Centrum Wiskunde & Informatica

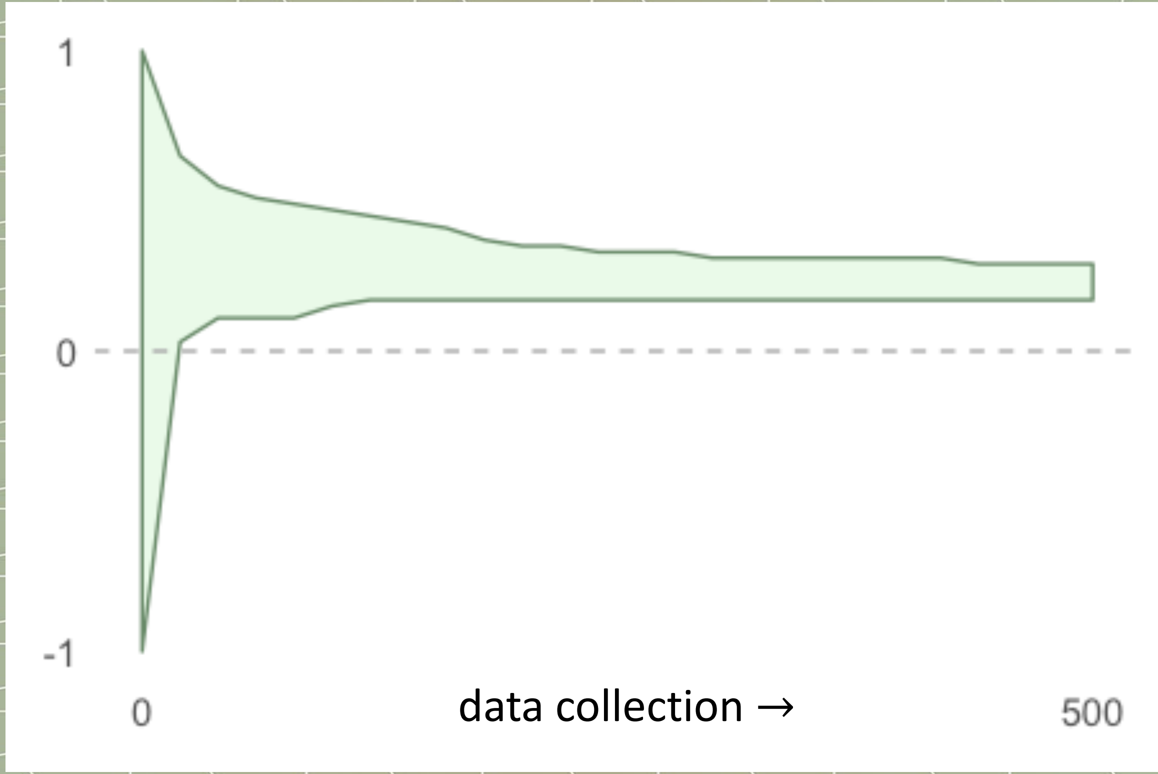
Anytime-valid Confidence Intervals for Contingency Tables and Beyond

Rosanne J. Turner

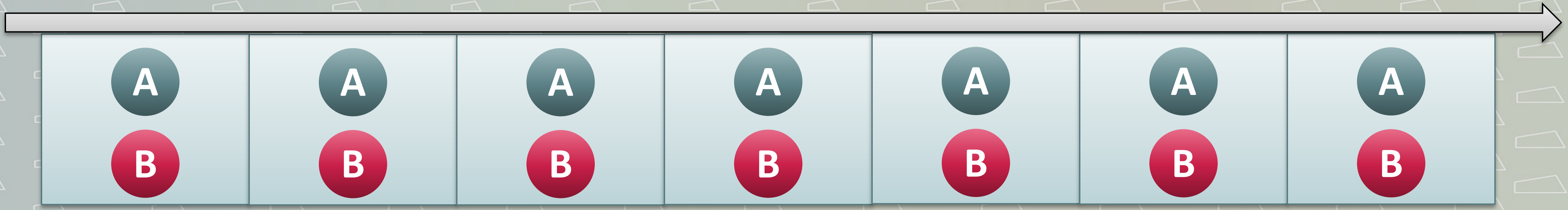
SAVI Worksop 2022

Joint work with Peter Grünwald

Goal: tests that can be used under optional stopping, *with* a notion of effect size



Setup



Setup

Setup


A B
A B
A B
A B
A B
A B
A B
A B B
A B B
A B B
A B B
A B B
A B B
A B B
A B B
A A B
A B B B
A A B
A B
A A A B
A B B

Setup

- $\mathcal{M} = \{P_\theta : \theta \in \Theta\}$ general parametric model, set of prob. distrs with densities or mass functions p_θ for random variable Y
- Two i.i.d. data streams $Y_{1,a}$, $Y_{2,a}$, ... and $Y_{1,b}, Y_{2,b}$, ...
- Want to create E-variable for block of n_a outcomes in group a, n_b outcomes in group b :

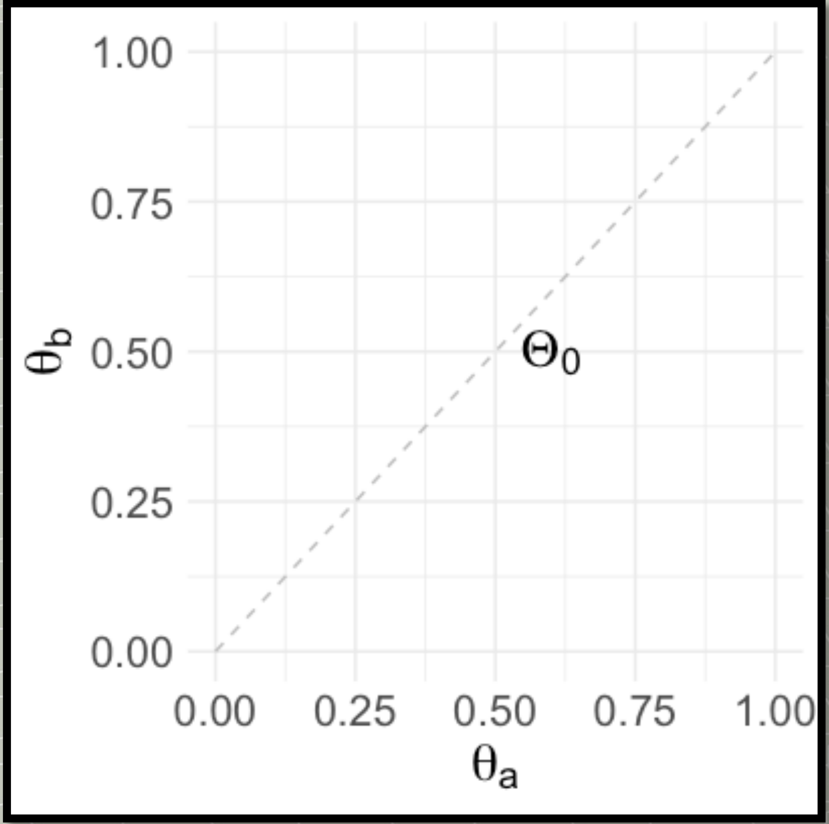
  $Y_a^{n_a} = (Y_{1,a}, \dots, Y_{n_a,a}), Y_b^{n_b} = (Y_{1,b}, \dots, Y_{n_b,b})$
- Take simple $\mathcal{H}_1$ indexed by (θ_a, θ_b) :

