Multiplicity considerations for analyses of non-exchangeable subgroups

Patrick Schnell

April 12, 2022

Many goals of subgroup analyses

- Estimate average treatment effects within subgroups
- Estimate consistency or heterogeneity among subgroups
- Identify population with exceptional effects
- **Identify population that benefits**
- Etc.

Benefiting subgroup identification

- Personalized treatment effect (PTE) $\Delta(\boldsymbol{x})$ depending on covariate vector $\boldsymbol{x}$, more positive is better
  - Could be CATE ($\mathsf{E}[Y^1 - Y^0|\boldsymbol{x}]$), hazard ratio, log odds ratio, etc.
- Benefiting subgroup $\mathrm{B} = \{\boldsymbol{x} : \Delta(\boldsymbol{x}) > \delta\}$
- Want some estimate $\widehat{\mathrm{B}}$ with nice properties
- Types of patients only identified as far as $\boldsymbol{x}$ allows, i.e., no statements about individual units, e.g., $\mathrm{P}[Y^1 > Y^0 \mid \boldsymbol{x} \in \mathrm{B}]$

Non-exchangeable subgroups

- Each subpopulation indexed by a covariate point x may be considered a subgroup
  - **Benefit:** subgroups, thresholds, etc. need not be prespecified, just variables
  - **Challenge:** many subgroups! Potentially inconvenient shape of identified benefiting subgroup
- PTE correlated between similar x, x' a priori (Bayesian) and as estimates
  - **Benefit:** accounting for positive correlation mitigates power loss from multiplicity adjustments
  - **Challenge:** difficult to get analytic results about joint distributions for multiplicity adjustments

Reasons for multiplicity control

- Need based on decisions being made
- **Regulatory:** limit probability of marketing to populations not benefiting from treatment (FWER)
  - $\mathrm{P}[\widehat{\mathrm{B}} \not\subseteq \mathrm{B}] < \alpha$, *not* $\mathrm{P}\left[\int_{\widehat{\mathrm{B}}} \Delta(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \leq \delta\right] < \alpha$
- **Cost-benefit thresholds:** limit proportion of identified subgroup not benefiting from treatment (FDR)
  - Proportion of the $\boldsymbol{x} \in \widehat{\mathrm{B}}$ weighted by population distribution not satisfying $\Delta(\boldsymbol{x}) > \delta$

FWER control

- **FDA on multiple endpoints (draft guidance):** FDA's concern for controlling the Type I error probability is to minimize the chances of a false favorable conclusion for any of the primary or secondary endpoints, regardless of which and how many endpoints in the study have no effect.
- **Analogy to subgroups:** Our concern for controlling the Type I error probability is to minimize the chances of a false favorable conclusion for any subgroups (here, x), regardless of which and how many subgroups in the study have no effect.

