Quantitative approaches to combining evidence across evidence streams

Stijn Vansteelandt

Ghent University, Belgium
London School of Hygiene and Tropical Medicine, U.K.
joint work with Vo Tat Thang

Outline

Outline

Introduction

- In assessing risks to human health
  from exposure to chemical substances in the environment, relevant evidence may come
  - from both randomised and observational studies;
  - from both animal and human research.
- *How to synthesise evidence across these studies?*

Exposure to trihalomethanes and low birth weight (1)

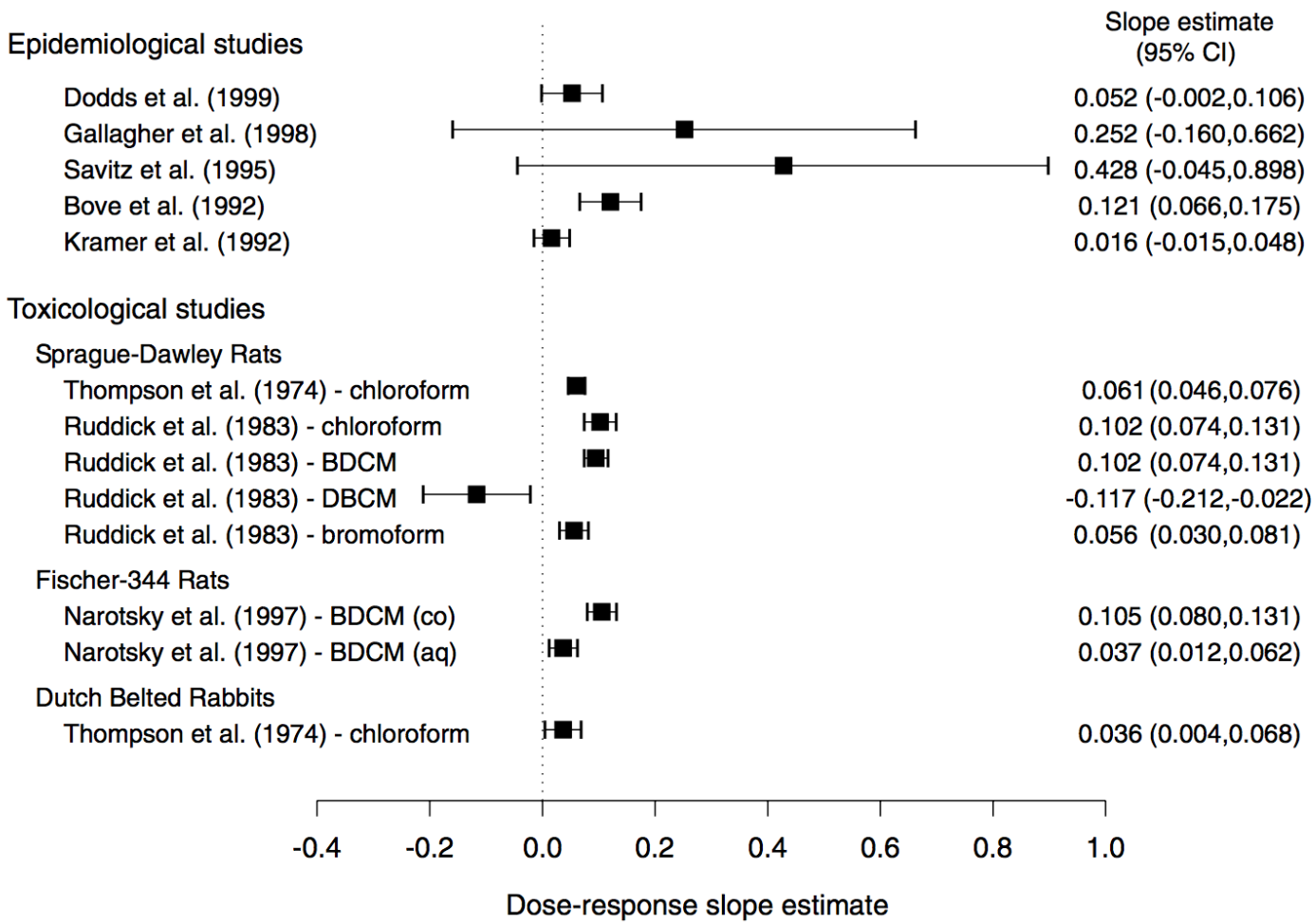


Fig. 1. Study-specific dose–response slope estimates β_i and 95% CIs from ln(OR) *versus* ln(dose) linear model (co, corn oil vehicle; aq, aqueous vehicle)

(Peters et al., Appl Stat 2005)

Meta-analysis

- Meta-analysis approaches synthesise evidence
  across these different types of studies.
- Simple approaches work by averaging the log odds ratios β_i
  from the 13 studies, weighting proportional to their precision.