  likelihood is $\prod_{i=1..n_a} p_{\theta_a}(Y_{i,a}) \cdot \prod_{i=1..n_b} p_{\theta_b}(Y_{i,b})$
- Classical $\mathcal{H}_0$ in this setting: $\theta_a = \theta_b$, i.e. the set of distributions indexed by $\{(\theta_0, \theta_0): \theta_0 \in \Theta\}$

Running example: 2x2 contingency table setting

Do success probabilities differ between 2 strategies?

- $\mathcal{H}_0$: observations $Y \in \{0,1\}$ independent of strategy $X \in \{a, b\}$
- Equivalently, when $Y_x \overset{i.i.d.}{\sim}$ Bernoulli(θ_x): $\mathcal{H}_0$: $\theta_a = \theta_b$.



Idea through numerical optimization for finding GROW E-variable

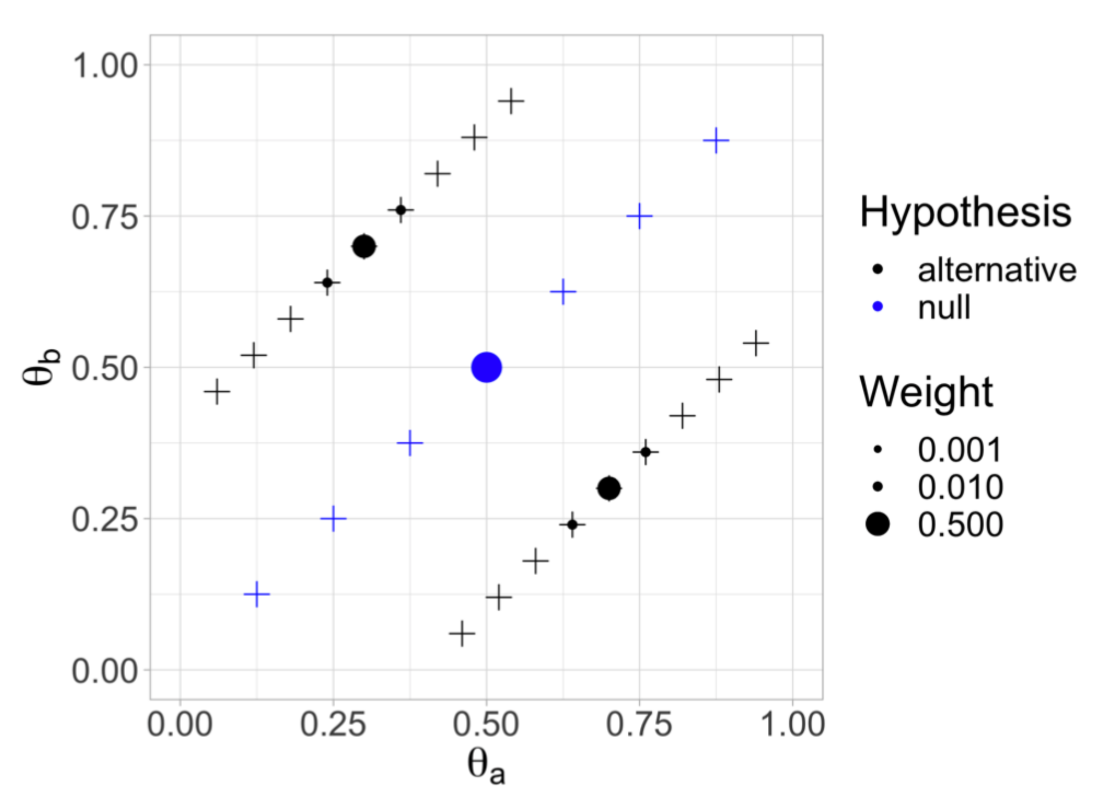


Figure 2.1a from Turner (2019), master thesis at Leiden University

Main Theorem of Turner et al. (2021)

Under no further regularity conditions, with $n = n_a + n_b$,

$$S^* := \prod_{i=1..n_a} \frac{p_{\theta_a}(Y_{i,a})}{\frac{n_a}{n} p_{\theta_a}(Y_{i,a}) + \frac{n_b}{n} p_{\theta_b}(Y_{i,a})} \cdot \prod_{i=1..n_b} \frac{p_{\theta_b}\left(Y_{i,b}\right)}{\frac{n_a}{n} p_{\theta_a}(Y_{i,b}) + \frac{n_b}{n} p_{\theta_b}(Y_{i,b})}$$

is an e-variable for the classical $\mathcal{H}_0$

If $\mathcal{M} = \{p_\theta : \theta \in \Theta\}$ is convex, S^* is the $(\boldsymbol{\theta_a}, \boldsymbol{\theta_b})$**-GRO e-variable,** achieving $\max_S \mathbf{E}_{Y_a^{n_a} \sim P_{\theta_a}, Y_b^{n_b} \sim P_{\theta_b}} [\log S]$ where the maximum is over all e-variables relative to $\mathcal{H}_0$

Proof sketch (i)

Let $G \in \{a, b\}$ satisfy $P(G = a) = \frac{n_a}{n}$ under both H_0 and H_1

- Apart from G **there is now just 1 (not n!) RV**, Y
- We observe (G, Y).
  - Under $\mathcal{H}_1$, (still a simple hypothesis indexed by (θ_a, θ_b)), $Y \sim P_{\theta_G}$
  - Under $\mathcal{H}_0$ (still a composite hypothesis with parameter $\theta_0 \in \Theta$) , $Y \sim P_{\theta_0}$ independently of G
- We will design an e-variable for this **modified testing problem** in which we randomize between observing an outcome from group a and b and then link it to our original problem in which we observe n_a and n_b of each (this proof technique may have broader applications…)

Proof sketch (ii)

