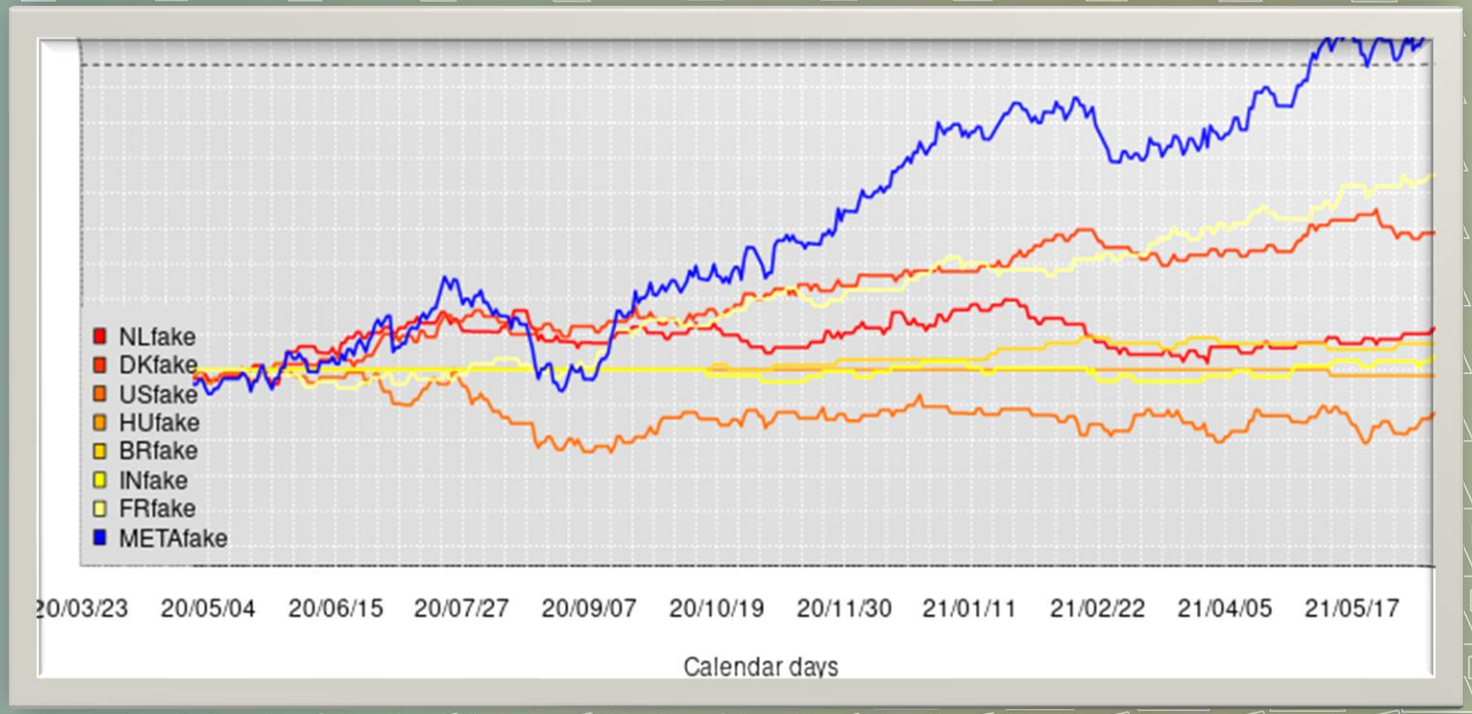
CWI Centrum Wiskunde & Informatica

MACHINE LEARNING

ALL-IN meta-analysis



Judith ter Schure

CWI Centrum Wiskunde & Informatica

MACHINE LEARNING

ALL-IN meta-analysis




Peter Grünwald Judith ter Schure Alexander Ly Muriel Pérez

ALL-IN meta-analysis properties

Type-I error whenever you analyze

- Continuous monitoring
- Unlimited horizon (no prespecified maximum sample size)
- Possibly 100% power
- Live meta-analysis (update after each new observation)
- Effortless cumulative meta-analysis (unlimited number of trials)