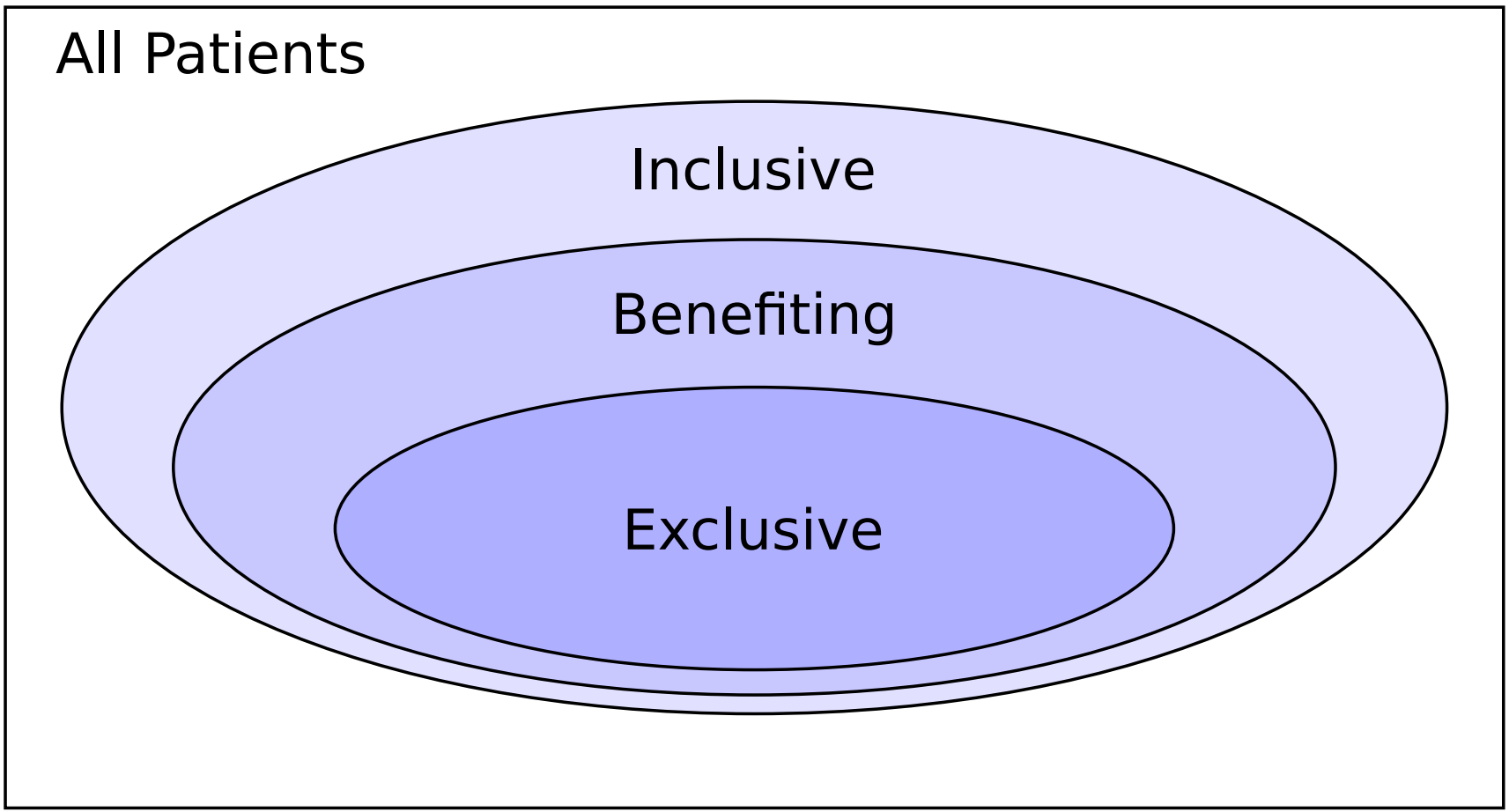
FWER control (continued)

- Estimator $\widehat{\mathrm{B}}$ will collect $\boldsymbol{x}$ for which we are confident that $\Delta(\boldsymbol{x}) > \delta$
- Controlling FWER means limiting probability that any $\boldsymbol{x} \in \widehat{\mathrm{B}}$ fails $\Delta(\boldsymbol{x}) > \delta$
- Result: FWER is probability that $\widehat{\mathrm{B}}$ is not a subset of B.

Confidence/credible subgroups

- **Benefiting subgroup:** types of patients who benefit from treatment
- **Exclusive confidence subgroup:** should contain *only* types of patients who benefit
- **Inclusive confidence subgroup:** should contain *all* types of patients who benefit