Let $G \in \{a, b\}$ satisfy $P(G = a) = \frac{n_a}{n}$ under both H_0 and H_1

- We observe (G, Y).
  - Under $\mathcal{H}_1$, (still a simple hypothesis indexed by (θ_a, θ_b)), $Y \sim P_{\theta_G}$
  - Under $\mathcal{H}_0$ (still a composite hypothesis with parameter $\theta_0 \in \Theta$) , $Y \sim P_{\theta_0}$ independently of G
- $s(G, Y) := \frac{p_{\theta_G}(Y)}{\frac{n_a}{n} p_{\theta_a}(Y) + \frac{n_b}{n} p_{\theta_b}(Y)}$ is an e-variable, since under all distributions in the null,

  i.e. for all $\theta_0 \in \Theta$,

  $\mathbf{E}_G \mathbf{E}_{Y \sim P_{\theta_0}}[s(G, Y)] = \frac{n_a}{n} \mathbf{E}_{Y \sim P_{\theta_0}}[s(a, Y)] + \frac{n_b}{n} \mathbf{E}_{Y \sim P_{\theta_0}}[s(b, Y)] = 1$

Proof sketch (iii)

We thus have $\frac{n_a}{n}\ \mathbf{E}_{Y\sim P_{\theta_0}}[s(a,Y)]+\frac{n_b}{n}\ \mathbf{E}_{Y\sim P_{\theta_0}}[s(b,Y)] = 1$.

Young's inequality now gives $(\mathbf{E}_{Y\sim P_{\theta_0}}[s(a,Y)])^{n_a}\cdot(\mathbf{E}_{Y\sim P_{\theta_0}}[s(b,Y)])^{n_b} \leq \mathbf{1}(*)$

In original problem, we observe n_a Y_a's and n_b Y_b's . We need to show

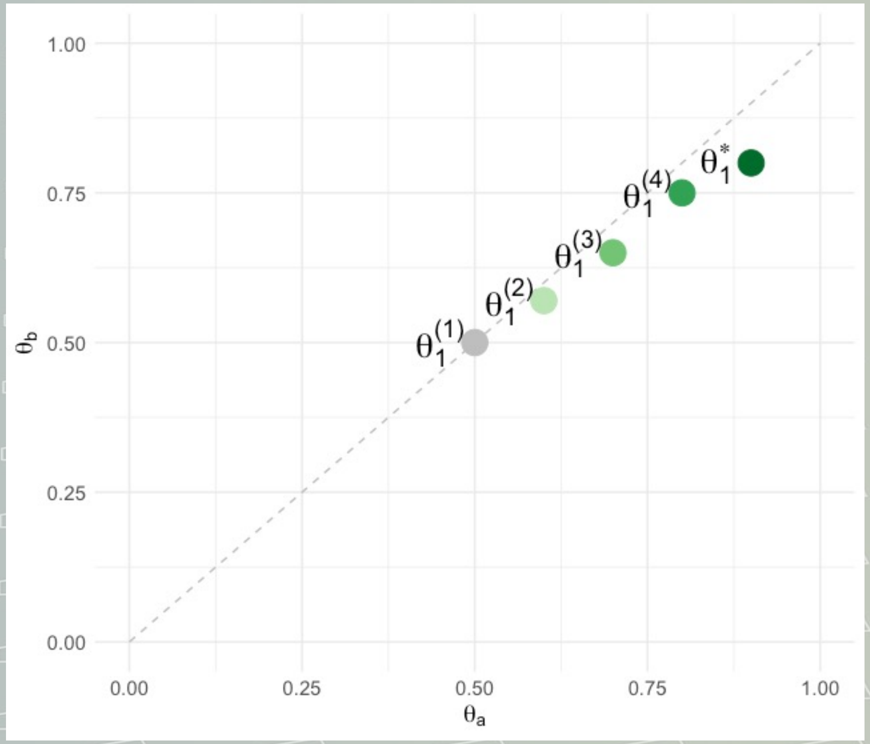
$$S^* := \prod_{i=1}^{n_a} \frac{p_{\theta_a}(Y_{i,a})}{\frac{n_a}{n}p_{\theta_a}(Y_{i,a})+\frac{n_b}{n}p_{\theta_b}(Y_{i,a})} \cdot \prod_{i=1}^{n_b} \frac{p_{\theta_b}(Y_{i,b})}{\frac{n_a}{n}p_{\theta_a}(Y_{i,b})+\frac{n_b}{n}p_{\theta_b}(Y_{i,b})}$$

is an e-variable. Using first independence and then (*) we get

$$\mathbf{E}_{Y^n\sim P_{\theta_0}}[S^*]=\left(\mathbf{E}_{Y\sim P_{\theta_0}}\left(\frac{p_{\theta_a}(Y)}{\frac{n_a}{n}p_{\theta_a}(Y)+\frac{n_b}{n}p_{\theta_b}(Y)}\right)\right)^{n_a}\cdot\left(\mathbf{E}_{Y\sim P_{\theta_0}}\left(\frac{p_{\theta_b}(Y)}{\frac{n_a}{n}p_{\theta_a}(Y)+\frac{n_b}{n}p_{\theta_b}(Y)}\right)\right)^{n_b} \leq 1$$

CWI Centrum Wiskunde & Informatica

Estimate (θ_a, θ_b) based on past blocks



- Allowed to estimate (θ_a, θ_b) for each new data block, based on past data
  - Maximum likelihood
  - MAP estimator
  - **Posterior mean**, …
- Restrict search space based on expert knowledge