Meta-analysis

- Meta-analysis approaches synthesise evidence
  across these different types of studies.
- Simple approaches work by averaging the log odds ratios β_i
  from the 13 studies, weighting proportional to their precision.
- More advanced approaches
  acknowledge heterogeneity between different types of studies.
- This can for instance be done
  through a hierarchical Bayesian analysis, assuming that

$$\beta_{ij} \sim N(\psi_{ij}, \sigma^2_{ij})$$

Meta-analysis

- Meta-analysis approaches synthesise evidence
  across these different types of studies.
- Simple approaches work by averaging the log odds ratios β_i
  from the 13 studies, weighting proportional to their precision.
- More advanced approaches
  acknowledge heterogeneity between different types of studies.
- This can for instance be done
  through a hierarchical Bayesian analysis, assuming that

$$\begin{aligned} \beta_{ij} &\sim N(\psi_{ij}, \sigma_{ij}^2) \\ \psi_{ij} &\sim N(\theta_j, \tau_j^2) \end{aligned}$$

Meta-analysis

- Meta-analysis approaches synthesise evidence
  across these different types of studies.
- Simple approaches work by averaging the log odds ratios β_i
  from the 13 studies, weighting proportional to their precision.
- More advanced approaches
  acknowledge heterogeneity between different types of studies.
- This can for instance be done
  through a hierarchical Bayesian analysis, assuming that

$$\begin{aligned} \beta_{ij} &\sim N(\psi_{ij}, \sigma_{ij}^2) \\ \psi_{ij} &\sim N(\theta_j, \tau_j^2) \\ \theta_j &\sim N(\mu, \nu^2) \end{aligned}$$

(Peters et al., Appl Stat 2005)

- Such analysis tends to acknowledge more uncertainty.

Exposure to trihalomethanes and low birth weight (2)

