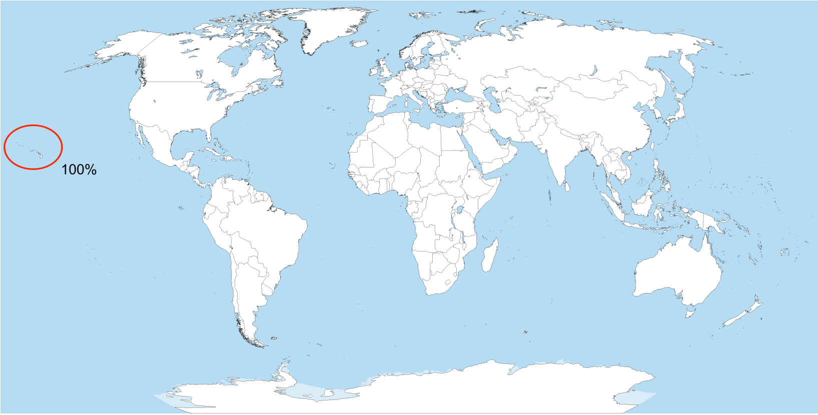
Using Bayesian sequential designs to efficiently estimate the effects of digital behaviour interventions

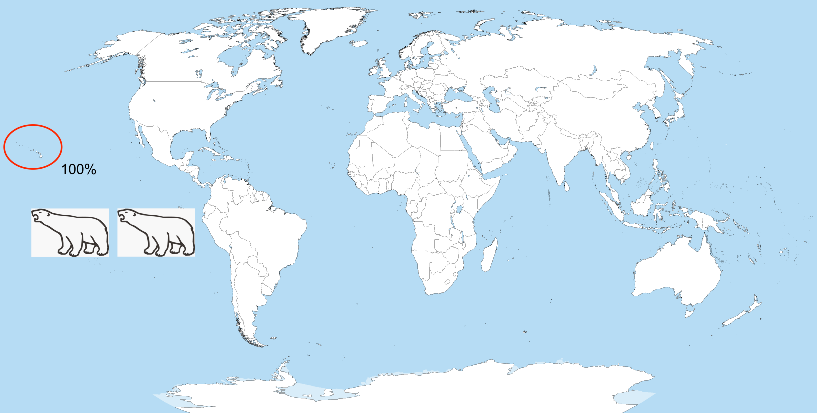
Marcus Bendtsen, PhD
Senior Associate Professor of Biostatistics in Public Health
marcus.bendtsen@liu.se

