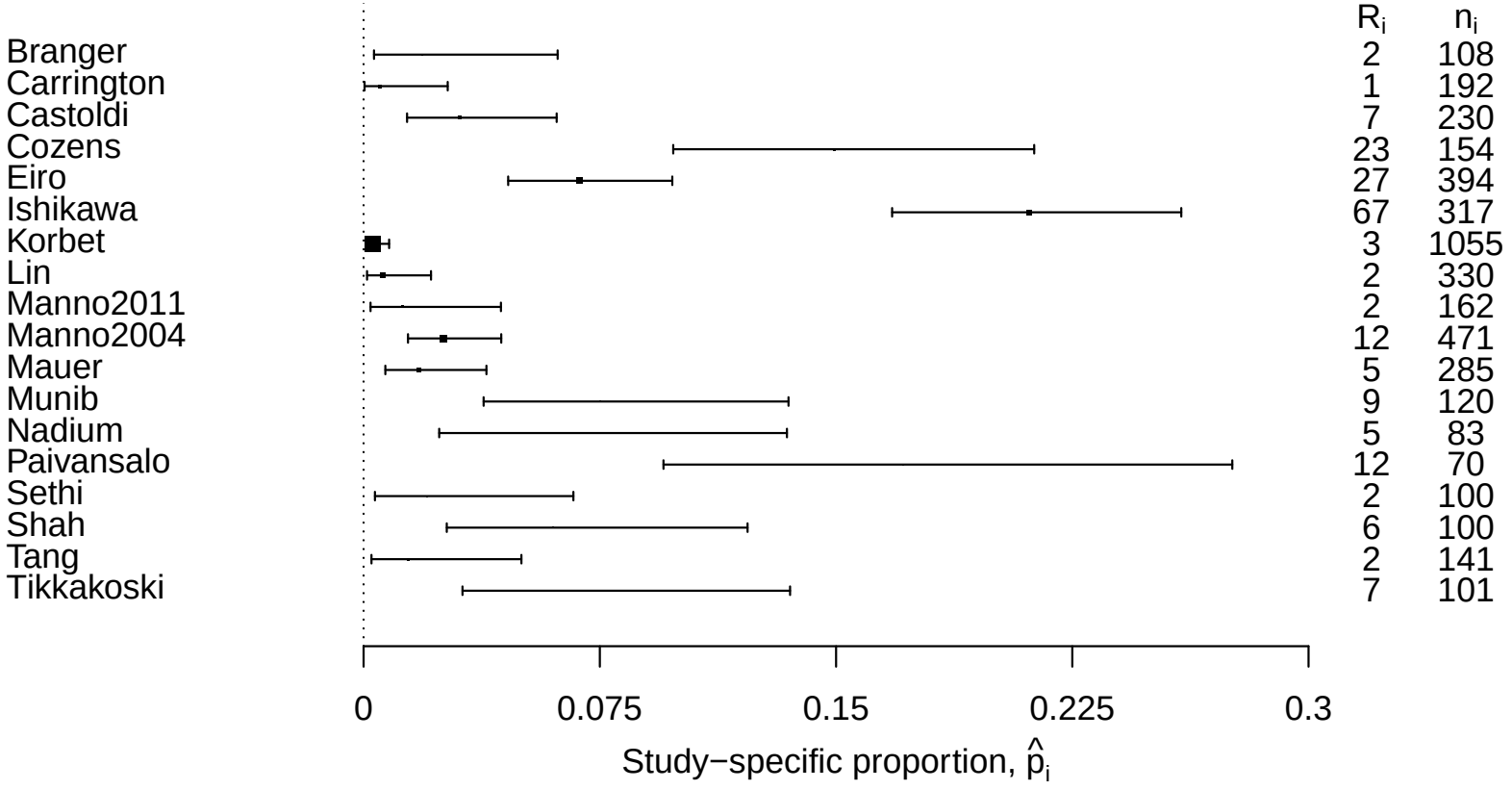
- Fixed-effect**S** meta-analysis, and how it differs from common effect/fixed effec**T**
- What are we estimating?
- Why might that be a good/bad idea?

http://tinyurl.com/fixef

has these slides and more.