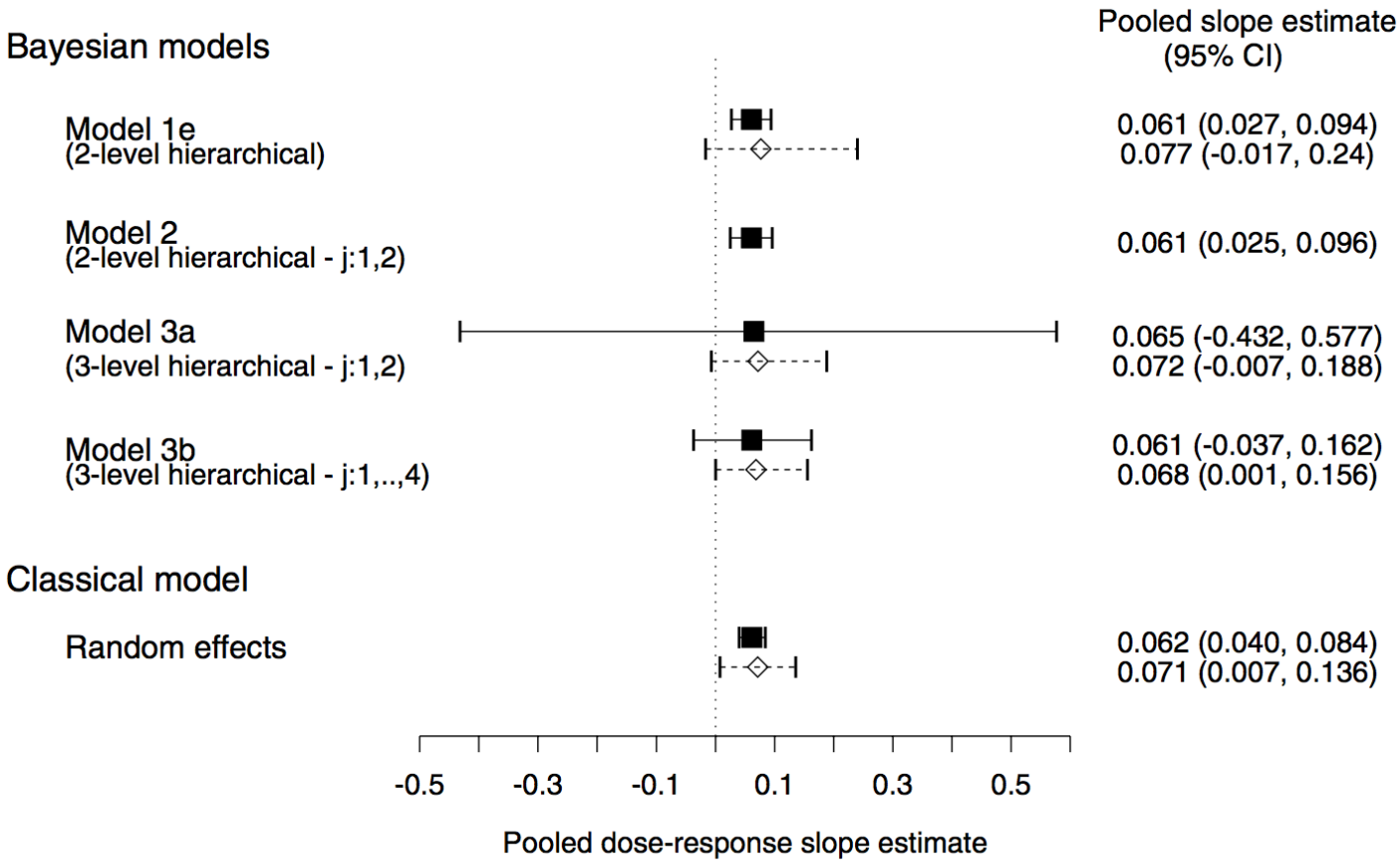


Fig. 2. Pooled dose–response slope estimates μ (and 95% CIs) obtained from the five synthesis models that were used to combine all 13 studies (model 1e, the human epidemiological estimate is μ from model 1a; model 3, the human epidemiological estimate is θ_1): ■, all-species estimate; ◇, human epidemiological estimate

(Peters et al., Appl Stat 2005)

Overview

- In this talk, I will reflect
  on whether we can ‘just’ pool the results
  from these different studies.
- *What population*
  does the summary effect refer to?

Overview

- In this talk, I will reflect
  on whether we can 'just' pool the results
  from these different studies.
- *What population*
  does the summary effect refer to?
- I will do this by first reflecting on
  cross-design synthesis:
  how to synthesise results
  from different study designs in humans?
- This will give insight into the more
  complex problem of how to synthesise
  evidence from human and animal studies.

Outline

Randomised experiments

- Randomised experiments are the gold standard for inferring the effects of exposures on the risk of adverse events.
- The fact that individuals are similar or exchangeable between different exposure groups enables fair comparisons.
- This moreover enables a simple presentation of results, e.g. in terms of the risk of adverse events in each of the exposure groups, possibly in function of time.

Observational studies

Observational studies often study populations of greater interest.

Observational studies

Observational studies often study populations of greater interest.

But exchangeability of exposure groups is not guaranteed.

Issues

- It is therefore common to collect data on pre-exposure characteristics (called confounders).
- This enables one to compare differently exposed individuals who are similar in terms of those characteristics.

Issues

- It is therefore common to collect data
  on pre-exposure characteristics (called confounders).
- This enables one to compare differently exposed individuals
  who are similar in terms of those characteristics.
- Often, we will not know what all confounders are,
  or have no data on all.
- This affects the quality of observational data analyses.
- Sensitivity analyses are therefore important,
  though not (yet) readily applicable.

Issues

- It is therefore common to collect data
  on pre-exposure characteristics (called confounders).
- This enables one to compare differently exposed individuals
  who are similar in terms of those characteristics.
- Often, we will not know what all confounders are,
  or have no data on all.
- This affects the quality of observational data analyses.
- Sensitivity analyses are therefore important,
  though not (yet) readily applicable.
- Approaches that do not demand data on confounders
  are also of interest (e.g. instrumental variables analyses).

Randomised experiments versus observational studies

- As we start to analyse observational studies, we typically report different effects measures than we would in randomised experiments.
- E.g. we tend to report the risk of adverse events for exposed and unexposed in randomised experiments, but an odds/hazard ratio in observational studies.
- *This renders interpretation more complicated.*

Randomised experiments versus observational studies

- As we start to analyse observational studies, we typically report different effects measures than we would in randomised experiments.
- E.g. we tend to report the risk of adverse events for exposed and unexposed in randomised experiments, but an odds/hazard ratio in observational studies.
- *This renders interpretation more complicated.*
- E.g. we report population-level effects in randomised experiments, but subgroup effects in observational studies.
- *Can we simply pool these different effects?*

Population-level versus subgroup effects...

... can be difficult to pool as a result of non-collapsibility, a 'dilution' effect.