Visualization



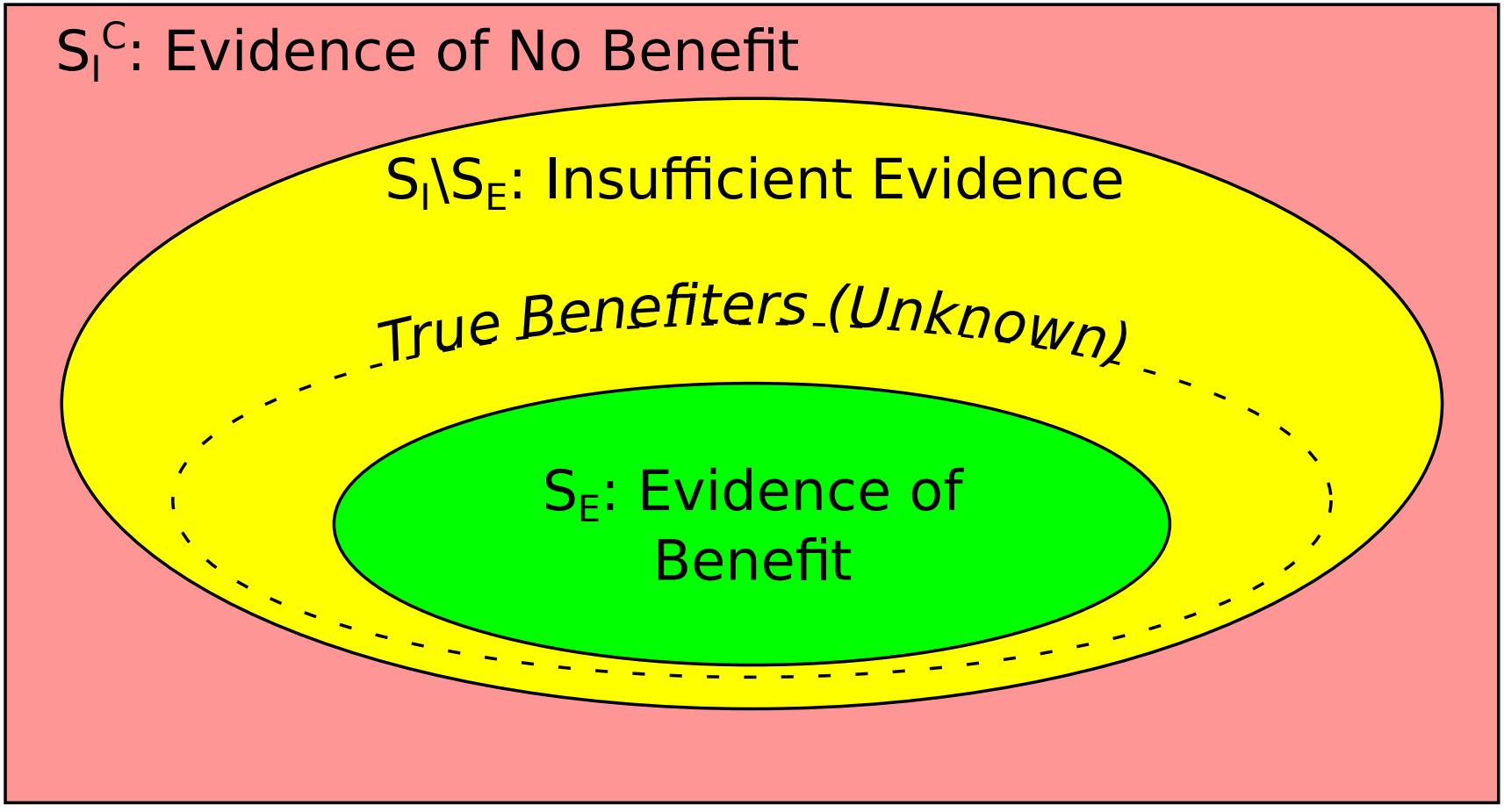
Frequentist definition

Definition (Confidence subgroups)

