

# RQ: Real-time evidence collection in data streams

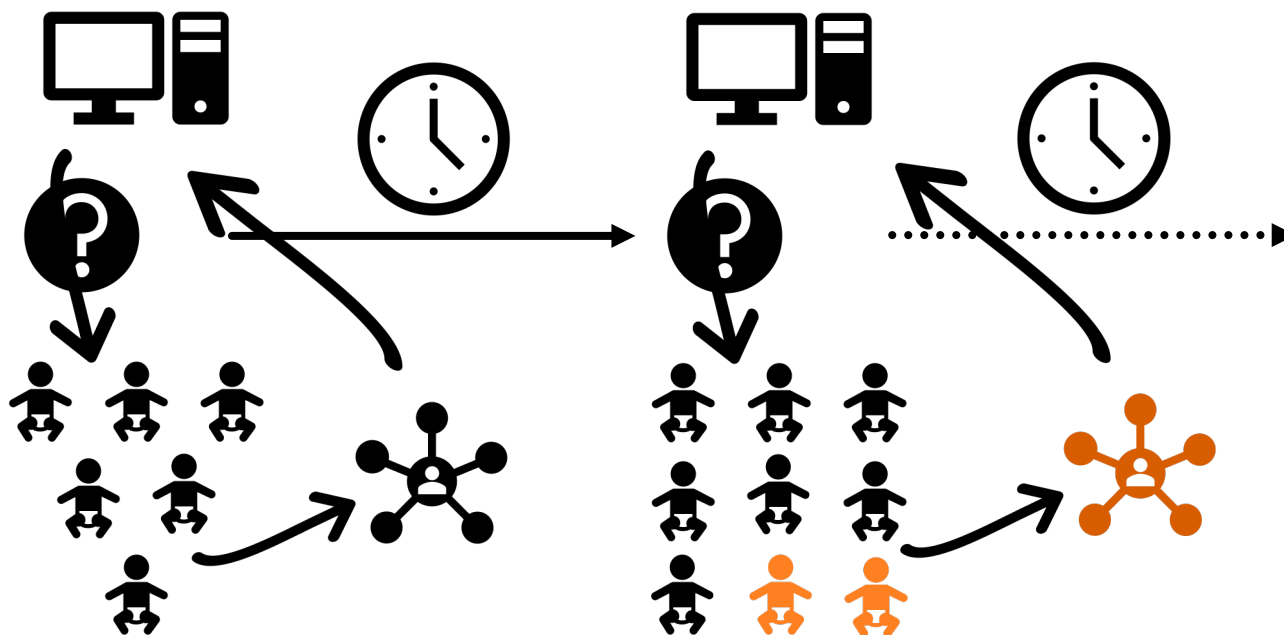
Updates april 2021

- ▶ Rosanne J. Turner
- ▶ Supervisors and collaborators within EPI:
  - ▶ Prof. Peter Grünwald (CWI)
  - ▶ Prof. Floor Scheepers (UMCU)
  - ▶ Karin Hagoort (UMCU)
  - ▶ Dr. Aki Harma (Philips)
  - ▶ Roel van Est (Parnassia)