Simulated example: 2x2 E-values vs classical counterpart

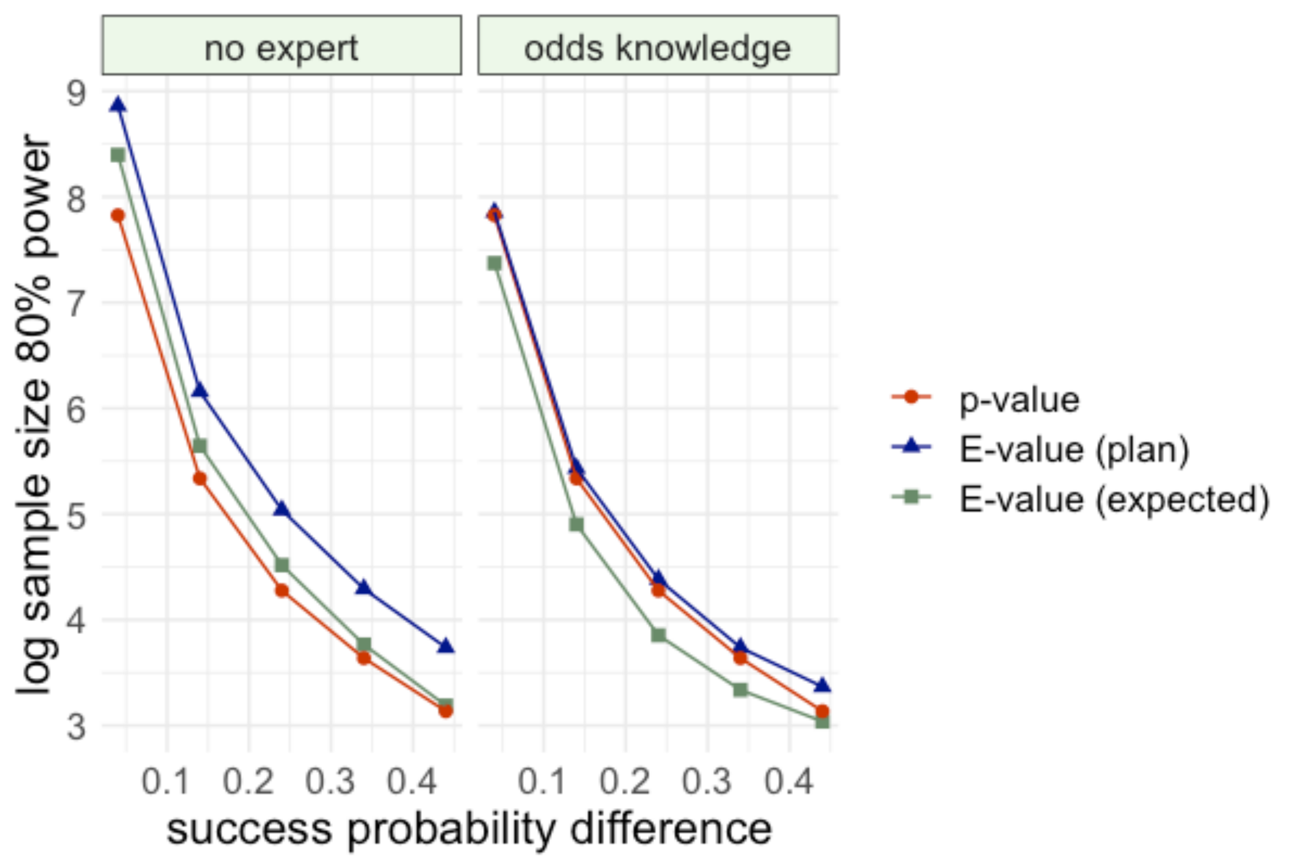
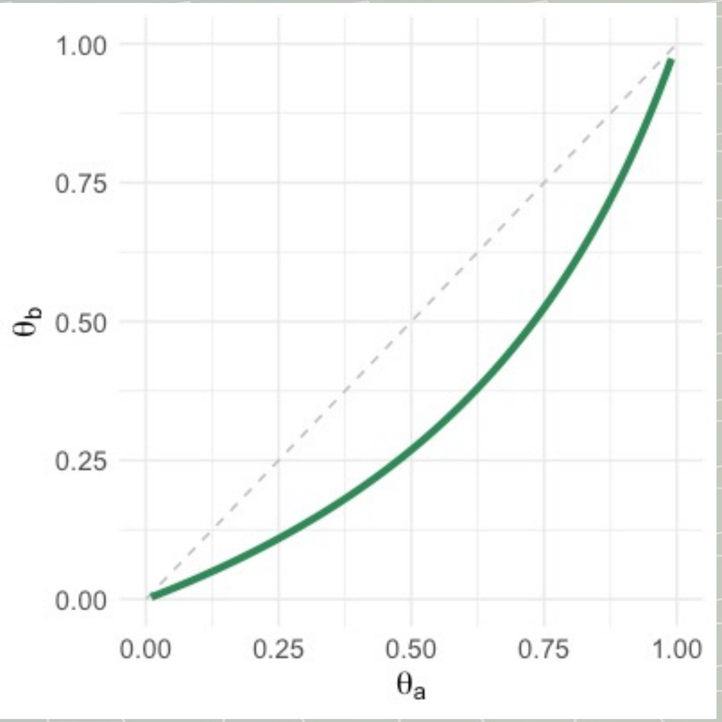


Figure adapted from Turner et al., 2021, figure 4

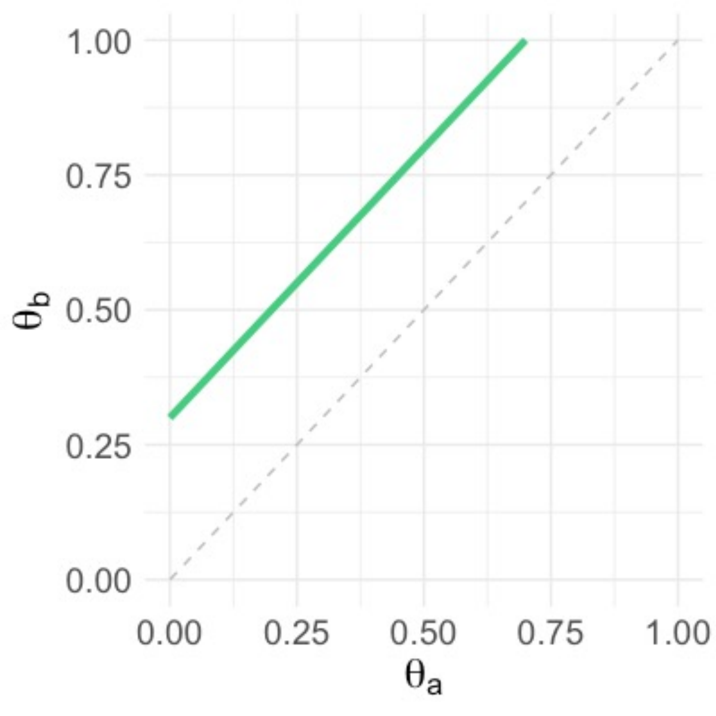
Centrum Wiskunde & Informatica

Extension to general $\mathcal{H}_0$

$\Theta_0(\delta) = \{(\theta_a, \theta_b): lOR(\theta_b, \theta_a) = -1\}$



$\Theta_0(\delta) = \{(\theta_a, \theta_b): \theta_b - \theta_a = 0.3\}$



E-variable for two-stream data, general $\mathcal{H}_0$

Theorem (Turner and Grünwald, 2022):

$$S_{\Theta_0}(Y^{(1)}) := \prod_{i=1}^{n_a} \frac{p_{\hat{\theta}_a}(Y_{i,a})}{p_{\theta_a^\circ}(Y_{i,a})} \prod_{i=1}^{n_b} \frac{p_{\hat{\theta}_b}(Y_{i,b})}{p_{\theta_b^\circ}(Y_{i,b})},$$ where $(\theta_a^\circ, \theta_b^\circ)$ achieve

$$\min_{(\theta_a,\theta_b)\in\Theta_0(\delta)} D(P_{\hat{\theta}_a,\hat{\theta}_b}(Y_a^{n_a}, Y_b^{n_b}) | P_{\theta_a,\theta_b}(Y_a^{n_a}, Y_b^{n_b})),$$

is an E-variable for $\mathcal{H}_0 := \{P_{\theta_a,\theta_b}: (\theta_a, \theta_b) \in \Theta_0(\delta)\}$