(Greenland, Robins and Pearl, Stat Sci 1998)

ALL ADULTS (n=2000)

Heart Disease

Vitamin D Level	Yes	No
Low	600	400
High	400	600

Odds ratio =2.25 (1.88, 2.69)

Risk ratio =1.50 (1.37, 1.64)

SEDENTARY ADULTS (n=1000)

Heart Disease

Vitamin D Level	Yes	No
Low	400	100
High	300	200

Odds ratio =2.67 (2.01, 3.54)

Risk ratio =1.33 (1.23, 1.45)

ACTIVE ADULTS (n=1000)

Heart Disease

Vitamin D Level	Yes	No
Low	200	300
High	100	400

Odds ratio =2.67 (2.01, 3.54)

Risk ratio =2.00 (1.63, 2.46)

Issues

- This is problematic.

Issues

- This is problematic.
- It hinders a good appreciation of the public health impact of certain exposures in terms of odds ratios, hazard ratios, ...
  - *Do we really understand how hazardous is an exposure with an odds ratio or hazard ratio of 1.2?*
  - 'One day scientists will look back and wonder why statisticians/epidemiologists spent decades reporting hazard ratios and not absolute risks'

    (Miguel Hernan, Twitter 22/10/17)

Issues

- This is problematic.
- It hinders a good appreciation of the public health impact of certain exposures in terms of odds ratios, hazard ratios, ...
  - *Do we really understand how hazardous is an exposure with an odds ratio or hazard ratio of 1.2?*
  - 'One day scientists will look back and wonder why statisticians/epidemiologists spent decades reporting hazard ratios and not absolute risks'

    (Miguel Hernan, Twitter 22/10/17)
- It makes standard meta-analysis a difficult exercise.
  - *Do we really understand the summary measure obtained by pooling log odds ratios for different populations with different degree of heterogeneity?*
  - *What is the use of a summary measure, if we don't know which population it refers to?*

Examples

- Consider synthesising the results of 2 randomised experiments,
  one in individuals aged 20-30
  and one in individuals aged 20-60.
  - Even if in both studies, the odds ratio of exposure
    in individuals of the same age is the same,
    population-level odds ratios will tend to differ.
  - *Can we just pool these results?*
  - *If we pool the results,*
    for what population are we then describing the effect?

Examples

- Consider synthesising the results of 2 randomised experiments,
  one in individuals aged 20-30
  and one in individuals aged 20-60.
  - Even if in both studies, the odds ratio of exposure
    in individuals of the same age is the same,
    population-level odds ratios will tend to differ.
  - *Can we just pool these results?*
  - *If we pool the results,*
    for what population are we then describing the effect?
- Consider synthesising the results of one observational study
  and one randomised experiment.

Examples

- Consider synthesising the results of 2 randomised experiments,
  one in individuals aged 20-30
  and one in individuals aged 20-60.
  - Even if in both studies, the odds ratio of exposure
    in individuals of the same age is the same,
    population-level odds ratios will tend to differ.
  - *Can we just pool these results?*
  - *If we pool the results,*
    for what population are we then describing the effect?
- Consider synthesising the results of one observational study
  and one randomised experiment.
  - Observational studies often report adjusted associations,
    which tend to appear 'stronger'.
  - *Can we just pool these results?*

Summary: meta-analysis is subtle business

Summary: meta-analysis is subtle business

- Standard approaches pool the results from different studies, but are silent as to how to interpret the summary result.
- In particular, for what population do they describe the risk of adverse effects?

Summary: meta-analysis is subtle business

- Standard approaches pool the results from different studies, but are silent as to how to interpret the summary result.
- In particular, for what population do they describe the risk of adverse effects?
- More heterogeneous populations often suggest weaker effects as a result of dilution.
- This can make results from different studies, even randomised experiments, difficult to pool when they differ in degree of heterogeneity, or adjust for different variables.

Outline

How to move forward?

How to move forward?

- When synthesising results from different studies,
  it is useful to first agree on the population
  for which we attempt to infer the exposure effect.
- e.g. we may aim to infer what the risk of adverse events for the participants of experiment 1 (aged 20-30) would be
  - if all were exposed;
  - if none were exposed.
- This is well-defined and simple to interpret.

How to move forward?

- When synthesising results from different studies, it is useful to first agree on the population for which we attempt to infer the exposure effect.
- e.g. we may aim to infer what the risk of adverse events for the participants of experiment 1 (aged 20-30) would be
  - if all were exposed;
  - if none were exposed.
- This is well-defined and simple to interpret.
- We may then attempt to use the data from the different studies to evaluate this same effect.
- The results from the different studies can now be pooled.

How to do this?

How to do this?