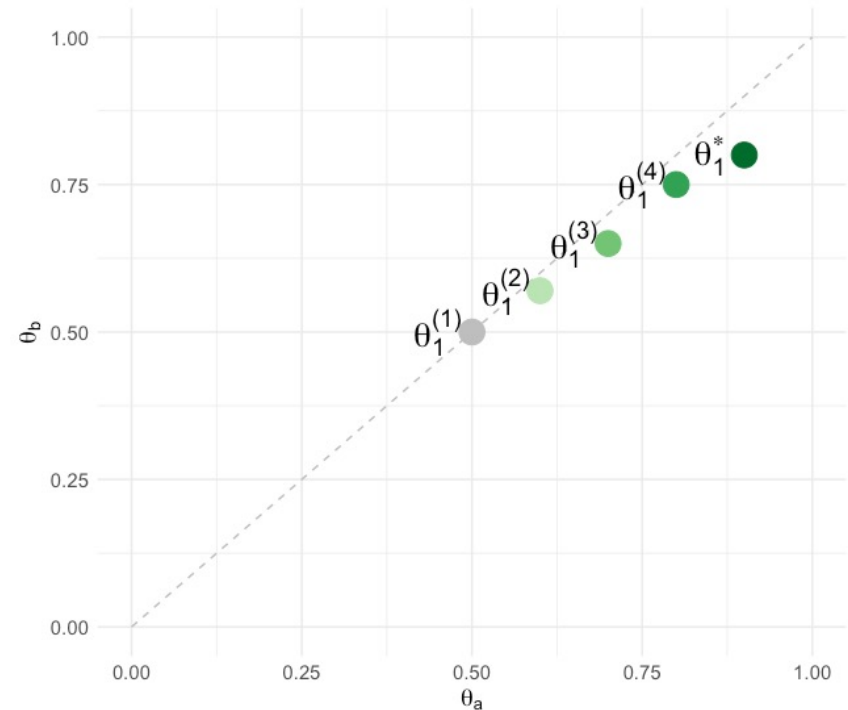
# Real-time analysis: safe statistics

- ▶ Collect and adjust inference about evidence for treatment strategies **in real time**



# Learn from data for adjusting evidence collection process

- ▶ With safe statistics can collect **“statistically sound” evidence for A/B testing (e.g. treatment recommendations!)** in real-time
- ▶ Learning approach: take safe test for next data block that is optimal relative to estimate  $\hat{\theta}_1$  based on data seen so far
- ▶ Technical detail: take  $\hat{\theta}_1$  as posterior mean for  $\theta_1$  with Beta prior on parameter space



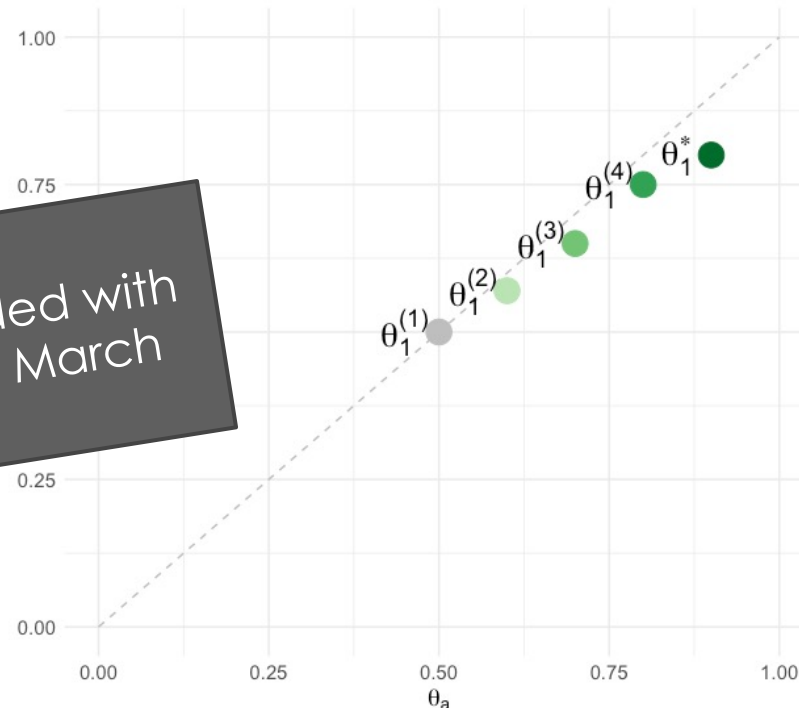
# Learn from data for adjusting evidence collection process

- ▶ With safe statistics can collect “statistically sound” evidence for A/B testing (e.g. treatment recommendations!) in real-time