- We will neither precisely state nor prove the general result, but give an idea of the general way that allows us to establish E-variables for general $\mathcal{H}_0$ / Θ_0 with $\theta_a \neq \theta_b$
- Once again, we do this for the modified problem in which we observe a single random variable rather than $n_a + n_b$ of them

General $\mathcal{H}_0$: proof idea

Let $G \in \{a, b\}$ satisfy $p(a) := P(G = a) = \frac{n_a}{n}$ under both H_0 and H_1

- Apart from G there is now just 1 RV, Y
- We observe (G, Y).
  - Under $\mathcal{H}_1$, (simple hypothesis indexed by (θ_a, θ_b)),
    $p_{\theta_a,\theta_b}(G, Y) := p(G) p_{\theta_a,\theta_b}(Y \mid G)$ with $p_{\theta_a,\theta_b}(Y|G = g) := p_{\theta_g}(Y)$
  - Similarly under $\mathcal{H}_0$ (composite hypothesis with free param. $(\theta_a^*, \theta_b^*) \in \Theta_0^* \subset \Theta^2$,
    $p_{\theta_a^*,\theta_b^*}(G, Y) := p(G) p_{\theta_a^*,\theta_b^*}(Y \mid G)$
    with $p_{\theta^*{}_a,\theta_b^*}(Y|G = g) := p_{\theta_g^*}(Y)$
  - Let W be prior on Θ_0^*. Let $p_W(G, Y) := \int p_{\theta_a^*,\theta_b^*}(G, Y)\, dW(\theta_a^*, \theta_b^*)$

Then $s(G, Y) := \frac{p_{\theta_g}(Y)}{p_{W_0^*}(Y)}$ is an e-variable,
where W_0^* is the RIPr of (G., De Heide, Koolen, 2019, Thm 1) of P_{θ_a,θ_b} onto Θ_0^*

General $\mathcal{H}_0$: proof idea

Theorem (Turner and Grünwald, 2022):

$S_{\Theta_0}(Y^{(1)}) \coloneqq \prod_{i=1}^{n_a} \frac{p_{\hat{\theta}_a}(Y_{i,a})}{p_{\theta_a^\circ}(Y_{i,a})} \prod_{i=1}^{n_b} \frac{p_{\hat{\theta}_b}(Y_{i,b})}{p_{\theta_b^\circ}(Y_{i,b})}$, where $(\theta_a^\circ, \theta_b^\circ)$ achieve

$$\min_{(\theta_a,\theta_b)\in\Theta_0(\delta)} D(P_{\hat{\theta}_a,\hat{\theta}_b}(Y_a^{n_a}, Y_b^{n_b}) | P_{\theta_a,\theta_b}(Y_a^{n_a}, Y_b^{n_b})),$$

is an E-variable for $\mathcal{H}_0 \coloneqq \{P_{\theta_a,\theta_b} : (\theta_a, \theta_b) \in \Theta_0(\delta)\}$

- It turns out that $s(G,Y) := \frac{p_{\theta_g}(Y)}{p_{W_0^*}(Y)}$ reduces to the previous construction for the classical $\mathcal{H}_0$
- It can once again be linked to an E-variable in the original problem
- In the Bernoulli case, with convex Θ_0 , we then get the stated result.

Anytime-valid confidence sequences

Goal: confidence sequence CS with coverage at level $(1-\alpha)$:

- $P_{\theta_a,\theta_b}\big($ for any $m = 1, 2, \ldots : \delta(\theta_a, \theta_b) \notin CS_{(m)}\big) \leq \alpha$
- $\delta(\theta_a, \theta_b)$: arbitrary notion of effect size

- Construct $CS_{\alpha,(m)} = \left\{\delta : S^{(m)}_{\Theta_0(\delta)} \leq \frac{1}{\alpha}\right\}$
- Gives desired coverage because $S^{(m)}_{\Theta_0(\delta)}$ is an E-variable and offers Type-I error guarantee at level α