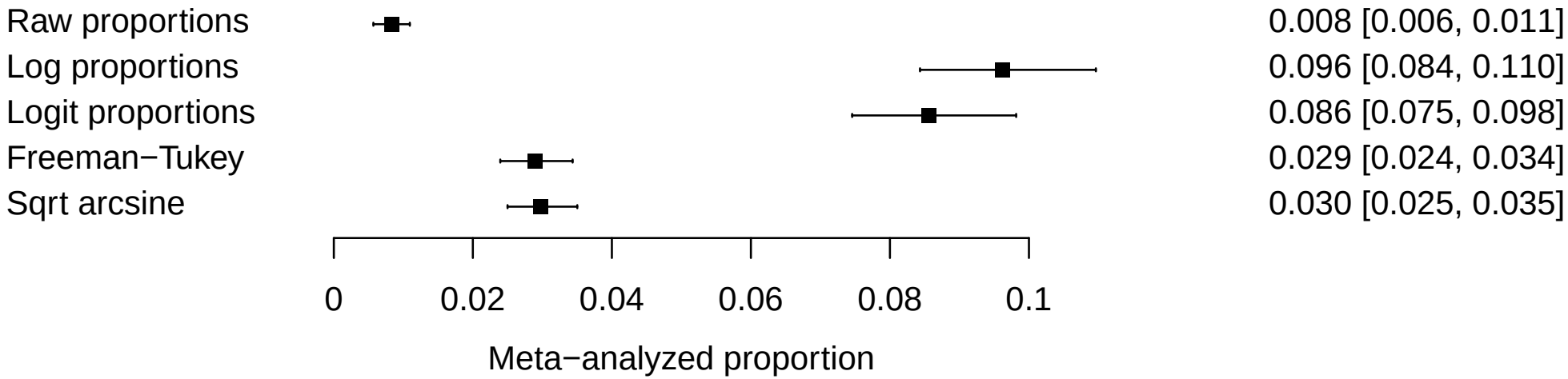
Is this going to hurt?

Kidney Precision Medicine Project (KPMP)'s meta-analysis of reported pain after native kidney biopsies: (k=18 studies... yes, some heterogeneity is obvious)



Is this going to hurt?

Embarrassing Q: How would these $\widehat{p}_i$ be meta-analyzed?



- Asking the vague Q (above) leads to a mess!
- Better Q: *What is the overall rate of pain, in people like those who contributed to these k studies?*

Is answering that Q also a pain? (no!)

Estimate for Better Q	Target parameter – overall rate
$\widehat{p}_F = \frac{\sum_{i=1}^k R_i}{\sum_{j=1}^k n_j} = \sum_{i=1}^k \frac{n_i}{\sum_{j=1}^k n_j} \widehat{p}_i$	$p_F = \sum_{i=1}^k \frac{n_i}{\sum_{j=1}^k n_j} p_i$

- Conditioning on n_i (in paper we use parameter η_i)
- There are k fixed, unknown rates p_i. This *fixed effect***S** *meta-analysis* estimates an average of them, **without assuming they are all identical**
- Through $\widehat{p}_F$ we learn about populations of the studies we have (See Hedges *et al* work, cited in paper)
- No, not generalizing to (predicting) the next study that comes along
- But why *this* overall average? Well...

What would you estimate if you **could** pool the data?

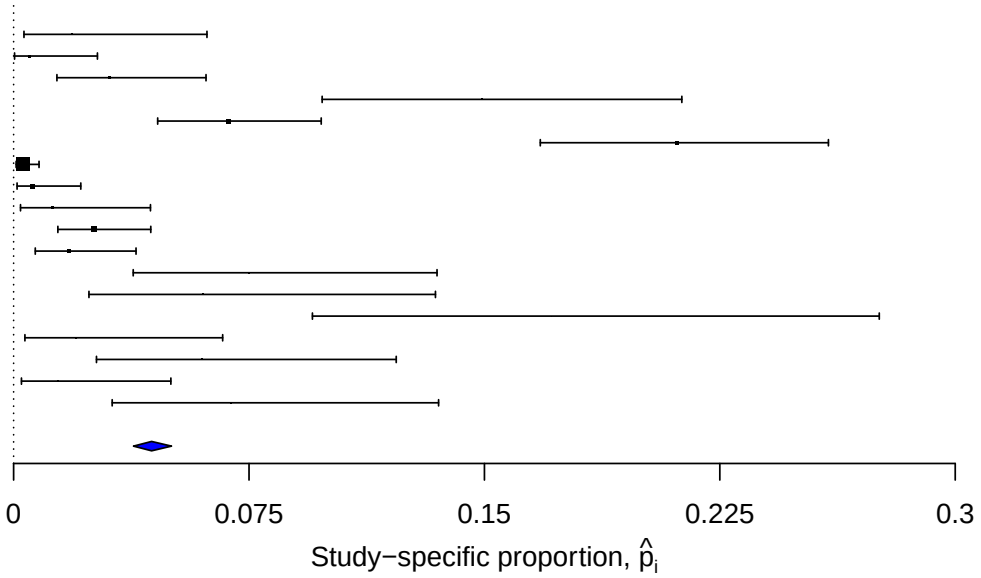
Is answering that Q also a pain? (no!)