CWI Centrum Wiskunde & Informatica

MACHINE LEARNING

ALL-IN meta-analysis

A*nytime*
L*ive* and
L*eading*
IN*terim* meta-analysis

ALL-IN meta-analysis properties

Type-I error whenever you analyze

- Continuous monitoring
- Unlimited horizon (no prespecified maximum sample size)
- Possibly 100% power
- Live meta-analysis (update after each new observation)
- Effortless cumulative meta-analysis (unlimited number of trials)

I'll give an example of time-to-event analysis (logrank test)
ALL-IN-META-BCG-CORONA

ALL-IN meta-analysis properties

Type-I error whenever you analyze

- Continuous monitoring
- Unlimited horizon (no prespecified maximum sample size)
- Possibly 100% power
- Live meta-analysis (update after each new observation)
- Effortless cumulative meta-analysis (unlimited number of trials)

I'll give an example of time-to-event analysis (logrank test)
ALL-IN-META-BCG-CORONA

Whenever you see *e*-value, you can read: likelihood ratio

ALL-IN meta-analysis

- Meta evidence by ***multiplying*** study e-values
- **Type-I error control** under continuous monitoring and **any decision** to start, stop or expand studies
- Evidence against
  the global null (= no risk difference ***in any study***)
  contribution by any study with hazard ratio < 0.8

e-values and Ville's inequality

$$E\langle i\rangle = \frac{\boldsymbol{P}(\ \textit{observed event } i\ \mid\ \textit{hazard ratio } = 0{,}8\ )}{\boldsymbol{P}(\ \textit{observed event } i\ \mid\ \textit{hazard ratio } = 1\ )} \quad \begin{matrix} H_1 \\ H_0 \end{matrix}$$

- If $E\langle i\rangle$ is an *e-value* for a single event i

$$E^{\langle n\rangle} = \prod_{i=1}^{n} E\langle i\rangle$$

is also an *e-value* for n events

- Ville's inequality (1939)

$$\boldsymbol{P}_{\boldsymbol{events} \sim \boldsymbol{H_0}}\Big(\ E^{\langle n\rangle} \geq \frac{1}{\alpha} \ \text{ for some } \ n\ \Big) \leq \alpha$$

e-values and Ville's inequality

$$E\langle i\rangle = \frac{\boldsymbol{P}(\ \textit{observed event } i\ \mid\ \textit{hazard ratio } = 0{,}8\ )}{\boldsymbol{P}(\ \textit{observed event } i\ \mid\ \textit{hazard ratio } = 1\ )} \quad \begin{matrix} H_1 \\ H_0 \end{matrix}$$

- If $E\langle i\rangle$ is an *e-value* for a single event i

$$E^{\langle n\rangle} = \prod_{i=1}^{n} E\langle i\rangle$$

is also an *e-value* for n events

- Ville's inequality (1939)

$$\boldsymbol{P}_{\boldsymbol{events} \sim \boldsymbol{H_0}}\Big(\ E^{\langle n\rangle} \geq \frac{1}{0{,}05} \ \text{ for some } \ n\ \Big) \leq 0{,}05$$

e-values and Ville's inequality

$$E\langle i\rangle = \frac{\boldsymbol{P}(\ \textit{observed event } i\ \mid\ \textit{hazard ratio } = 0{,}8\ )}{\boldsymbol{P}(\ \textit{observed event } i\ \mid\ \textit{hazard ratio } = 1\ )} \quad \begin{matrix} H_1 \\ H_0 \end{matrix}$$