Linköping University
Department of Health, Medicine and Caring Sciences

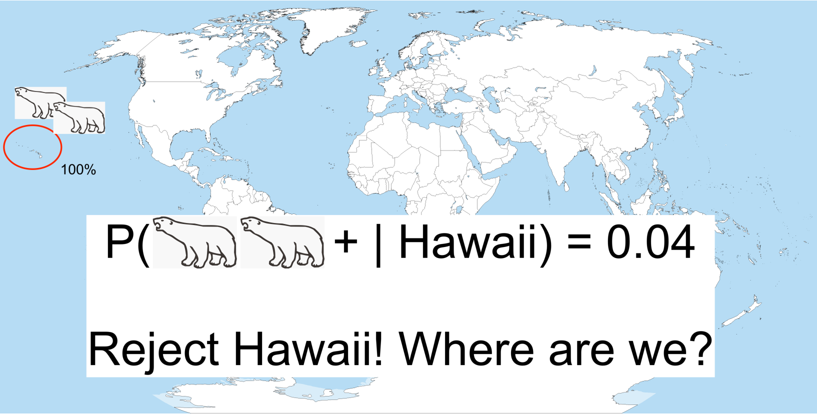
Polar bears and null-hypothesis testing



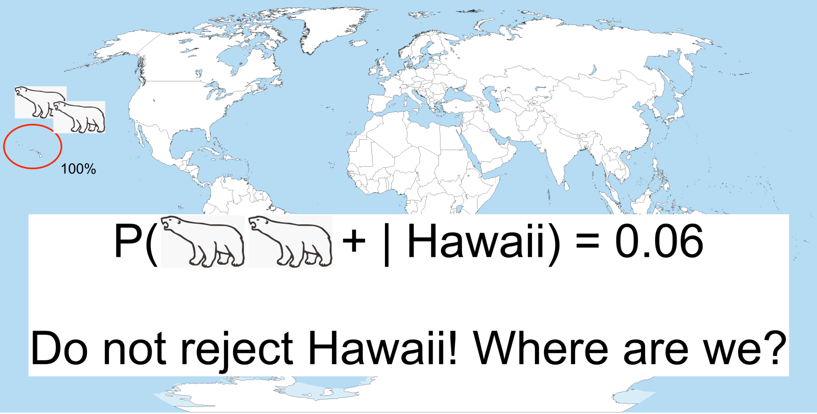
Polar bears and null-hypothesis testing



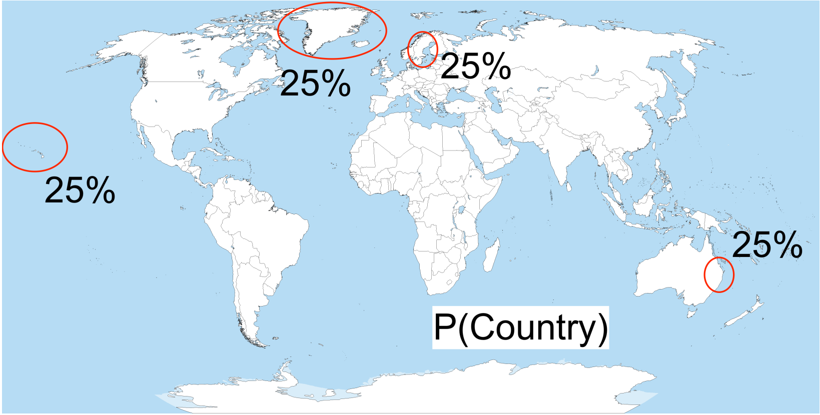
Polar bears and null-hypothesis testing



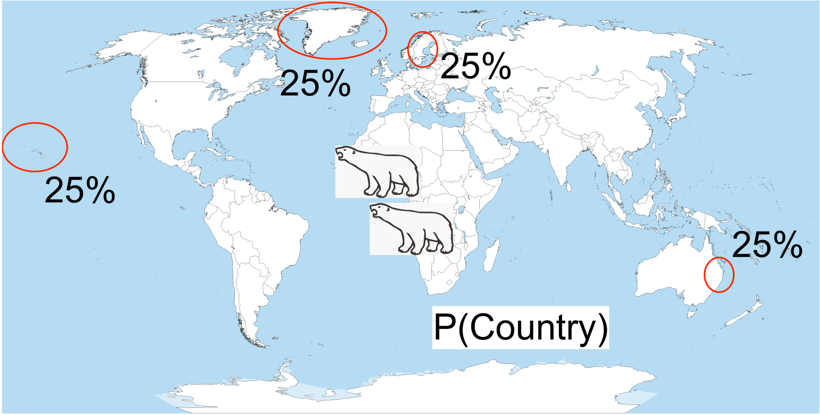
Polar bears and null-hypothesis testing



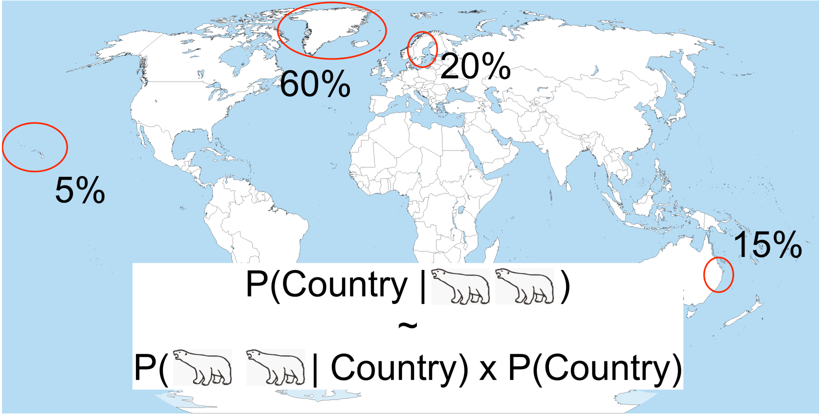
Polar bears and Bayes



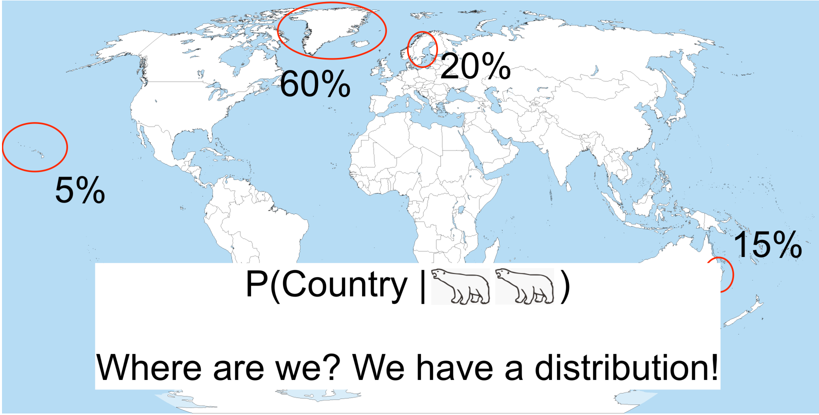
Polar bears and Bayes



Polar bears and Bayes



Polar bears and Bayes



Why does the distinction matter?

- Baysian inference is about forward probabilities (predictions)
- Allows you to calculate any probabilities that you care about:
  - Probability that the effect is greater than 0
  - Probability that the effect is between -0.5 and 0.5
  - Probability that the effect is greater than current practice
- Does not care about data that you could have collected
- Null-hypothesis is about how extreme your data are conditional on a particular effect size
  - Requires thinking about data that you could have collected
  - P-value only concerns one specific effect size

Bayesian sequential design