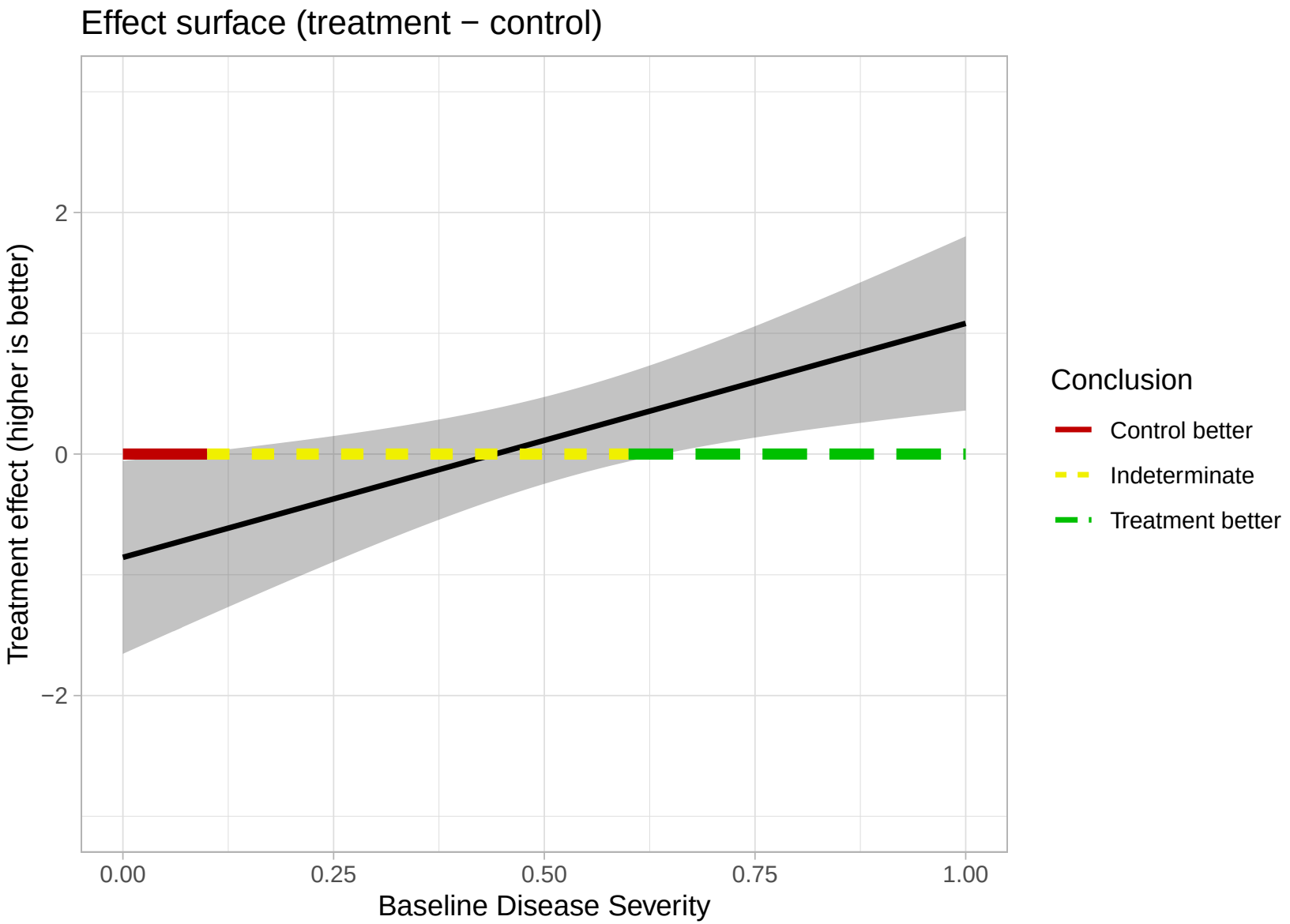
If $S_B(\boldsymbol{\theta})$ is the benefiting subgroup, then α-level inclusive confidence subgroups $S_I(\boldsymbol{Y})$ and exclusive confidence subgroups $S_E(\boldsymbol{Y})$ are subsets of the population such that under repeated sampling:

$$P_{\boldsymbol{Y}}\left[S_E(\boldsymbol{Y}) \subseteq S_B(\boldsymbol{\theta}) \subseteq S_I(\boldsymbol{Y})\right] \geq 1 - \alpha$$