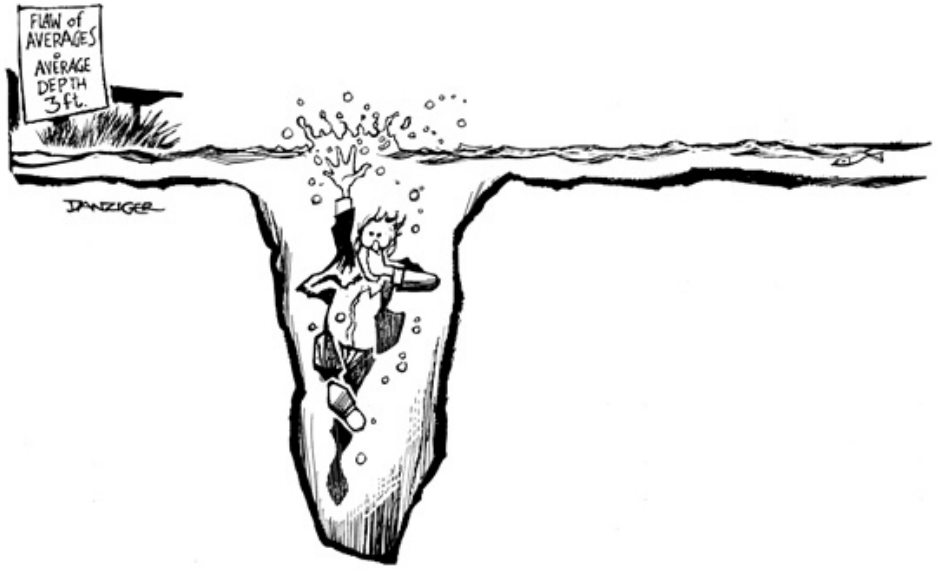
Not a focus today, but:

- Exact frequentist inference on p_F is easily available (Clopper-Pearson/Blaker intervals) **and has been for decades!** (Hoeffding 1956)
- No continuity corrections, adding 1/2, arcsine transformations etc
- The inference can be slightly conservative – though estimates of *how* conservative are available

But what about heterogeneity?

Considering **only** $\widehat{p}_F$=0.043 (0.038, 0.05) we'd fall into the *flaw of averages*:





- Overall p_F **still meaningful**, inference on $\widehat{p}_F$ **still valid**, but heterogeneity is also important.
  (And this is okay! One dataset can address two questions!)

But what about heterogeneity?

To summarize spread, suggest estimating

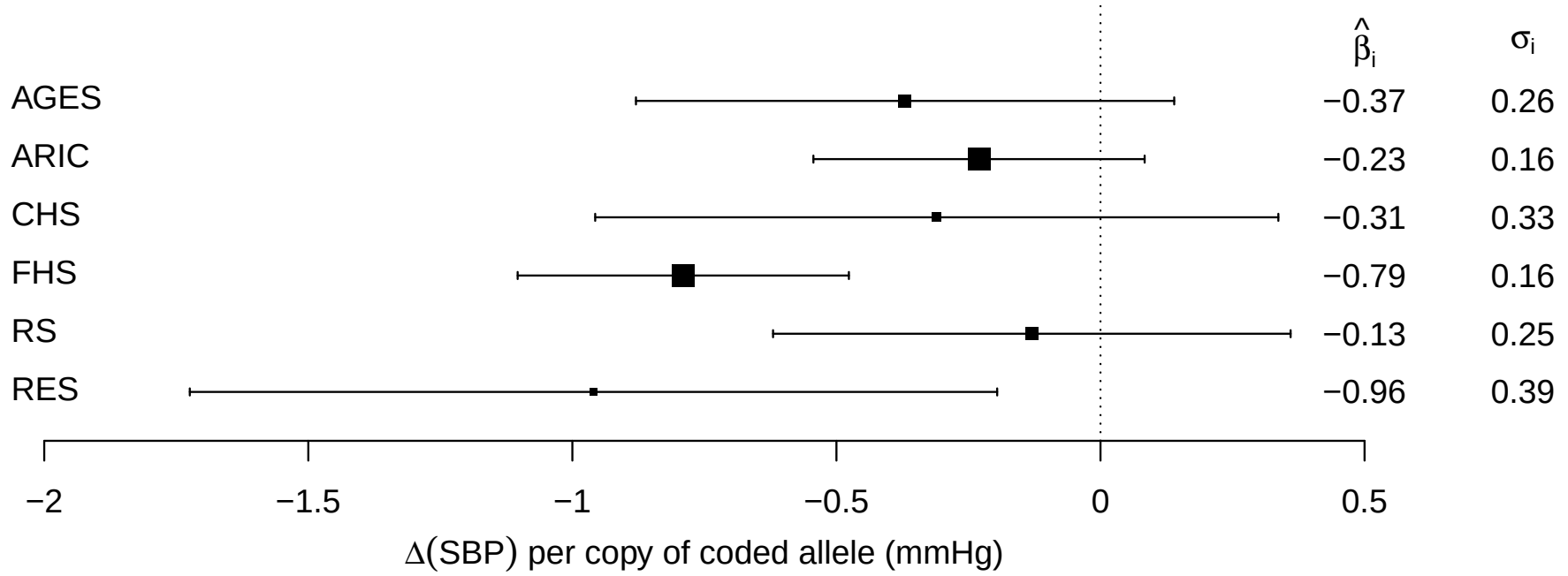
$$\delta^2 = \sum_{i=1}^{k} \frac{n_i}{\sum_{j=1}^{k} n_j} (p_i - p_F)^2$$