- If $E\langle i\rangle$ is an *e-value* for a single event i

$$E^{\langle n\rangle} = \prod_{i=1}^{n} E\langle i\rangle$$

is also an *e-value* for n events

- Ville's inequality (1939)

$$\boldsymbol{P}_{\boldsymbol{events} \sim \boldsymbol{H_0}}\big(\ \ E^{\langle n\rangle} \geq 20 \ \text{ for some } n\ \big) \leq 0{,}05$$

e-values and Ville's inequality

$$E^{\langle n \rangle} = \prod_{i=1}^{n} \frac{\boldsymbol{P}(\ \textit{observed event } i \ \mid \ \textit{hazard ratio } = 0{,}8\ )}{\boldsymbol{P}(\ \textit{observed event } i \ \mid \ \textit{hazard ratio } = 1\ )} \quad \begin{matrix} H_1 \\ H_0 \end{matrix}$$

- If $E_1^{\langle n_1(t) \rangle}$ is an *e-value* for n_1 events at time t in study 1, testing H_0

  $E_K^{\langle n_K(t) \rangle}$ is an *e-value* for n_K events at time t in study K, testing H_0

e-values and Ville's inequality

$$E^{\langle n \rangle} = \prod_{i=1}^{n} \frac{\boldsymbol{P}(\ \textit{observed event } i \ \mid \ \textit{hazard ratio } = 0{,}8\ )}{\boldsymbol{P}(\ \textit{observed event } i \ \mid \ \textit{hazard ratio } = 1\ )} \quad \begin{matrix} H_1 \\ H_0 \end{matrix}$$

- If $E_1^{\langle n_1(t) \rangle}$ is an *e-value* for n_1 events at time t in study 1, testing H_0

  $E_K^{\langle n_K(t) \rangle}$ is an *e-value* for n_K events at time t in study K, testing H_0

$$E^{\langle t \rangle} = \prod_{k=1}^{K} E_k^{\langle n_k(t) \rangle}$$

is also an *e-value for* $n_1(t) + \cdots + n_K(t)$ events at time t, testing H_0

e-values and Ville's inequality

$$E^{\langle n \rangle} = \prod_{i=1}^{n} \frac{\boldsymbol{P}(\ \textit{observed event } i \ \mid \ \textit{hazard ratio } = 0{,}8\ )}{\boldsymbol{P}(\ \textit{observed event } i \ \mid \ \textit{hazard ratio } = 1\ )} \quad \begin{matrix} H_1 \\ H_0 \end{matrix}$$

- If $E_1^{\langle n_1(t) \rangle}$ is an *e-value* for n_1 events at time t in study 1, testing H_0

  $E_K^{\langle n_K(t) \rangle}$ is an *e-value* for n_K events at time t in study K, testing H_0

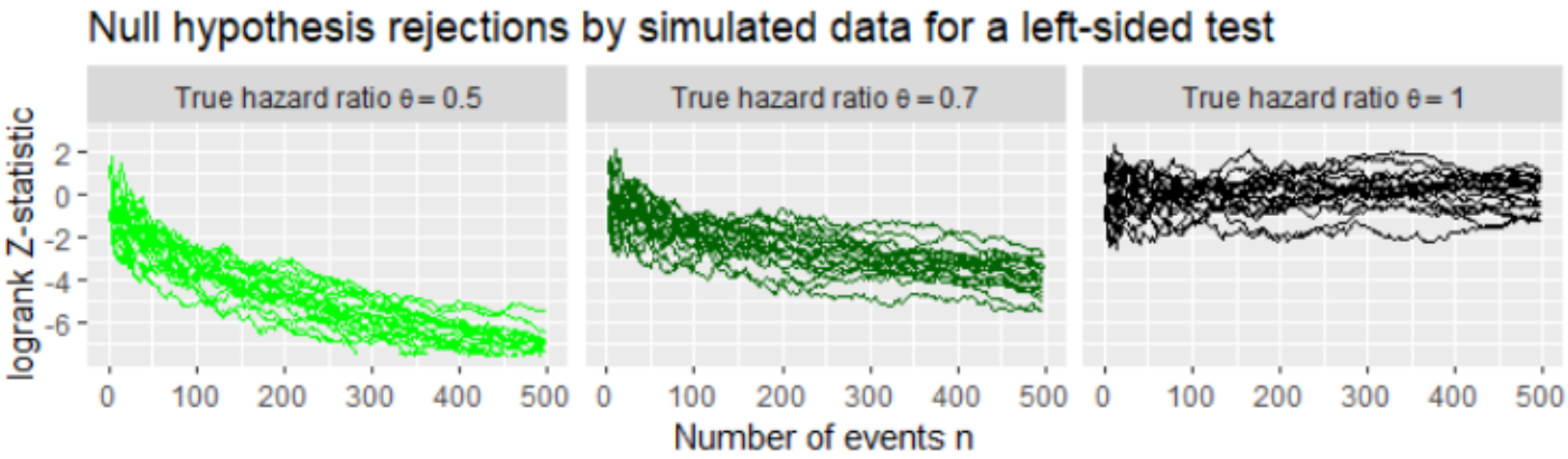
$$E^{\langle t \rangle} = \prod_{k=1}^{K} E_k^{\langle n_k(t) \rangle}$$