Suppose our aim is to use the data from experiment 2 to infer what the risk of adverse events for the participants of experiment 1 would be if all were exposed.

- use the data from experiment 2 to build a prediction model for the risk of adverse events in function of exposure, all baseline covariates which capture between-study differences (and extraneous variables);

$$P(Y = 1|X, Z, S = 2) = \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ),$$

where S is the number of the study in which one is included.

How to do this?

Suppose our aim is to use the data from experiment 2 to infer what the risk of adverse events for the participants of experiment 1 would be if all were exposed.

- use the data from experiment 2 to build a prediction model for the risk of adverse events in function of exposure, all baseline covariates which capture between-study differences (and extraneous variables);

$$P(Y = 1|X, Z, S = 2) = \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ),$$

  where S is the number of the study in which one is included.

- use the model to make a prediction for each individual in experiment 1, as if exposed.

$$\text{expit}(\beta_0 + \beta_1 + \beta_2 Z + \beta_3 Z)$$

- average this across all subjects in experiment 1.

Direct standardisation

- This can also be used in observational studies.
- In that case, the prediction model must additionally include all relevant confounders of the exposure - outcome association.
- Readily available using software for direct standardisation: `stdreg` in R or `teffects` in Stata.

Summary: direct standardisation

- Standardisation maps the results from different studies onto the same estimand,
  e.g. the risk for individuals in study 1, if exposed.
- These results can therefore be meaningfully pooled.
- There are also a number of caveats.

Summary: direct standardisation

- Standardisation maps the results from different studies onto the same estimand,
  e.g. the risk for individuals in study 1, if exposed.
- These results can therefore be meaningfully pooled.
- There are also a number of caveats.
- Standardisation requires individual data.
- In particular, on all prognostic factors of outcome, which are differentially distributed between studies.
- All such characteristics can be difficult to find, especially when combining animal and human studies.