Interpretation



Inverting simultaneous interval estimates



Simultaneous interval estimates

- M-draw Monte Carlo estimate of sampling distribution of estimated effect surface $\widehat{\Delta}(x)$ approximately normal
- Restrict covariate space to C
- Find the $1-\alpha$ restricted-space confidence band:

$$\widehat{\Delta}(x) \pm W^*_{\alpha,\mathrm{C}} \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left\{\widehat{\Delta}^{(m)}(x) - \widehat{\Delta}(x)\right\}^2}$$

where $W^*_{\alpha,C}$ is the $1-\alpha$ quantile of the distribution of

$$W_C^{(m)} = \sup_{x\in\mathrm{C}} \frac{\left|\widehat{\Delta}^{(m)}(x) - \widehat{\Delta}(x)\right|}{\sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left\{\widehat{\Delta}^{(m)}(x) - \widehat{\Delta}(x)\right\}^2}}$$

Weighted FDR/PPV control

- Measure μ on covariate space
  - E.g., population covariate distribution
  - Could be estimated by empirical distribution
- FDR control:

$$\mathsf{E}\left[\frac{\mu(\widehat{\mathrm{B}} \setminus \mathrm{B})}{\mu(\widehat{\mathrm{B}})}\right] < \alpha$$

- Positive predictive value analogue: $1 - \mathsf{E}\left[\frac{\mu(\widehat{\mathrm{B}} \setminus \mathrm{B})}{\mu(\widehat{\mathrm{B}})}\right]$ is PPV of "diagnostic test" assigning patients to benefiting subgroup

Estimated FDR for inclusion in $\widehat{\mathrm{B}}$

- Yekutieli & Benjamini, "Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics." *JSPI* 82(1-2): 171–196, 1999.
- Bootstrap

$$Z^{(k)}(x_i) = \left\{\widehat{\Delta}^{(k)}(x_i) - \widehat{\Delta}(x_i)\right\} \Big/ \sqrt{\sum_{k=1}^{K} \left\{\widehat{\Delta}^{(k)}(x_i) - \widehat{\Delta}(x_i)\right\}^2}$$

- Unadjusted bootstrap p-values from Z scores
- For fixed p estimate

$$Q_{\widehat{\mathrm{B}}}(p) = \frac{\mu(\widehat{\mathrm{B}}(p) \setminus \mathrm{B})}{\mu(\widehat{\mathrm{B}}(p) \setminus \mathrm{B}) + \mu(\widehat{\mathrm{B}}(p) \cap \mathrm{B})}, \tag{1}$$

the weighted FDR of the testing procedure which rejects the null hypotheses with p-values less than p