is also an *e-value for* $n_1(t) + \cdots + n_K(t)$ events at time t, testing H_0

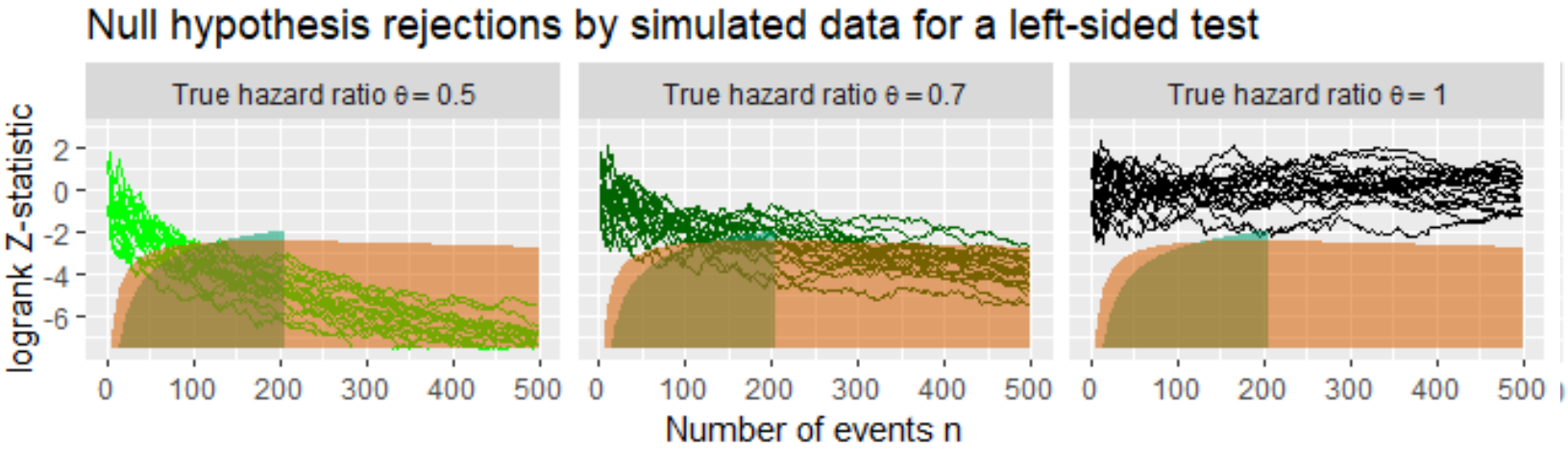
- Ville's inequality (1939)

$$\boldsymbol{P}_{\boldsymbol{events} \sim H_0}\left(\quad E^{\langle t \rangle} \geq \frac{1}{\alpha} \quad \text{for some } \ t \ \right) \leq \alpha$$