- Bayesian inference opens up a new world of design choices
- Combine superiority and non-inferiority in the same trial
- No fixed sample size or allocation ratio between arms
- Drop arms that are ineffective or harmful
- You still just calculate $P(Effect|D)$
- No more power calculation *voodoo*!

Bayesian sequential design

> *(In a Bayesian analysis) It is entirely appropriate to collect data until a point has been proven or disproven, or until the data collector runs out of time, money, or patience.*

— Edwards, Lindman, Savage (Psychological Review, **1963**)

Case study: Digital alcohol intervention - Null-hypothesis testing

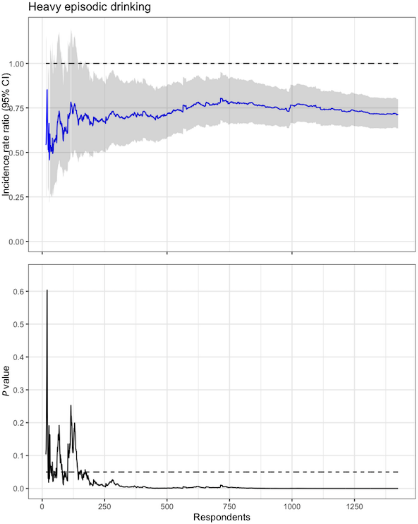
Effectiveness of a digital intervention versus alcohol information for online help-seekers in Sweden: a randomised controlled trial

Marcus Bendtsen, Katarina Åsberg & Jim McCambridge

BMC Medicine **20**, Article number: 176 (2022) | Cite this article

Minimal relevant effect … intervention group consuming 15% less alcohol per week at the 4-month follow-up in comparison to the control group. We aimed for an expected power of 80% at the 0.05 significance threshold. … attrition rate between 5 and 25% … expected sample size of 2126 individuals.