Simulation

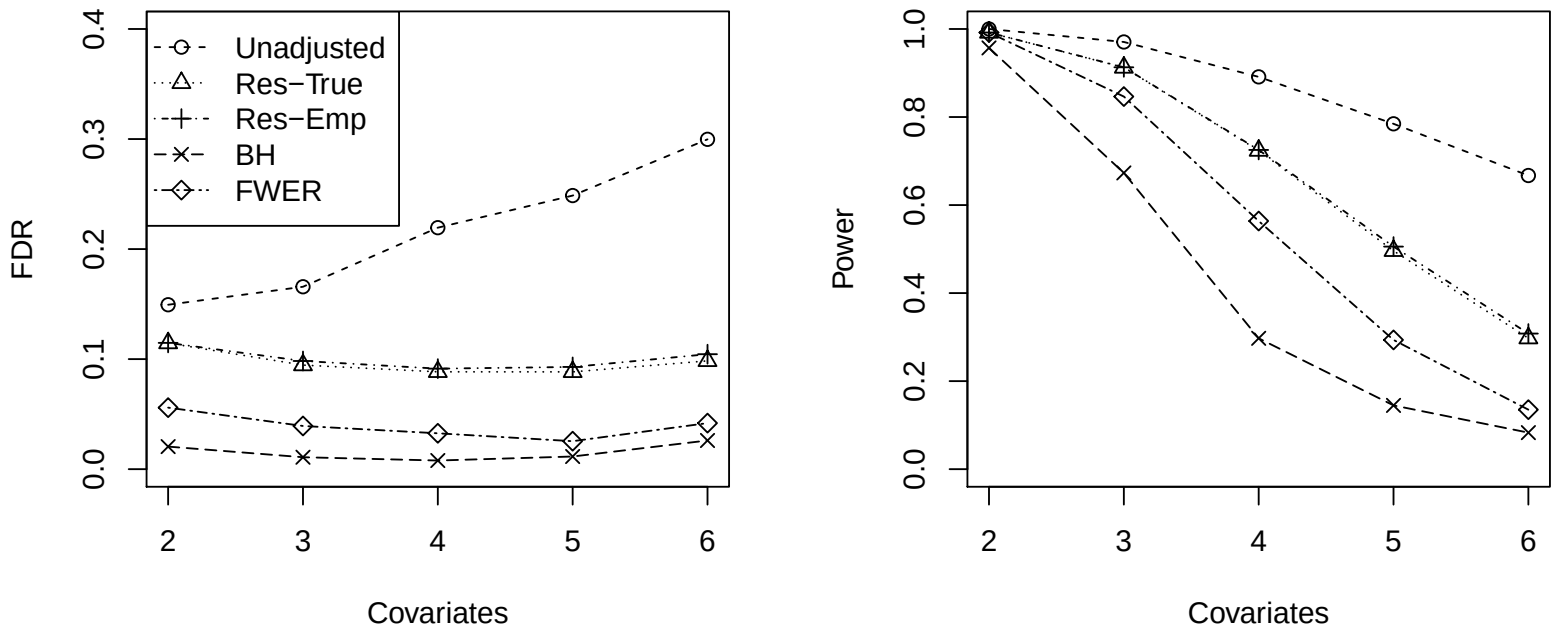


Figure: N = 300, varying number of uniform binary covariates, effect is 1 iff first two covariates are 1. Target FDR is 10%. Res-True: resampling with true weights; Res-Emp: resampling with empirical weights; BH: Benjamini-Hochberg FDR procedure; FWER: FWER-controlling confidence subgroups.

Simulation

Generating mechanism	Characteristic	Unadjusted	Res-True	Res-Emp	BH	FWER
Null effect	FWER	0.441	0.104	0.107	0.010	0.086
	FDR	0.441	0.104	0.107	0.010	0.086
	Power	—	—	—	—	—
Sigmoidal effect	FWER	0.400	0.346	0.349	0.019	0.108
	FDR	0.031	0.027	0.028	0.000	0.007
	Power	0.855	0.837	0.839	0.567	0.730
Quadratic effect	FWER	0.168	0.092	0.101	0.000	0.005
	FDR	0.009	0.005	0.028	0.000	0.000
	Power	0.570	0.404	0.839	0.083	0.210

Table: $N = 300$; thin plate spline fits (`mgcv`); target 10% for respective error rates

- Resampling procedures conservative when effect not uniformly null
- Asymptotically, points with $\Delta(\boldsymbol{x}) > \delta$ are correctly identified almost surely
- Res-True very conservative for quadr. case (investigating)

Example

- Data from sequence of four clinical trials of Alzheimer's disease treatments with same placebo control and SoC active control
- Endpoint: change in ADAS-Cog disease severity over 12 weeks
- Covariates: sex, APoE4 carrier status, baseline disease severity, rate of decline from diagnosis to baseline