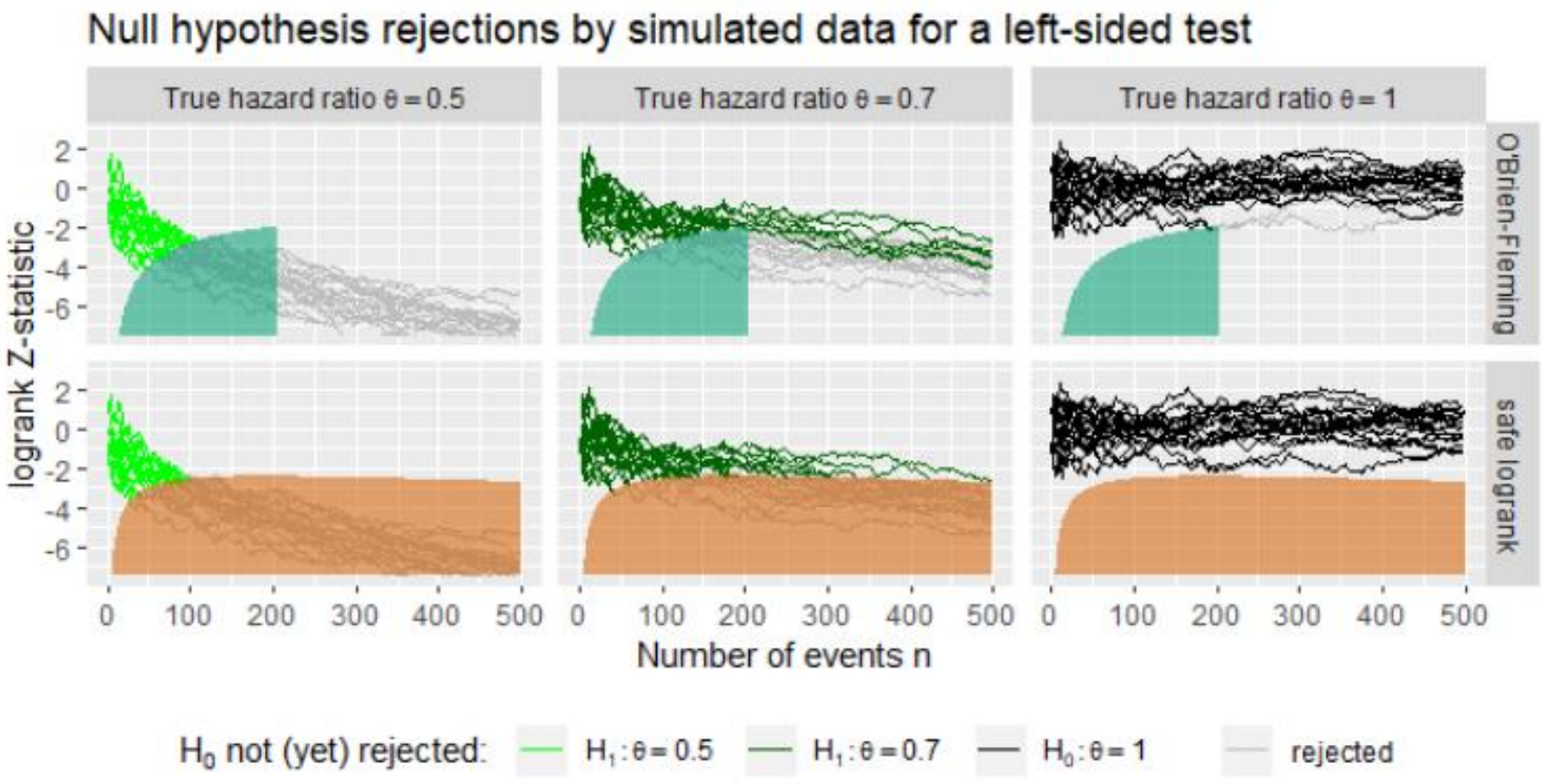
Possibly 100% power



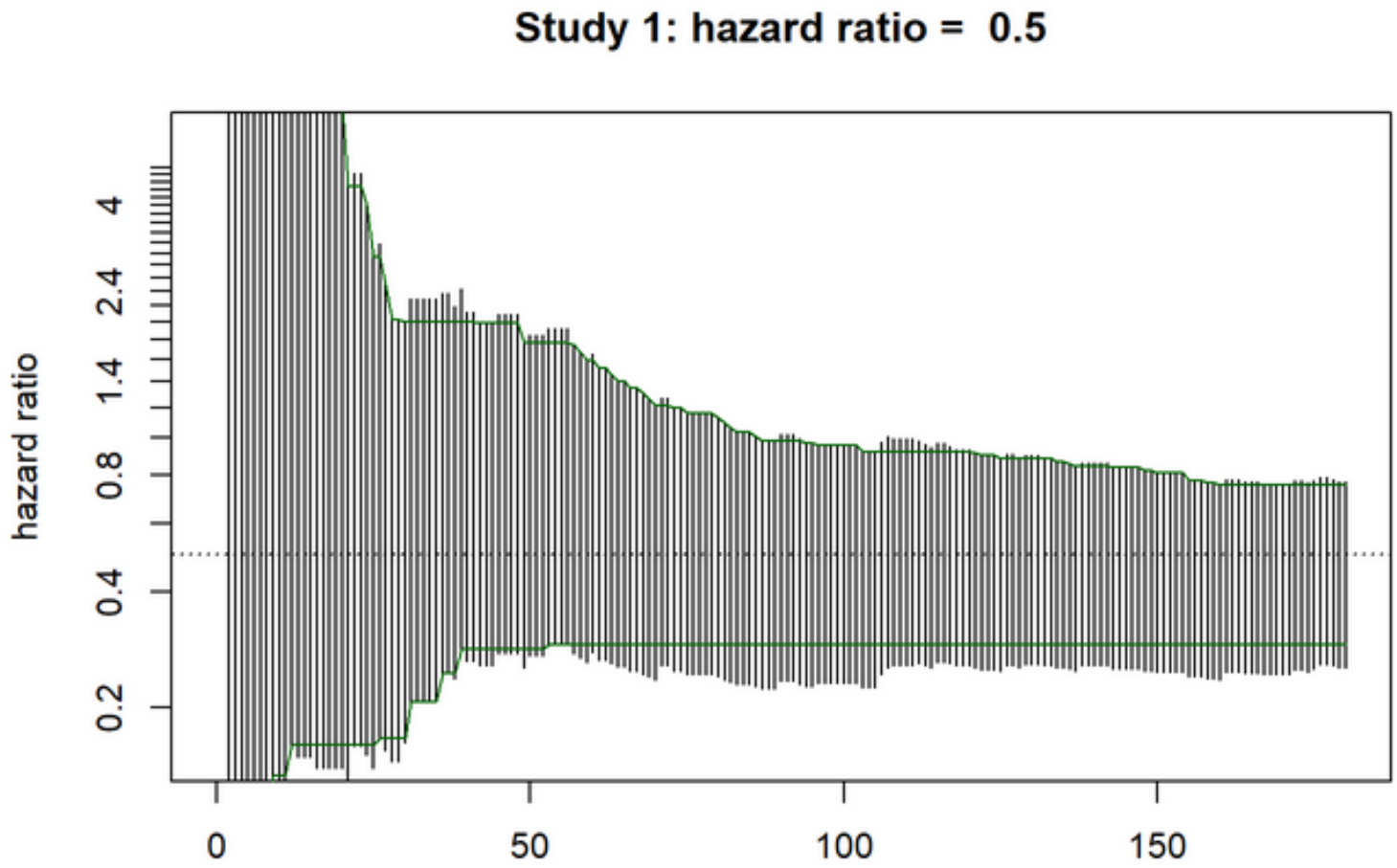
Possibly 100% power



Possibly 100% power



Anytime-valid confidence sequences



What are we assuming? Points for discussion

Heterogeneity

Simplicity

Anytime-valid confidence sequences

What are we assuming? Points for discussion

Heterogeneity

- I believe we should not be testing whether some mean of true effects is 0, but whether all are 0 (global null)