Case study: Digital alcohol intervention - Null-hypothesis testing



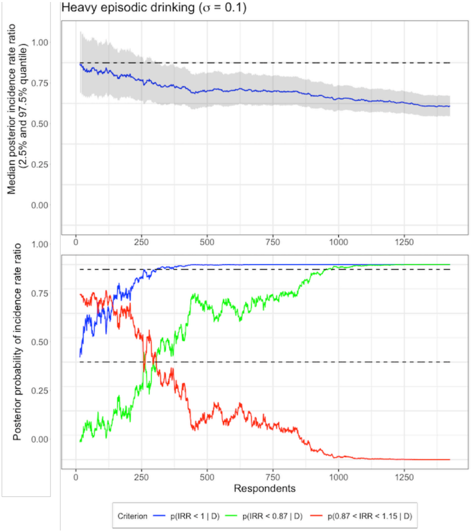
Case study: Digital alcohol intervention - Bayesian sequential design

Avoiding Under- and Overrecruitment in Behavioral Intervention Trials Using Bayesian Sequential Designs: Tutorial

Marcus Bendtsen[1]

- Define target criteria for stopping recruitment
- Effectiveness:
  - $P(IRR < 1|D) > 97.5\%$
  - $P(IRR < 0.87|D) > 50\%$
- Futility: $P(0.87 < IRR < 1.15|D) > 97.5\%$

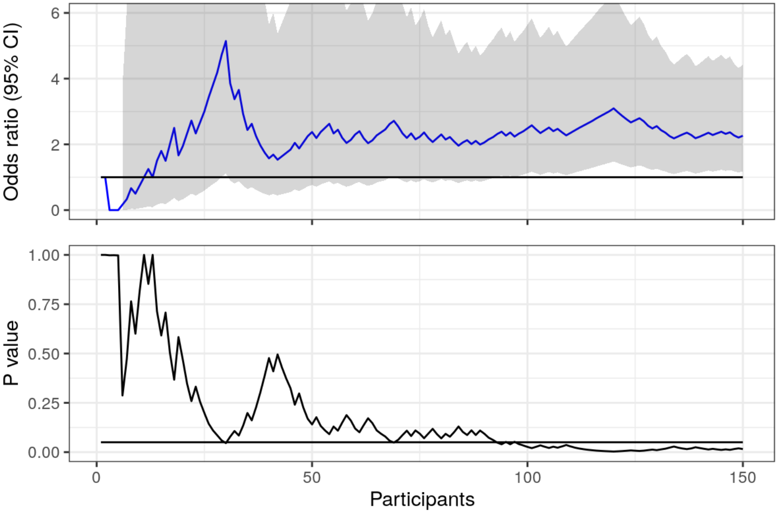
Case study: Digital alcohol intervention - Bayesian sequential design