... a weighted variance of the p_i's; it's also the excess variance in $\widehat{p}_F$ compared to what we'd get under homogeneity, i.e. all $p_i = p_F$

- Weighted SD δ is **on the proportion scale**: estimated here to be 0.06 – which matters, if overall rate $p_F \approx 0.04$
- *Estimate that, sampling pairs of individuals from overall population, we'd expect their study-specific rates to differ by* ≈ 0.06 *(6 percentage points)*

Wasn't your paper about regression estimates?

Same ideas in classic meta-analysis setup, from a blood pressure GWAS:

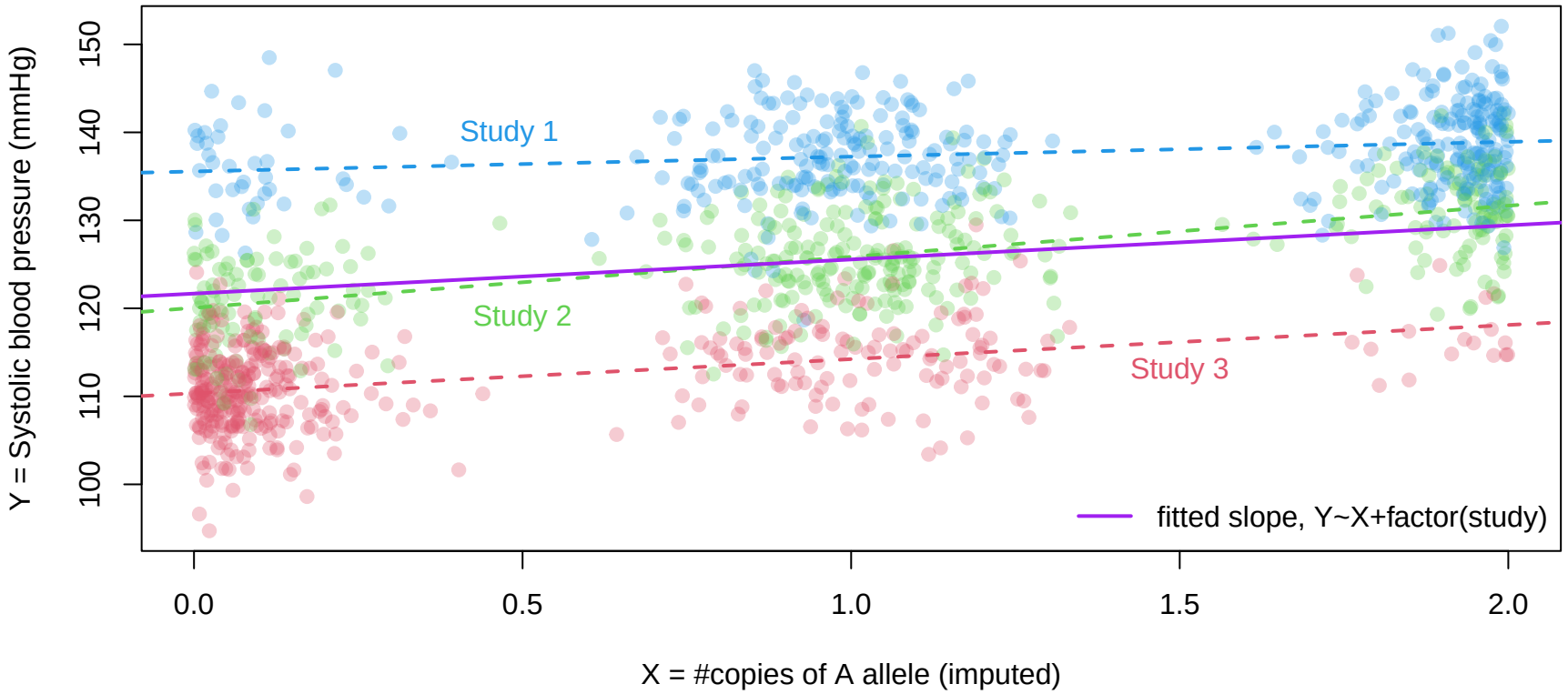


Linear regression estimates from k=6 large cohorts: CLT means $\widehat{\beta}_i \sim N(\beta_i, \sigma_i^2)$, and standard errors σ_i known with negligible error.