Simplicity

Anytime-valid confidence sequences

What are we assuming? Points for discussion

Heterogeneity

- I believe we should not be testing whether some mean of true effects is 0, but whether all are 0 (global null)

Simplicity

- Likelihood ratios are easy to interpret,
  if you can specify an effect size of minimal clinical relevance

Anytime-valid confidence sequences

What are we assuming? Points for discussion

Heterogeneity

- I believe we should not be testing whether some mean of true effects is 0, but whether all are 0 (global null)

Simplicity

- Likelihood ratios are easy to interpret,
  if you can specify an effect size of minimal clinical relevance
- If you do not want to focus on such decision making
  (one vs the other hypothesis) go for estimation
  -> be prepared to need more trials

Anytime-valid confidence sequences

What are we assuming? Points for discussion

Heterogeneity

- I believe we should not be testing whether some mean of true effects is 0, but whether all are 0 (global null)

Simplicity

- Likelihood ratios are easy to interpret,
  if you can specify an effect size of minimal clinical relevance
- If you do not want to focus on such decision making
  (one vs the other hypothesis) go for estimation
  -> be prepared to need more trials

Anytime-valid confidence sequences

- We follow Peto in this regard: e.g. ‘typical hazard ratio’ interval

What are we assuming? Points for discussion

Heterogeneity

- I believe we should not be testing whether some mean of true effects is 0, but whether all are 0 (global null)

Simplicity

- Likelihood ratios are easy to interpret,
  if you can specify an effect size of minimal clinical relevance
- If you do not want to focus on such decision making
  (one vs the other hypothesis) go for estimation
  -> be prepared to need more trials

Anytime-valid confidence sequences

- We follow Peto in this regard: e.g. ‘typical hazard ratio’ interval
- Future work confidence sequences based on random-effects model

References

- ter Schure, J.A., Pérez-Ortiz, M.F., Ly, A., & Grünwald, P.D. (2021). The Safe Logrank Test: Error Control under Continuous Monitoring with Unlimited Horizon. arXiv:2011.06931

- ALL-IN meta-analysis webinar (2020) https://ir.cwi.nl/pub/30338
originally for trials contributing to ALL-IN-META-BCG-CORONA

- ter Schure, J.A, & Grünwald, P.D. (2019). Accumulation Bias in meta-analysis: the need to consider time in error control [version 1; peer review: two approved]. *F1000Research*, *8*. doi:10.12688/f1000research.19375.1

- Peto, Richard. Why do we need systematic overviews of randomized trials? (Transcript of an oral presentation, modified by the editors). *Statistics in medicine* 6.3 (1987): 233-240.

Thank you!

References

- ter Schure, J.A., Pérez-Ortiz, M.F., Ly, A., & Grünwald, P.D. (2021). The Safe Logrank Test: Error Control under Continuous Monitoring with Unlimited Horizon. arXiv:2011.06931

- ALL-IN meta-analysis webinar (2020) https://ir.cwi.nl/pub/30338
originally for trials contributing to ALL-IN-META-BCG-CORONA

- ter Schure, J.A, & Grünwald, P.D. (2019). Accumulation Bias in meta-analysis: the need to consider time in error control [version 1; peer review: two approved]. *F1000Research*, *8*. doi:10.12688/f1000research.19375.1

- Peto, Richard. Why do we need systematic overviews of randomized trials? (Transcript of an oral presentation, modified by the editors). *Statistics in medicine* 6.3 (1987): 233-240.