Simulations: risk difference

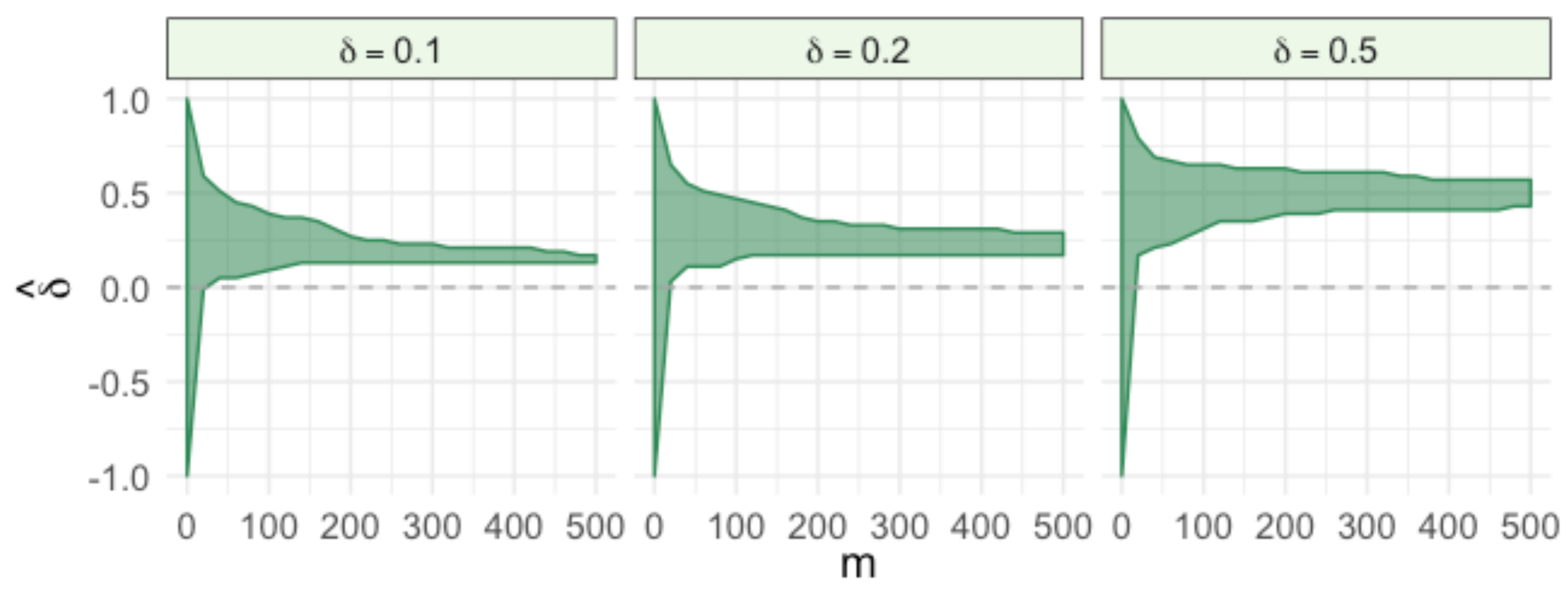
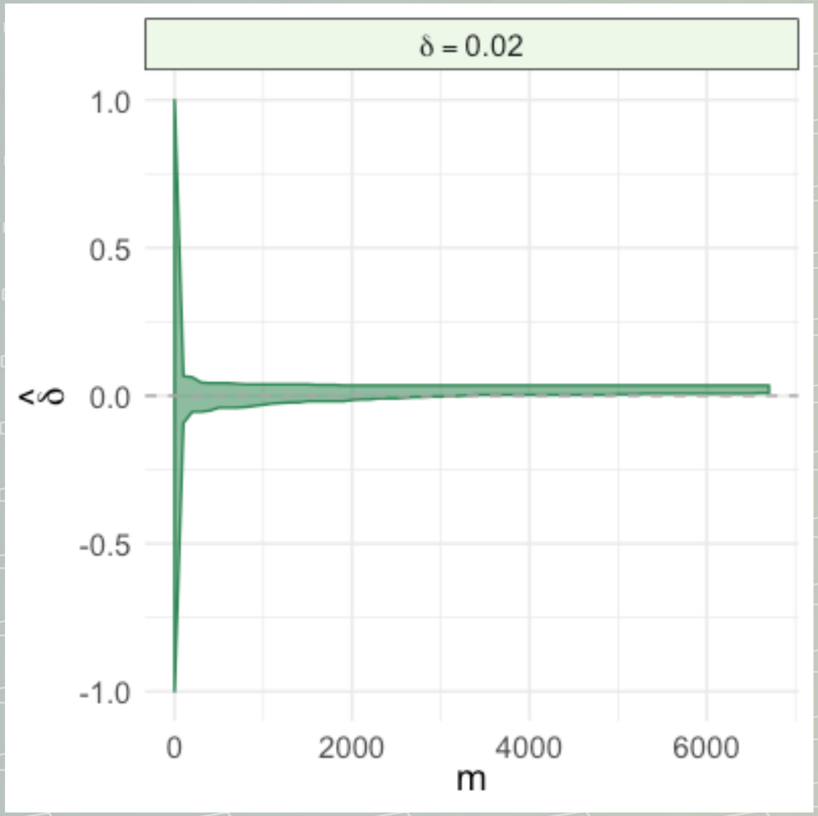
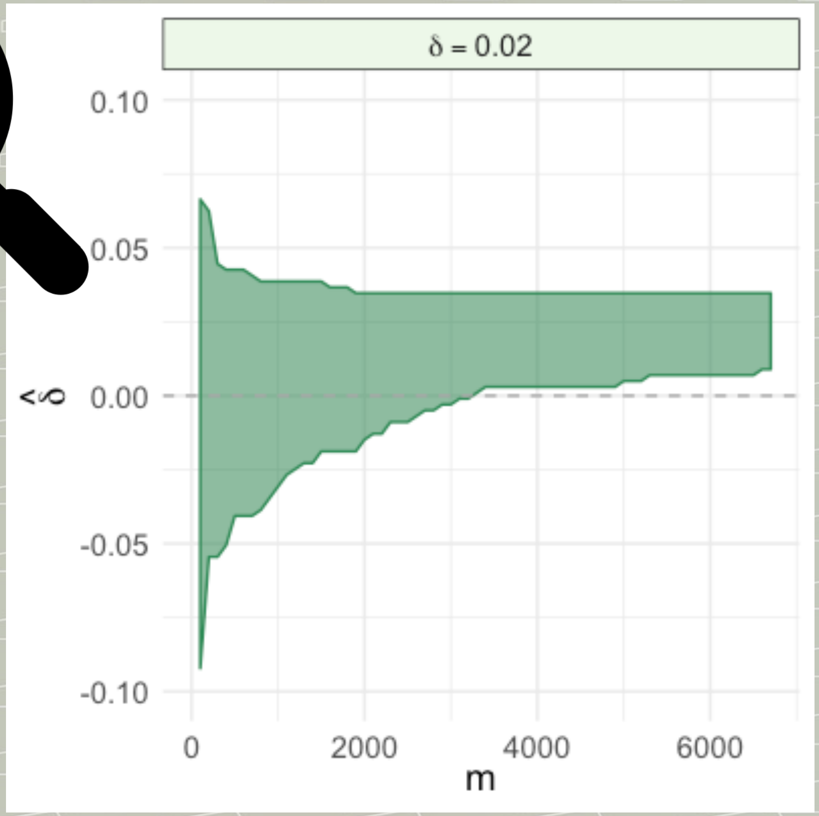


Figure adapted from Turner et al., 2022

Simulations: risk difference





Tricky case: odds ratio and convexity of $\mathcal{H}_0$

- Need convexity of $\Theta_0(\delta)$ to construct E-variable
- $\delta > 0 \rightarrow$ can estimate lower bound (see figure)
- $\delta < 0 \rightarrow$ can estimate upper bound