Example FWER

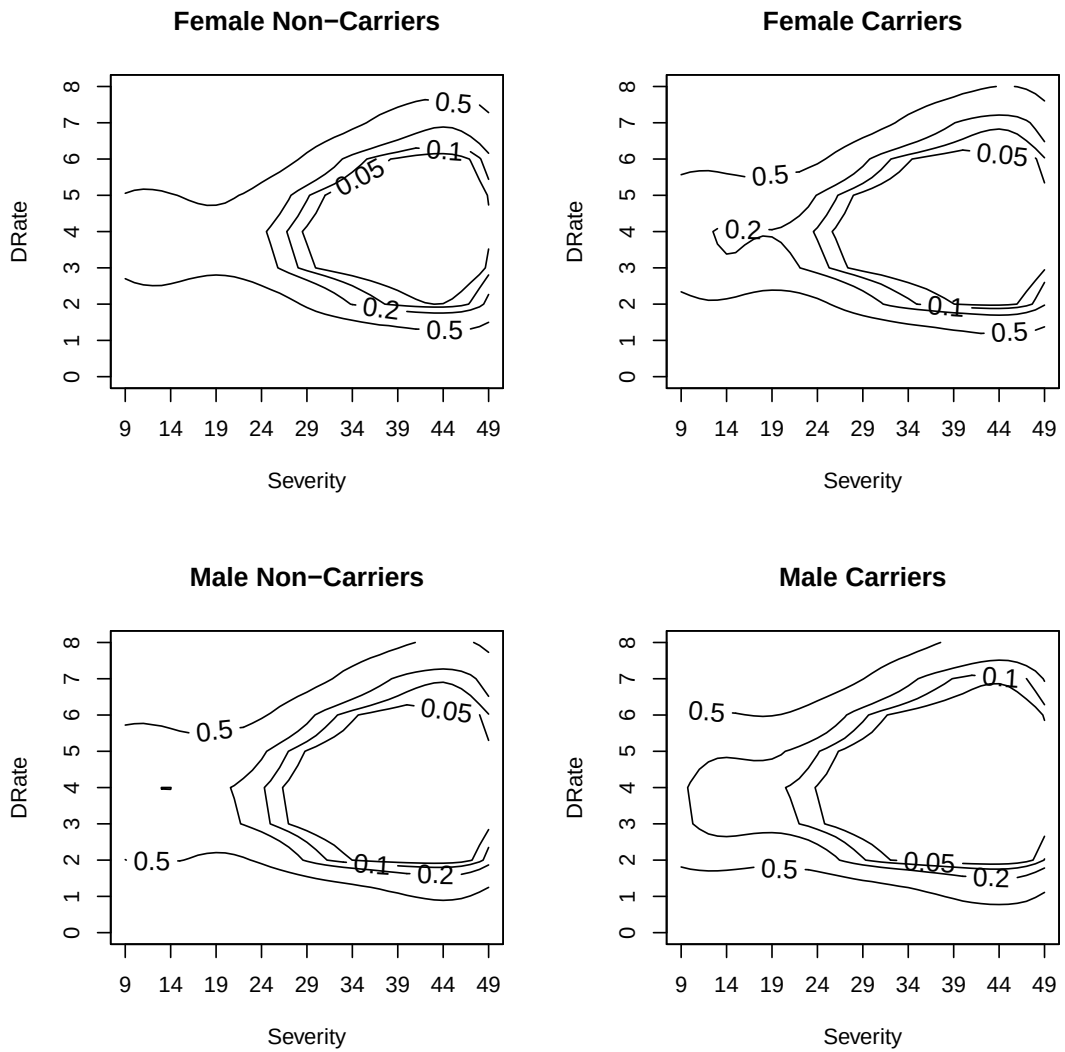


Figure: FWER contours. The exclusive confidence subgroup controlling FWER at or below 5% is the region within the 0.05 contour, etc.