What question should I answer with them?

What would you estimate if you **could** pool the data?

Arguably, regress Y on X *and adjust for study* (here for $k = 3$):



How do we get that from a meta-analysis?

We get *essentially* this from *fixed-effects meta-analysis*: (conditioning on n_i)

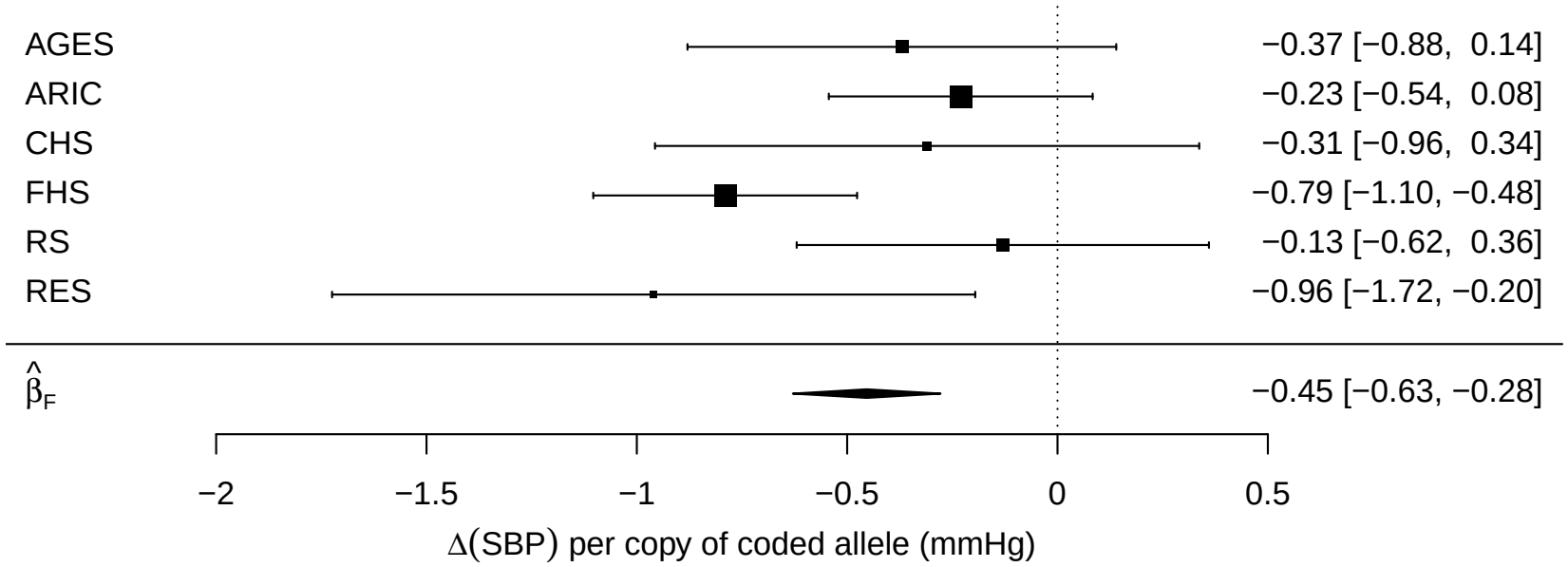
Estimate	Target parameter
$\sum_{i=1}^{k} \frac{\sigma_i^{-2}}{\sum_{j=1}^{j} \sigma_j^{-2}} \widehat{\beta}_i$	$\sum_{i=1}^{k} \frac{n_i \phi_i}{\sum_{j=1}^{j} n_j \phi_j} \beta_i$

...where ϕ_i is Fisher information about β_i per observation from study population i and standard errors σ_i ($\propto 1/\sqrt{n_i \phi_i}$) are known with negligible error.

- Exact confidence intervals straightforward (standard FE formula)
- Under homogeneity, estimates **same parameter** as pooled study-adjusted analysis, with no efficiency loss
- Under heterogeneity, parameter is same as pooled study-adjusted analysis, weighted for each study's precision; still useful, efficiency loss minimal

Our 2017 paper argues that it's a sane default, as an **overall summary** of what happens in people like those in the k studies.