- ▶ Learning and

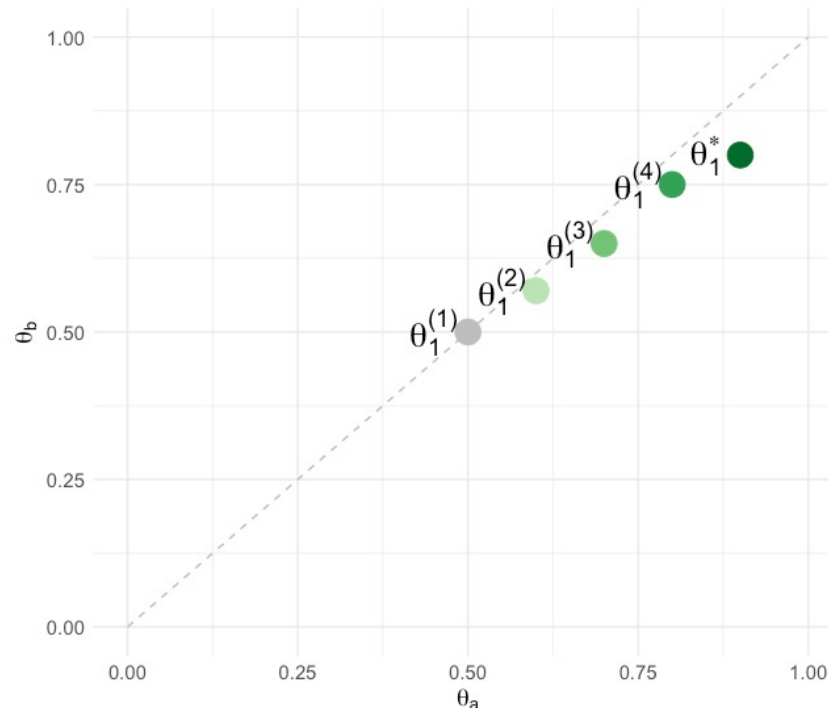
Work partly from master thesis; awarded with the VVSOR Jan Hemelrijk award last March

- ▶ In more detail: take  $\hat{\theta}_1$  as posterior mean for  $\theta_1$  with Beta prior on parameter space



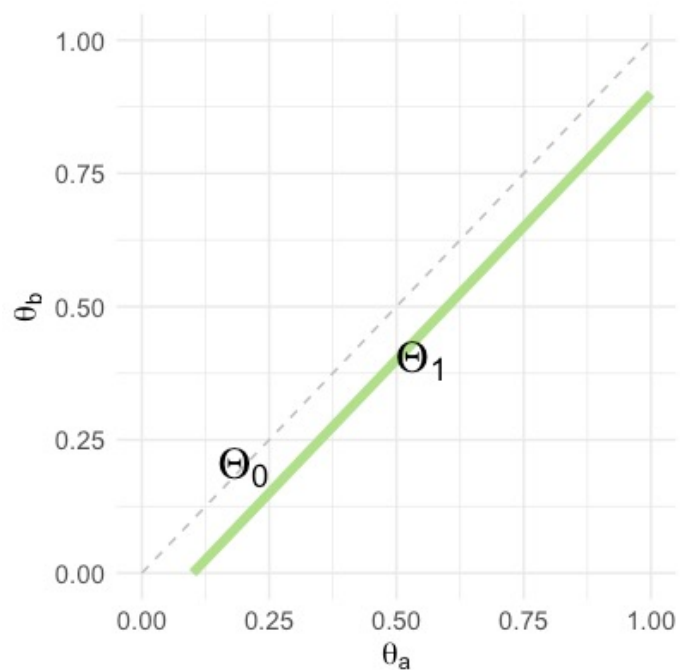
# Learn from data for adjusting evidence collection process

- ▶ With safe statistics can collect **“statistically sound” evidence for A/B testing (e.g. treatment recommendations!)** in real-time
- ▶ Learning approach: take safe test for next data block that is optimal relative to estimate  $\hat{\theta}_1$  based on data seen so far
- ▶ Technical detail: take  $\hat{\theta}_1$  as posterior mean for  $\theta_1$  with Beta prior on parameter space