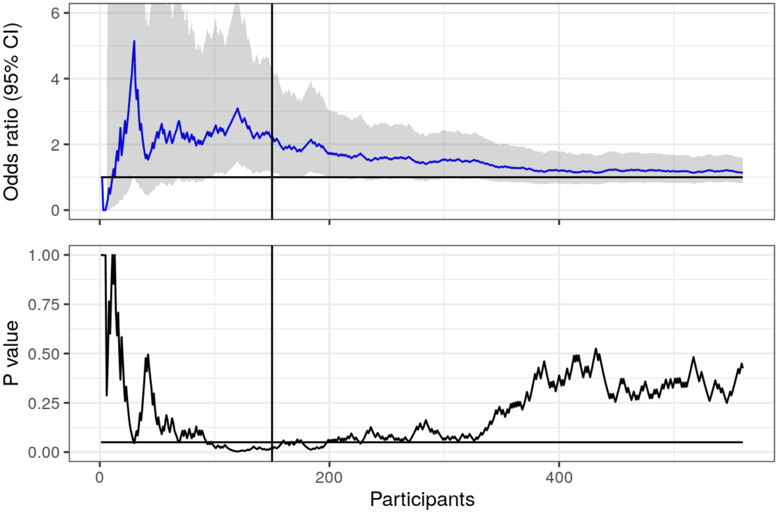
Public Health vs Industry

- **Public health arm**: Public health message regarding alcohol, violence, and cancer.
- **Industry arm**: Information about alcohol, violence, and cancer worded in an alcohol industry manner, focusing on responsible drinking and downplaying the evidence on the risks of alcohol.
- At the end of both text messages was a hyperlink, which lead to more information about alcohol and health
- Experiment outcome was whether or not participants pressed the hyperlink

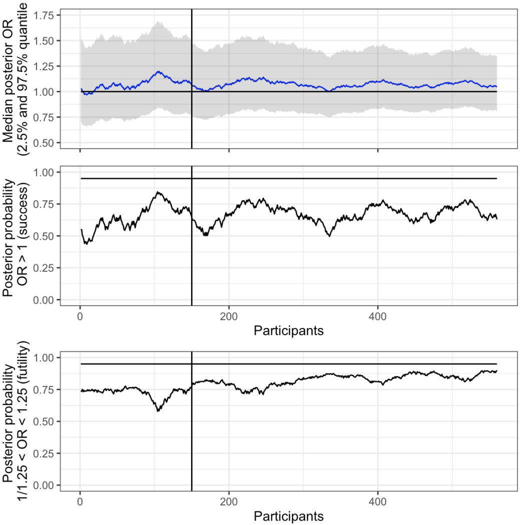
Public Health vs Industry - Null-hypothesis testing



Public Health vs Industry - Null-hypothesis testing



Public Health vs Industry - Bayesian inference



Bayesian sequential designs

- Define a set of target criteria, for instance effect, harm, and futility:
  - Effect: $P(Effect > 0|D) > 95\%$ and $P(Effect > \delta|D) > 50\%$
  - Harm: $P(Effect < 0|D) > 95\%$ and $P(Effect < -\delta|D) > 50\%$
  - Futility: $P(-\delta < Effect < \delta|D) > 95\%$
- Continuously estimate the posterior distribution over effects and monitor the criteria when data becomes available
- Stop recruitment when criteria fullfilled or *until the data collector runs out of time, money, or patience*
- Do analysis on all of data collected at end of follow-up

Reading

Bayesian sequential designs

- D Berry. Bayesian statistics and the efficiency and ethics of clinical trials - *Statistical Science, 2004.*
- D Berry. Bayesian clinical trial - *Nature Reviews Drug Discovery, 2006.*
- G Sponer. A practical guide to Bayesian group sequential designs - *Pharmaceutical Statistics, 2013.*
- M Bendtsen. Avoiding under- and overrecruitment in behavioral intervention trials using Bayesian sequential designs: tutorial - *JMIR, 2022.*
- M Bendtsen. The p value line dance: when does the music stop? - *JMIR, 2022.*
- Prof. Frank Harrell https://www.fharrell.com/ - Personal website with vast knowledge on biostatistics and design.

Examples of trials using Bayesian sequential designs

- J Blomqvist. Effects of a text messaging smoking cessation intervention amongst online help-seekers and primary health care visitors: findings from a randomised controlled trial - *BMC Medicine, 2023.*
- K Åsberg. Digital multiple health behaviour change intervention targeting online help seekers: protocol for the COACH randomised factorial trial - *BMJ Open, 2022.*
- J Crawford. Effects of a drinking motives and readiness to change tailored digital alcohol intervention among online help-seekers: protocol for a randomised controlled trial - *BMJ Open, 2025.*
- (Bayesian but not sequential) C Lilliecreutz. SPARK: an mHealth intervention for self-management and treatment of gestational diabetes mellitus in Sweden – protocol for a randomised controlled trial - *BMJ Open, 2025.*