How do we get that from a meta-analysis?

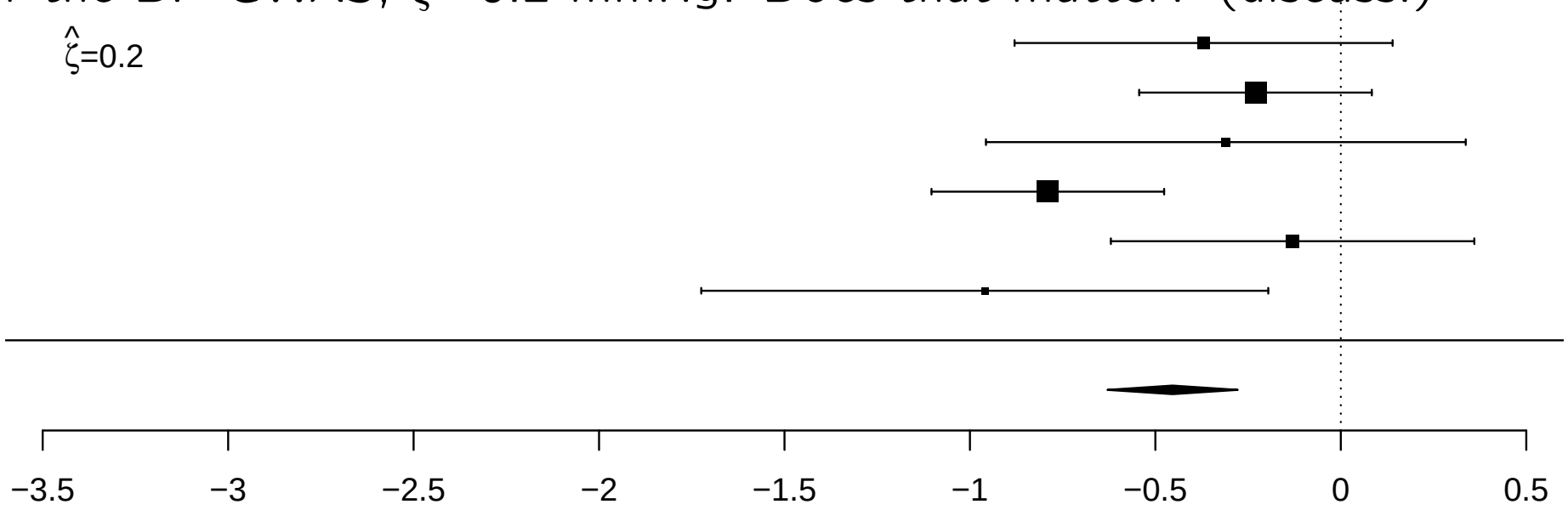


- *Combining the 6 populations* (weighting $\propto$ precision) *on average each extra A allele is associated with 0.45mmHg (0.28, 0.63) lower blood pressure*
- In this example, focus is testing the *strong null*, that all β_i=0. Wald tests of $\widehat{\beta}_F = 0$ do this efficiently
- Heterogeneity still worth addressing: how much do the effects differ? Are there patterns by study?

So what should we do about heterogeneity?

As before we suggest a weighted variance; $\zeta^2 = \sum_{i=1}^{k} \frac{n_i \phi_i}{\sum_{j=1}^{k} n_j \phi_j} (\beta_i - \beta_F)^2$.