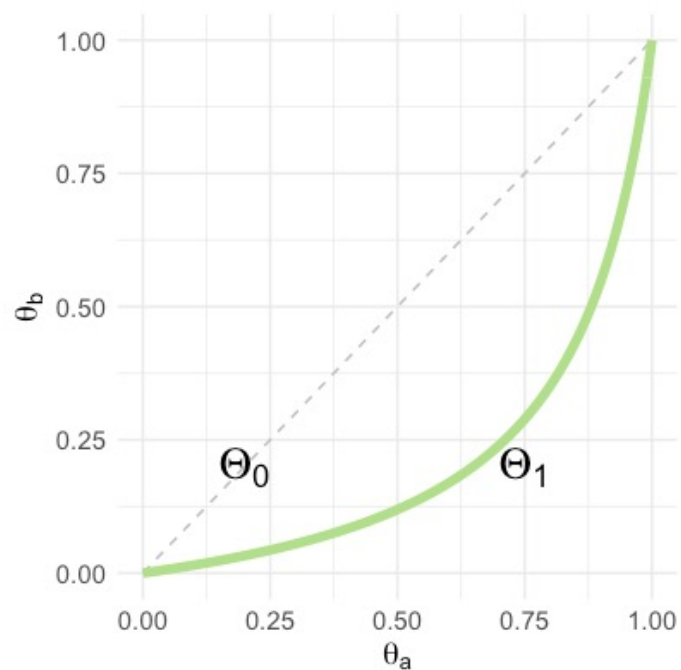


# Can also add expert/ prior knowledge: restrict options!

Absolute difference



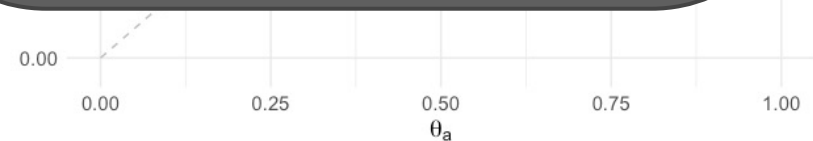
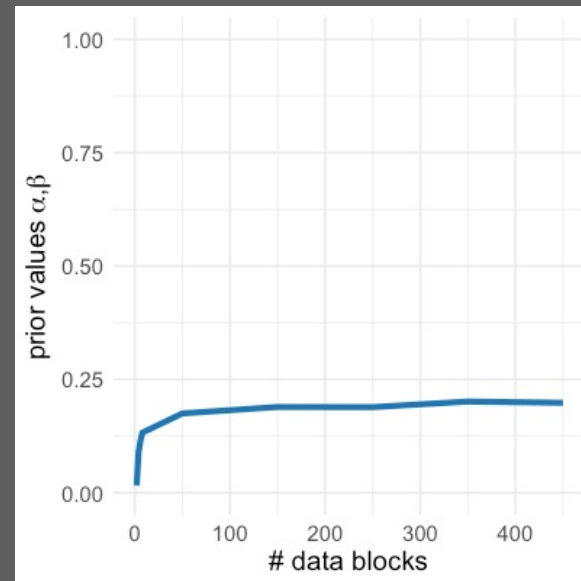
Odds ratio/ relative risk



# Learn from data for evidence collection

Progress since last meeting: Beta priors that are “optimal” w.r.t. REGRET and data collected

- ▶ With safe statistics can collect “**statistically sound**” evidence for A/B testing (e.g. **treatment recommendations!**) in real-time
- ▶ Learning approach: take safe test for next data block that is optimal relative to estimate  $\hat{\theta}_1$  based on data seen so far
- ▶ Technical detail: take  $\hat{\theta}_1$  as posterior mean for  $\theta_1$  with Beta prior on parameter space



# Real-life example: SWEPIs<sup>1</sup> perinatal death study

- ▶ Comparing perinatal death in labour induction at 41 or 42 weeks
- ▶ “All significance tests were two sided at the 0.05 level.”