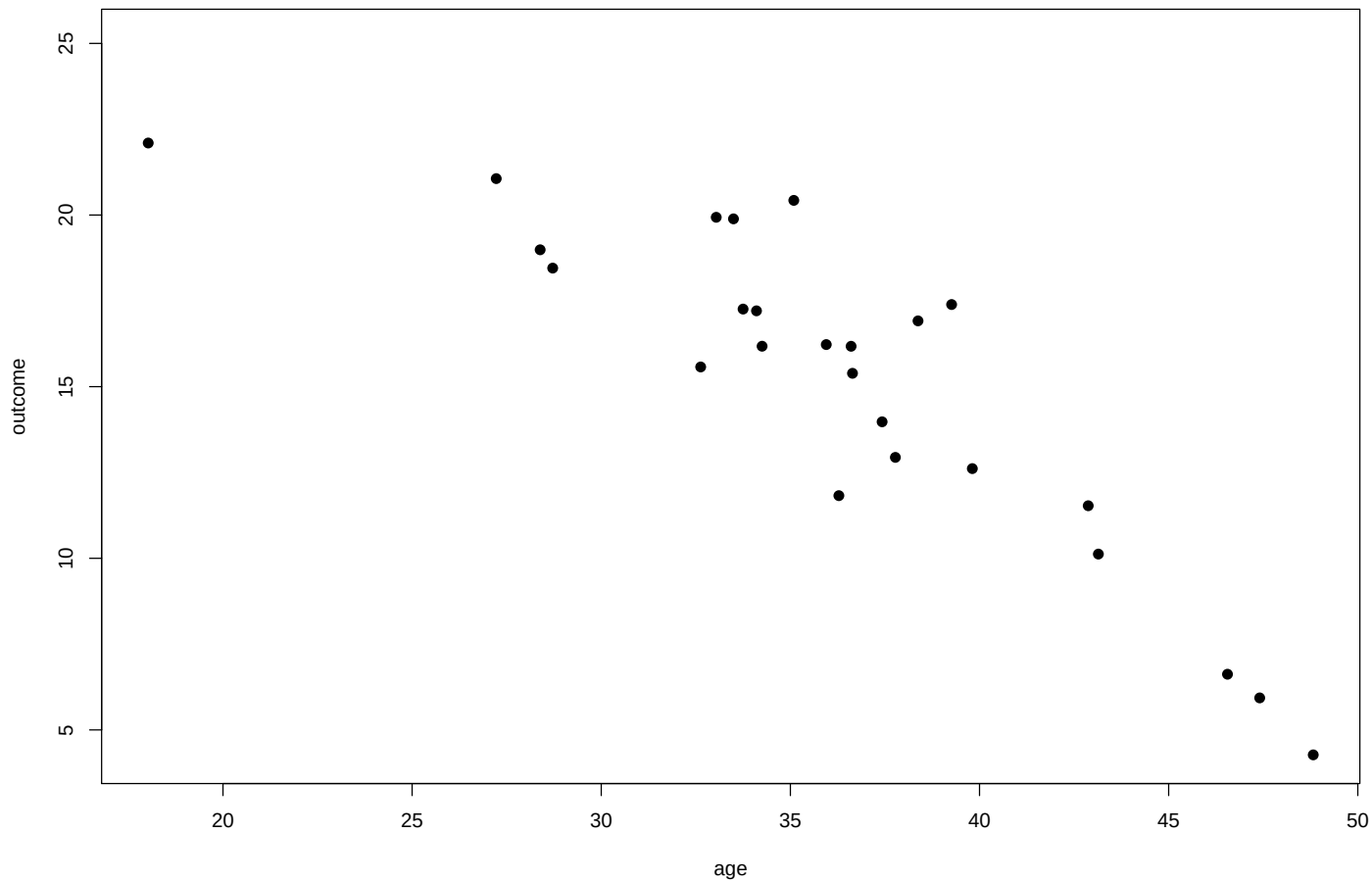
Summary: direct standardisation

- Standardisation maps the results from different studies onto the same estimand,
  e.g. the risk for individuals in study 1, if exposed.
- These results can therefore be meaningfully pooled.
- There are also a number of caveats.
- Standardisation requires individual data.
- In particular, on all prognostic factors of outcome, which are differentially distributed between studies.
- All such characteristics can be difficult to find, especially when combining animal and human studies.
- Even when they can all be measured, there is a danger of extrapolation...