Example FDR

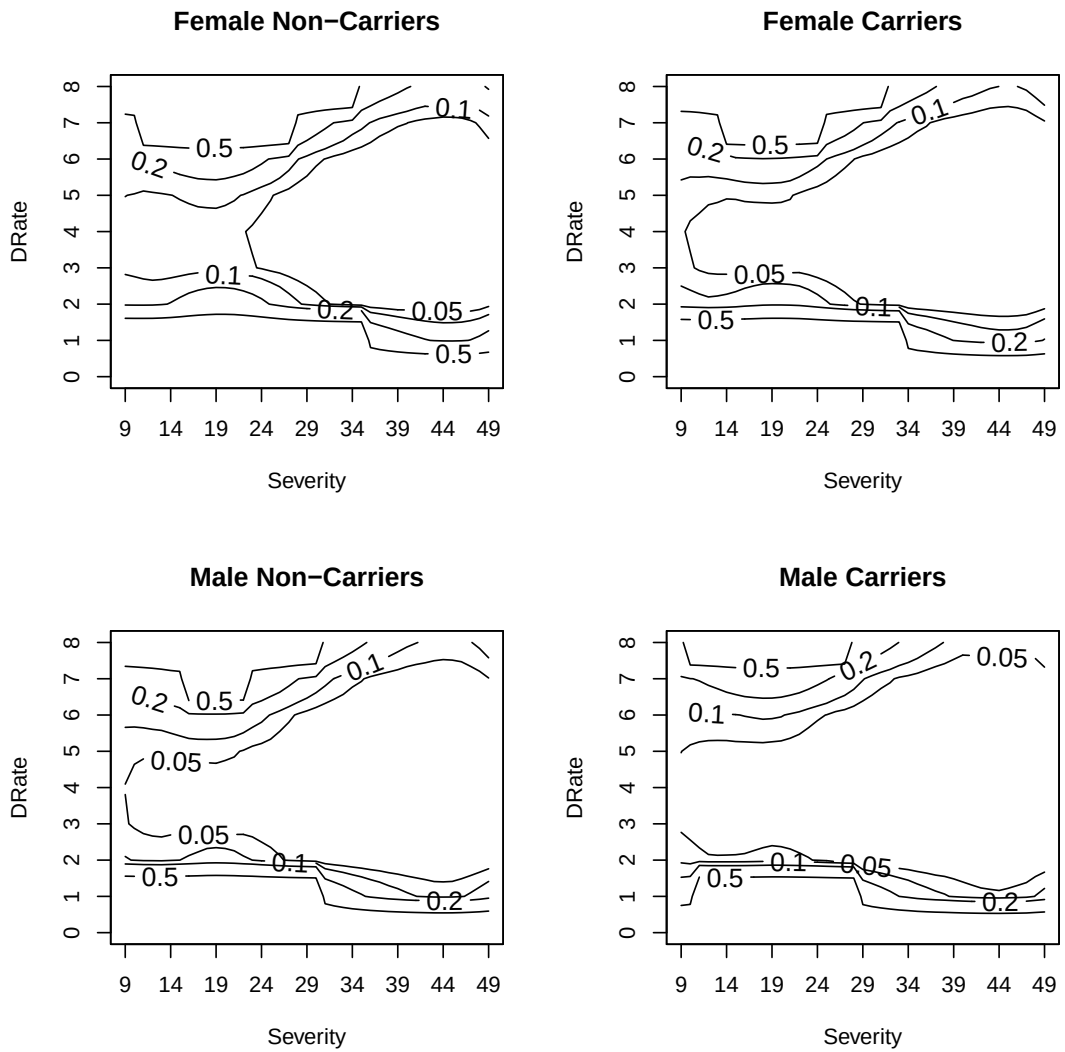


Figure: FDR contours. The exclusive confidence subgroup controlling FDR at or below 5% is the region within the 0.05 contour, etc.

Conclusion

- Benefiting subgroups can be built up from atomic subgroups (e.g., covariate points)
- Multiplicity control goals have useful intuitive interpretations in terms of benefiting subgroups (*FWER*: estimator contained in benefiting subgroup, 1 – *FDR*: PPV of assignment)
- Bayesian or resampling methods allow flexible model choice decoupled from multiplicity handling
- Building subgroups from non-exchangeable atoms benefits from correlation between similar covariate points

Selected references

1. Schnell. "Monte Carlo approaches to frequentist multiplicity-adjusted benefiting subgroup identification." *Stat Methods Med Res* 2020.
2. Schnell, Tang, Müller, Carlin. "Semiparametric benefiting subgroup identification via credible subgroups." *Clin Trials* 2018.
3. FDR-controlling method to be submitted to *Clin Trials* special issue.
4. `credsubs` R package on CRAN (FWER only for now)

email: `schnell.31@osu.edu`