**After observing approx. 1380 births in each group:**

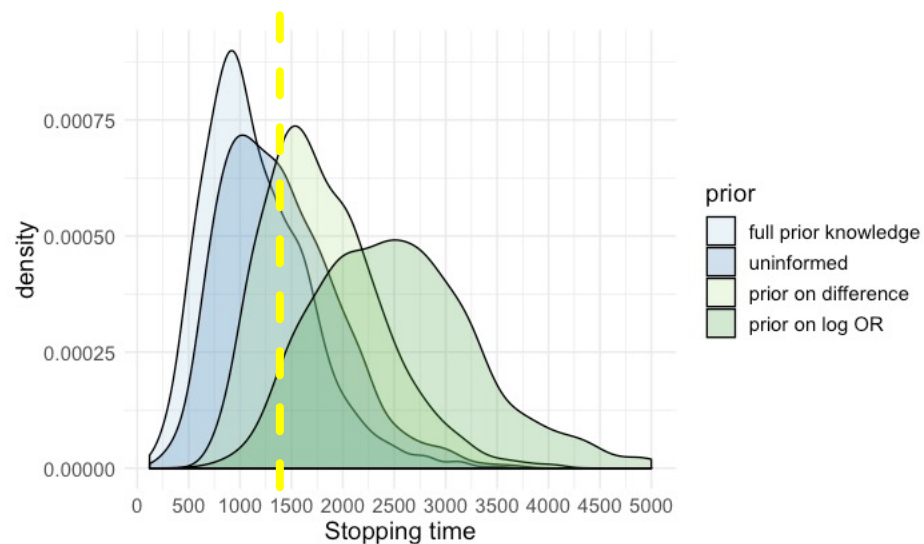
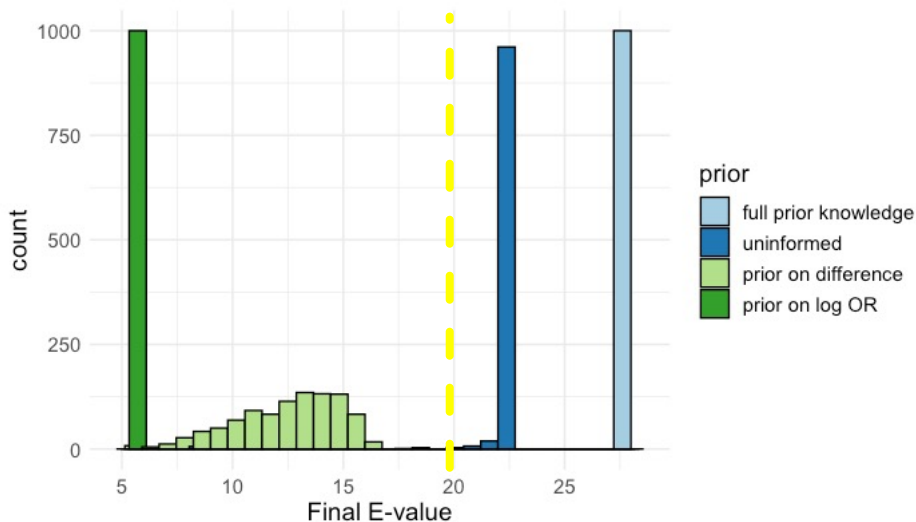
“On 2 October 2018 the Data and Safety Monitoring Board strongly recommended the SWEPIs steering committee to **stop the study owing to a statistically significant higher perinatal mortality in the expectant management group**. Although perinatal mortality was a secondary outcome, it was not considered ethical to continue the study. **No perinatal deaths occurred in the early induction group but six occurred in the expectant management group (five stillbirths and one early neonatal death; P=0.03).**”<sup>1</sup>



# Safe testing applied to collect evidence in the SWEPIIS scenario

Optionally: use knowledge from previous studies in prior of safe test

- ▶ Mean perinatal death rate at 41 weeks: 0.0001
- ▶ Difference risk between 42 and 41 weeks: 0.00318