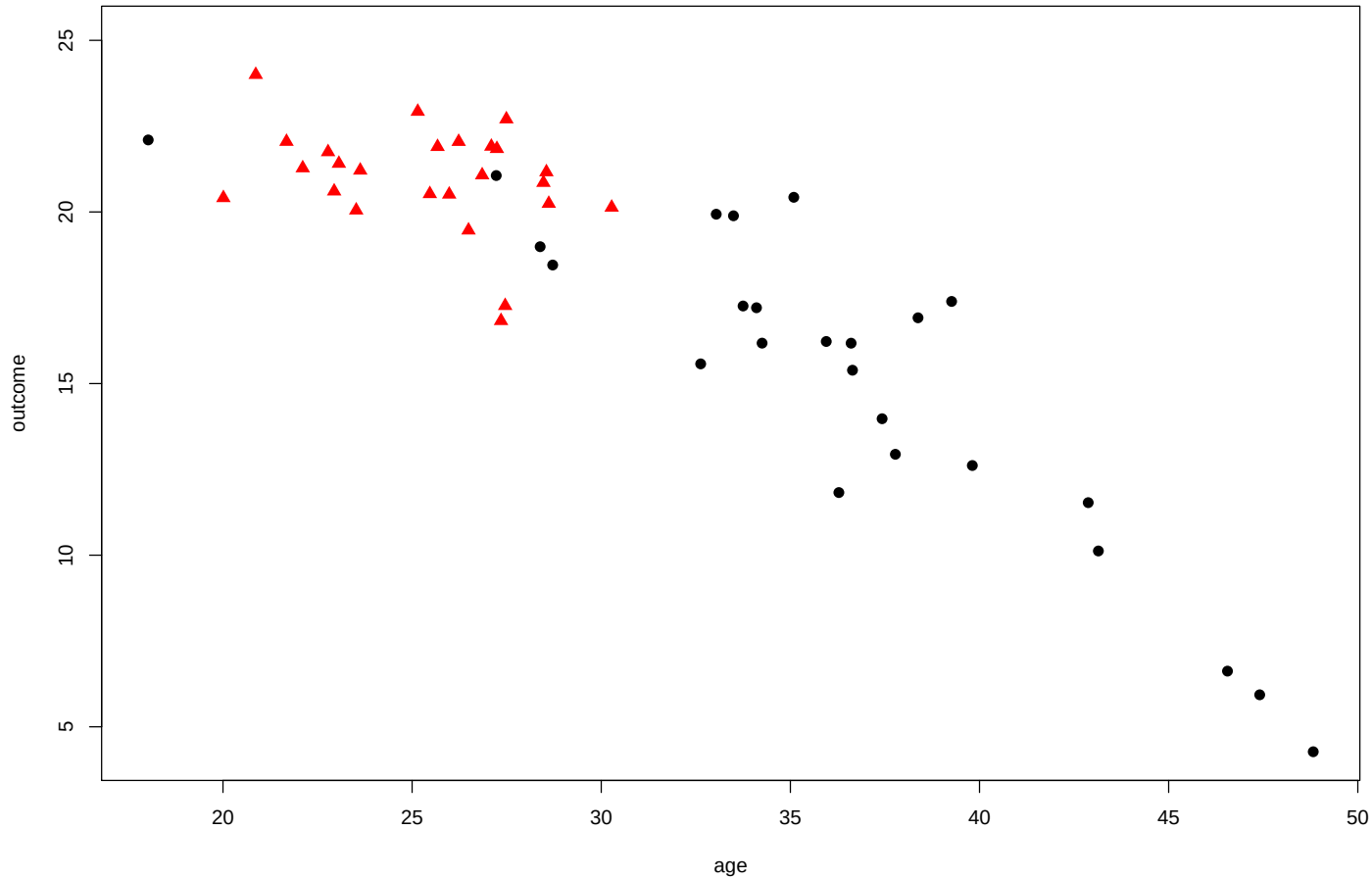
Outline

Ill overlapping experiments (1)

Consider data for the 'exposed' in 2 randomised experiments.

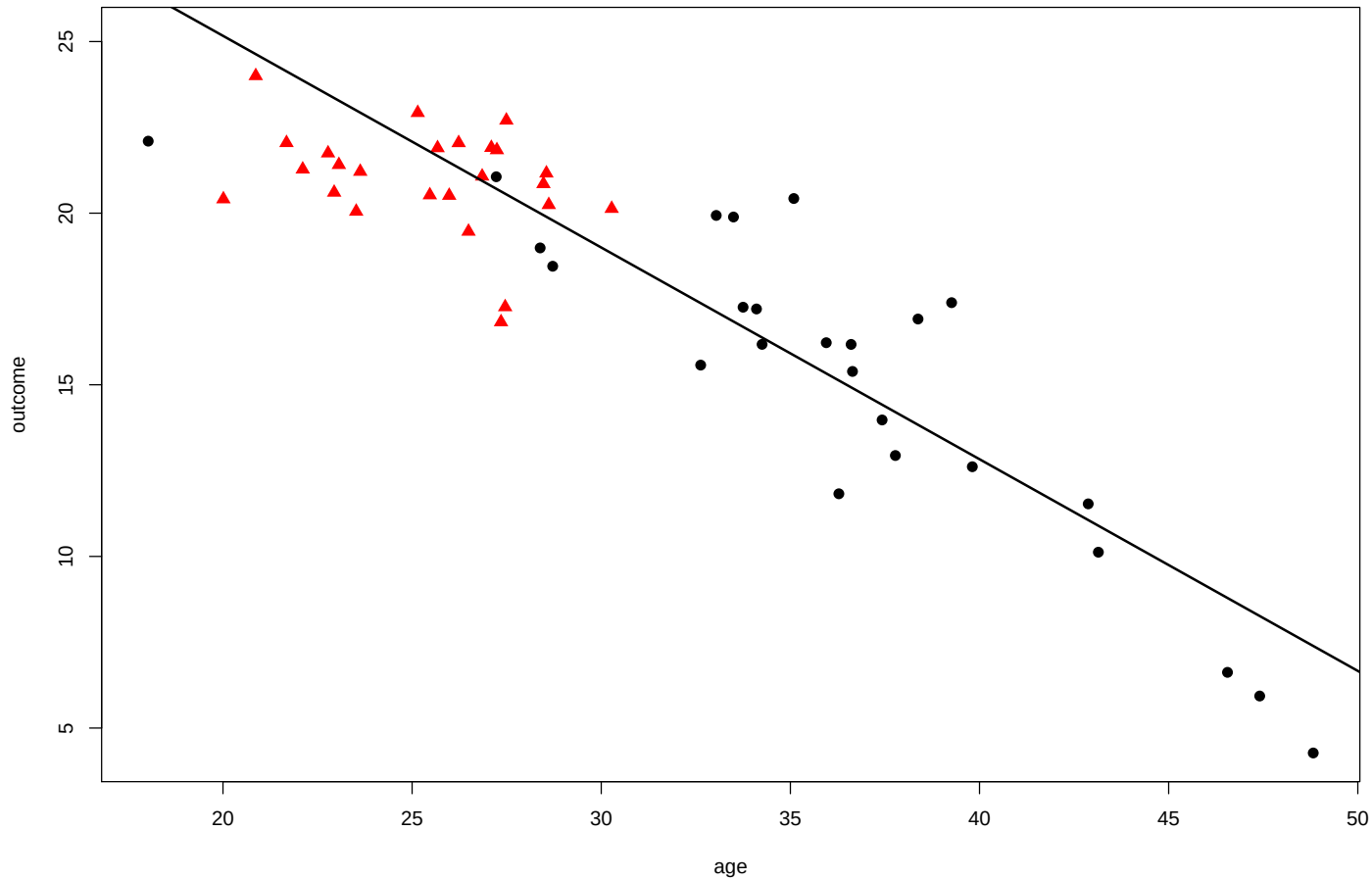


Ill overlapping experiments (2)