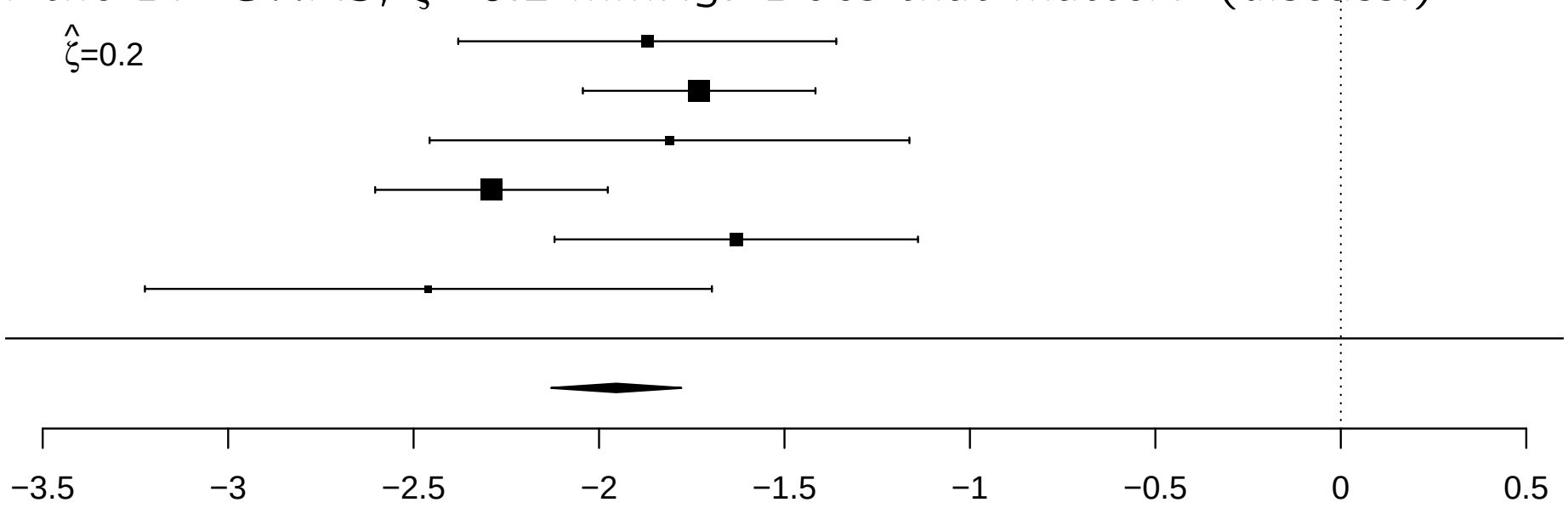
- Empirical estimate just a rescaled Cochran's Q; easy connection to tests
- $\hat{\zeta}^2$ much stabler than RE versions (e.g. DerSimonian-Laird, REML $\hat{\tau}^2$)
- For the BP GWAS, $\hat{\zeta}$=0.2 mmHg. Does that matter? (discuss!)



So what should we do about heterogeneity?

As before we suggest a weighted variance; $\zeta^2 = \sum_{i=1}^{k} \frac{n_i \phi_i}{\sum_{j=1}^{k} n_j \phi_j} (\beta_i - \beta_F)^2$.

- Empirical estimate just a rescaled Cochran's Q; easy connection to tests
- $\hat{\zeta}^2$ much stabler than RE versions (e.g. DerSimonian-Laird, REML $\hat{\tau}^2$)
- For the BP GWAS, $\hat{\zeta}$=0.2 mmHg. Does that matter? (discuss!)



So what else should we do about heterogeneity?

Fixed-effect**S** ideas extend to meta-regression on study-specific x_i:

$$\hat{\beta}_{MR} = \sum_{i=1}^{k} \frac{\sigma_i^{-2}(x_i - x_F)^2}{\sum_j \sigma_j^{-2}(x_j - x_F)^2} \left(\frac{\hat{\beta}_i - \hat{\beta}_F}{x_i - x_F} \right), \text{ where } x_F = \sum_{i=1}^{k} \frac{\sigma_i^{-2}}{\sum_j \sigma_j^{-2}}$$

$$Q_{MR} = \sum_{i=1}^{k} \sigma_i^{-2}(\hat{\beta}_i - \hat{\beta}_F - \hat{\beta}_{MR}(x_i - x_F))^2$$

Simple ANOVA/ANODev-style decompositions, for testing the β_{MR} **trend**:

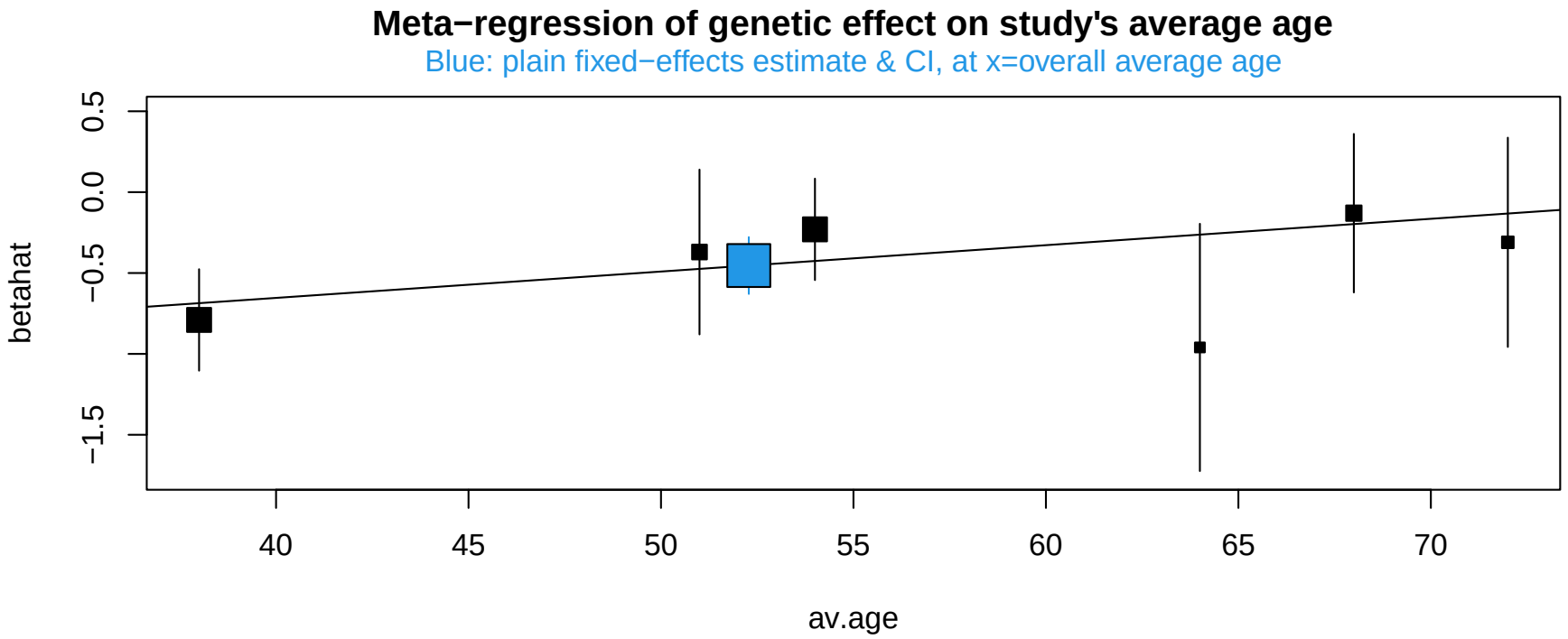
$$\sum_{i=1}^{k} \hat{\beta}_i^2/\sigma_i^2 \equiv \sum_{i=1}^{k} Z_i^2 = Z_F^2 + Q$$

$$= Z_F^2 + Z_{MR}^2 + Q_{MR}$$

where Z_F^2, Z_{MR}^2 are Wald statistics testing $\beta_F = 0$ and $\beta_{MR} = 0$

So what else should we do about heterogeneity?

Regressing the genetic BP effect on x=study's average age:

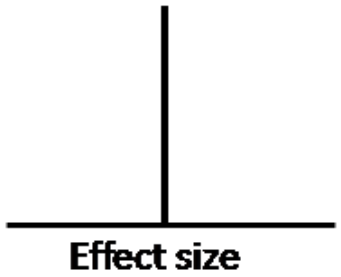
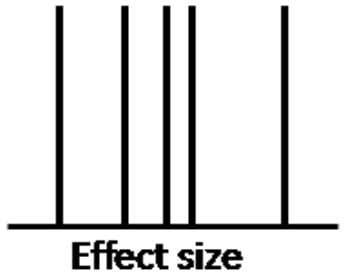


Averaging over all studies, we estimate effect on BP is higher by 0.016 mmHg per allele (0.001, 0.032, p=0.036) per 1 year difference in average age.

Why not use random effects (RE) meta-analysis?

This comes up a lot!

But...

Name:	Common effect	Fixed effect**S**	Random effects
Assumptions:			
	All $\beta_i = \beta_0$	β_i unrestricted	$\beta_i \sim (\mu, \tau^2)$

- Unlike RE models, the β_i are typically **not** randomly sampled
  ... though for prediction, few better options are available
- Appealing to exchangeability motivates *fitting* an RE model (De Finetti) but **does not** motivate estimating its μ
- "Because it gives wider intervals"
- Cochrane agrees: *"...a random-effects model does not 'take account' of the heterogeneity, in the sense that it is no longer an issue."*

What did you miss out?

Other work on the same theme:

- Bayesian approaches to fixed-effect**S** meta-analysis, that are careful to decouple exchangeability from target parameters
- Relaxing the assumption that σ_i have negligible error (see also Hoaglin 2015)
- Addressing small-sample issues when combining 2×2 tables, via existing methods, or a new exact method
- Shrinkage/clustering to summarize heterogeneity (WIP)
- Decision theory: for what questions are FE/RE analyses the optimal answers? (WIP)

Are you going to stop now?

Summary:

- FE methods have **simple, useful** interpretations, even under heterogeneity
- What would you do if you could combine the data?
- What question **do** you want to answer?

Many thanks to:

- Thomas & Julian, and Clara, Kendrick, Spencer and Albert
- RSS reviewers, discussants
- Everyone who attended!

http://tinyurl.com/fixef
has these slides and more.