# Application to use case: evaluating the usefulness of a recommender system for treatment of depression

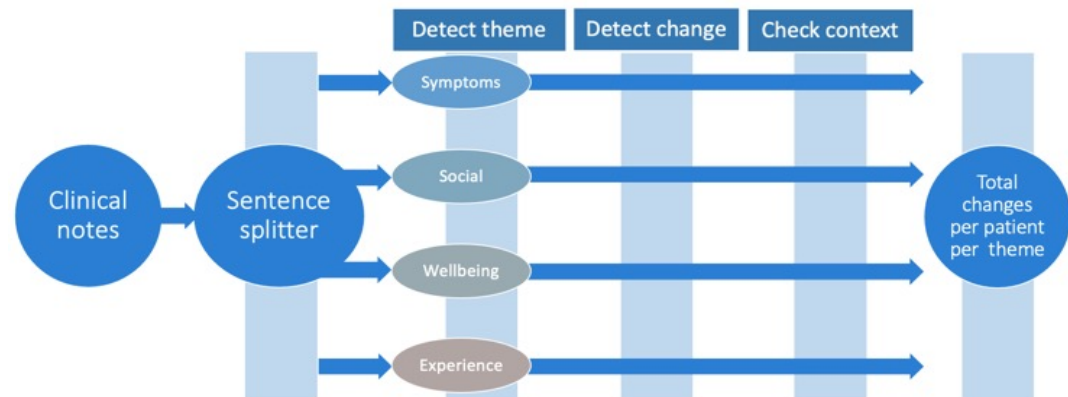
## Extracting information on the outcome of treatment trajectories from electronic health records in psychiatry

R.J. Turner<sup>1,2</sup>, F. Coenen<sup>2</sup>, K. Hagoort<sup>2</sup>, F.E. Scheepers<sup>2</sup>, P.D. Grünwald<sup>1</sup>, and A. Härmä<sup>3</sup>

<sup>1</sup>CWI, Amsterdam, NL

<sup>2</sup>UMC Utrecht, Utrecht, NL

<sup>3</sup>Philips Research, Eindhoven, NL



# Application to use case: evaluating the usefulness of a recommender system for treatment of depression

## Extracting information on the outcome of treatment trajectories from electronic health records in psychiatry

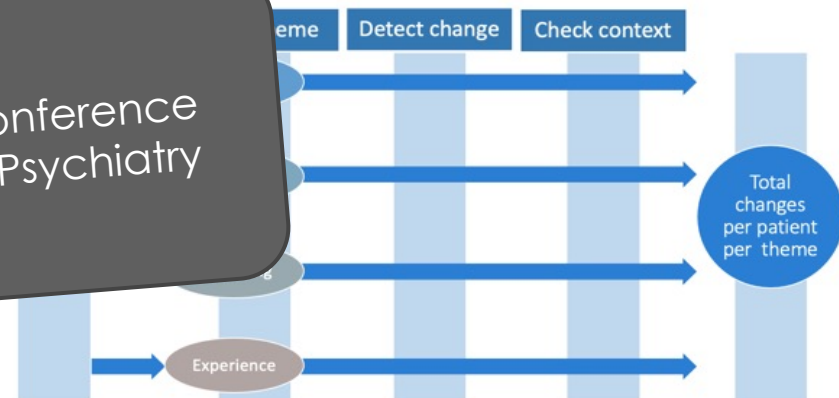
R.J. Turner<sup>1,2</sup>, F. Coenen<sup>2</sup>, K. Hagoort<sup>2</sup>, F.E. Scheepers<sup>2</sup>, P.D. Grünwald<sup>1</sup>, and A. Härmä<sup>3</sup>

<sup>1</sup>CWI, Amsterdam, NL

<sup>2</sup>UMC Utrecht, Utrecht, NL

<sup>3</sup>Philips Research, Eindhoven, NL

NLP part submitted to OCUPAI'21 conference  
Full draft will be submitted to JAMA Psychiatry



# Application to use case: evaluating the usefulness of a recommender system for treatment of depression

- ▶ Which types of recommendations assist clinicians the best?
- ▶ Plan: offer, in (micro-)randomized format, different forms of recommendations to clinicians based on the four outcome measures extracted from free text
  - ▶ ECT
  - ▶ Antidepressants
- ▶ Continuously analyze results with safe tests

ARTICLE

Open Access

## How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection

Maia Jacobs<sup>1</sup>, Melanie F. Pradier<sup>1</sup>, Thomas H. McCoy Jr.<sup>2,3</sup>, Roy H. Perlis<sup>2,3</sup>, Finale Doshi-Velez<sup>1</sup> and Krzysztof Z. Gajos<sup>1</sup>