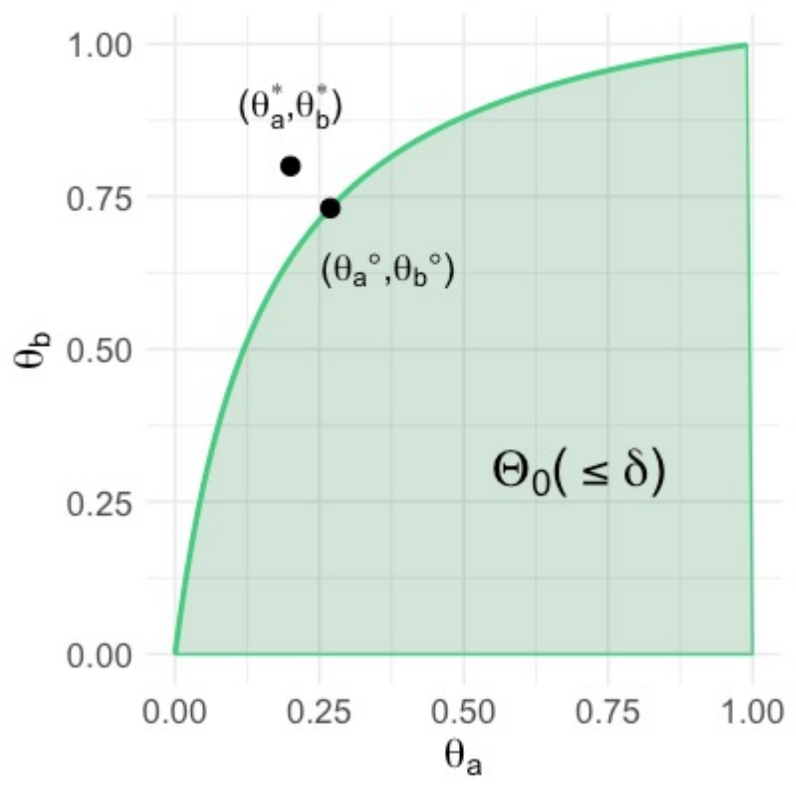


Figure adapted from Turner et al., 2022

Centrum Wiskunde & Informatica

Tricky case: odds ratio and convexity of $\mathcal{H}_0$

- Need convexity of $\Theta_0(\delta)$ to construct E-variable
- $\delta > 0 \rightarrow$ can estimate lower bound
- $\delta < 0 \rightarrow$ can estimate upper bound (see figure)

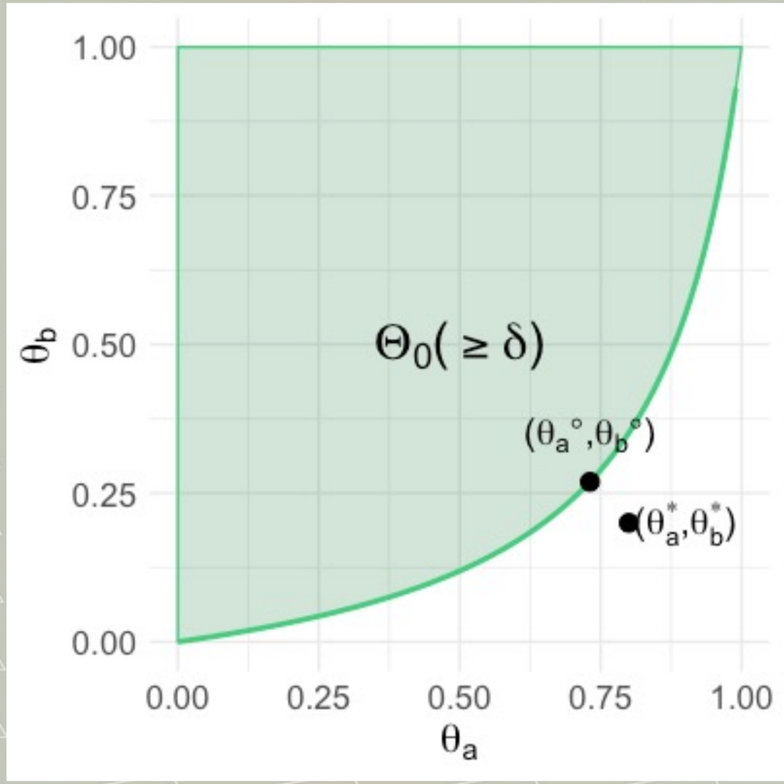
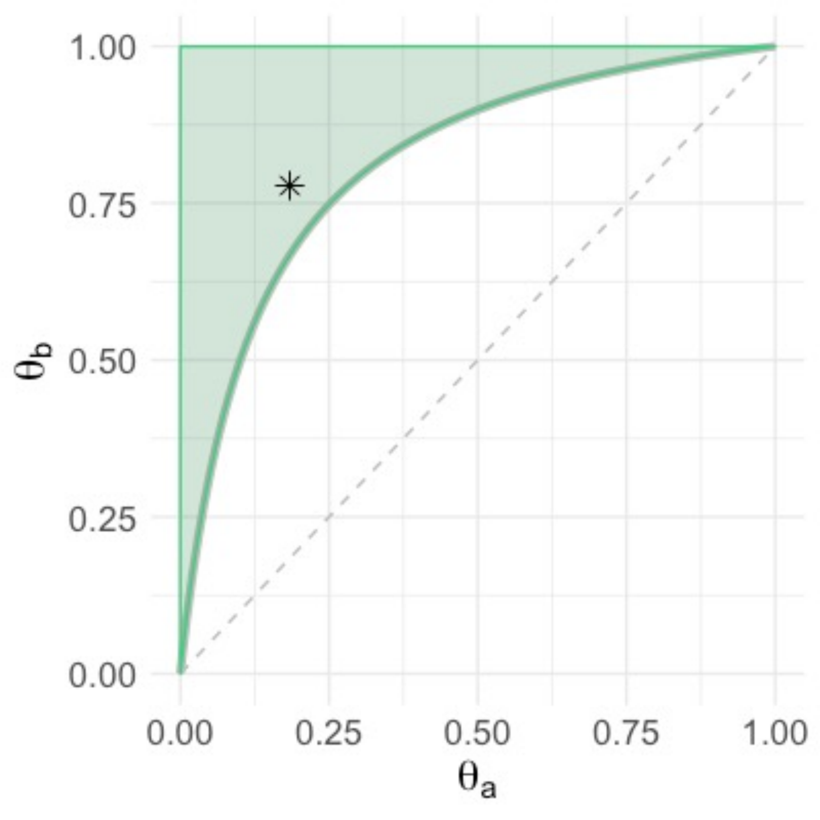
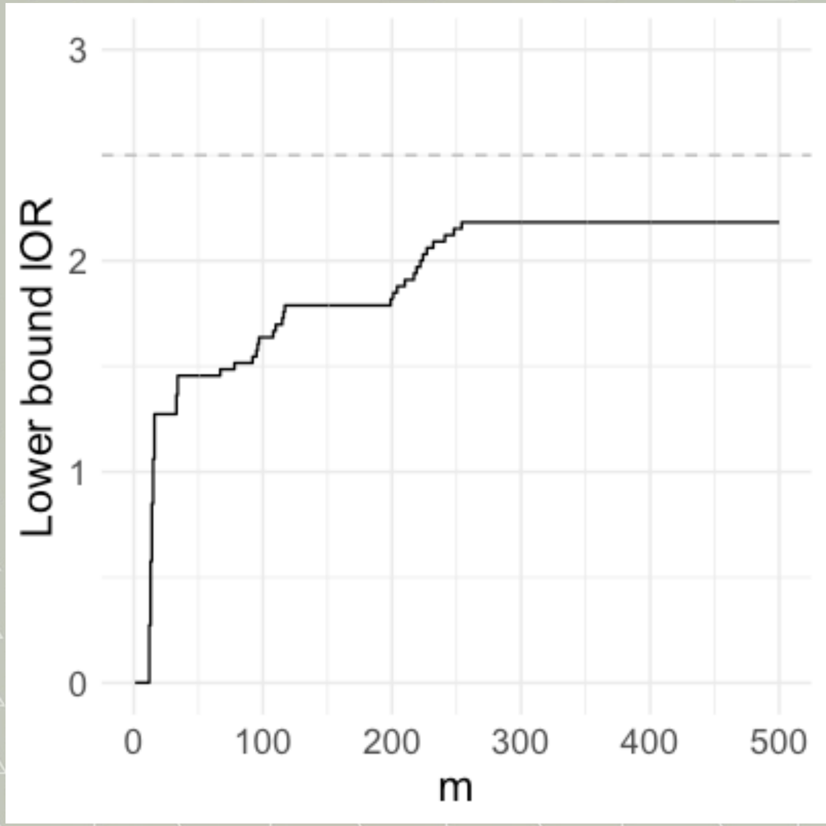


Figure adapted from Turner et al., 2022

Simulation: log of the odds ratio



One-sided CS^{+} at data block $m = 500$

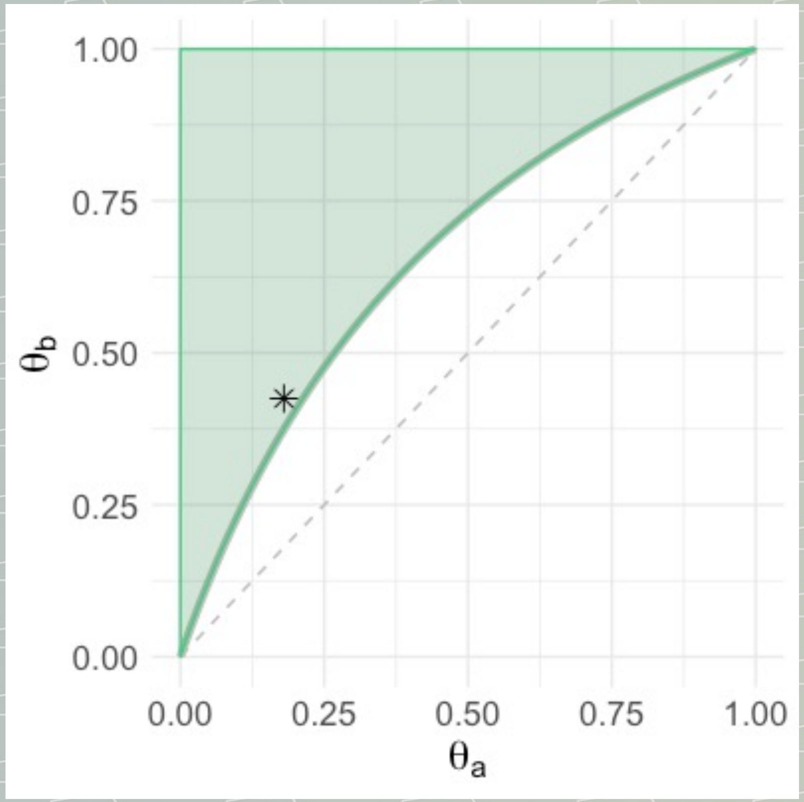


lower bound over time

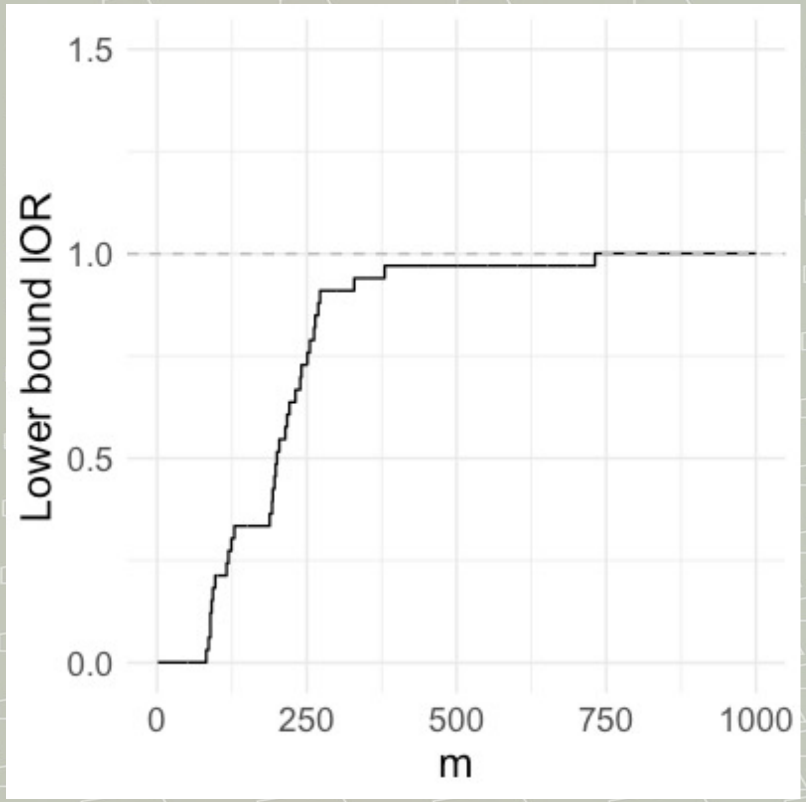
Figure adapted from Turner et al., 2022

Centrum Wiskunde & Informatica

Simulation: log of the odds ratio



One-sided CS^{+} at data block $m = 500$



lower bound over time

Figure adapted from Turner et al., 2022

Conclusion and novelty

- To our knowledge, really new:
  - **flexibility** (block size, user-specified notions of effect size)
  - **growth rate optimality**: expect evidence for H1 to **grow as fast as possible** during data collection
- Wald's sequential probability ratio test:
  - Probability ratios can be interpreted as "alternative" E-variables
  - Not growth-rate optimal
  - Only allow for testing odds ratio effect size

CWI Centrum Wiskunde & Informatica

Extensions

- Beyond Bernoulli: GRO property? (work by Y. Hao and others)
- Stratified data and conditional independence
  - Use case at UMC Utrecht: real-time psychiatry research and recommendations

		Strategy A	Strategy B
Stratum 1	Success	S(A1)	S(B1)
Stratum 1	Failure	F(A1)	F(B1)
Stratum 2	Success	S(A2)	S(B2)
Stratum 2	Failure	F(A2)	F(B2)
Stratum 3	Success	S(A3)	S(B3)
Stratum 3	Failure	F(A3)	F(B3)

Further reading and references

- On the theory of E-values:
  - P.D. Grünwald, R. de Heide and W. Koolen (2019) on ArXiv:
- On implementations of E-values:
  - R.J. Turner, A. Ly and P.D. Grünwald (2021) on ArXiv:2106.02693
  - R.J. Turner and P.D. Grünwald (2022) on ArXiv:2203.09785
  - R software: https://CRAN.R-project.org/package=safestats

In R console:
install.packages("safestats")