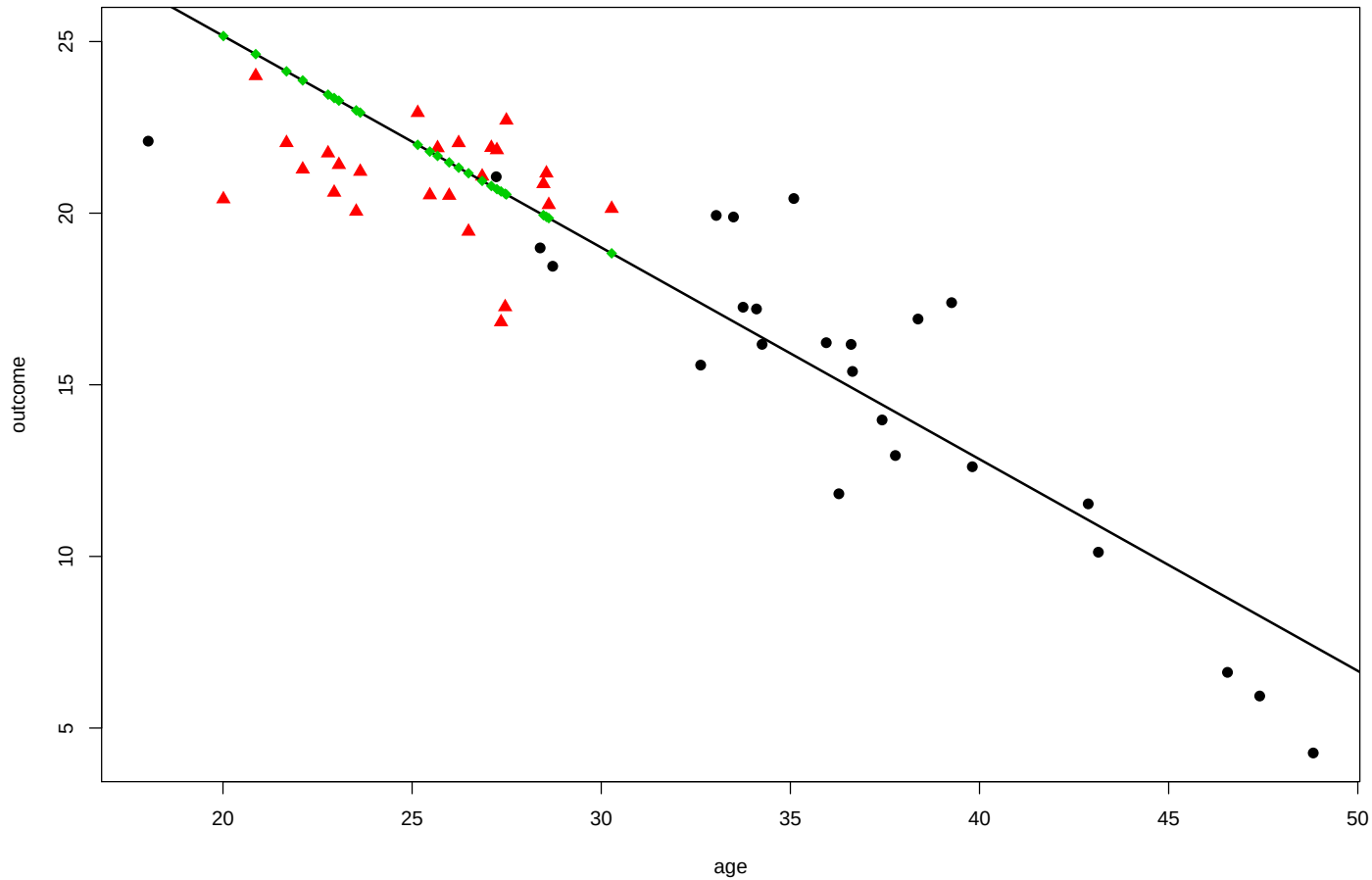
The data from experiment 1 carry no information about the effect for participants of experiment 2. We will therefore only transport the information from experiment 2 to participants of experiment 1.

Linear standardisation (1)



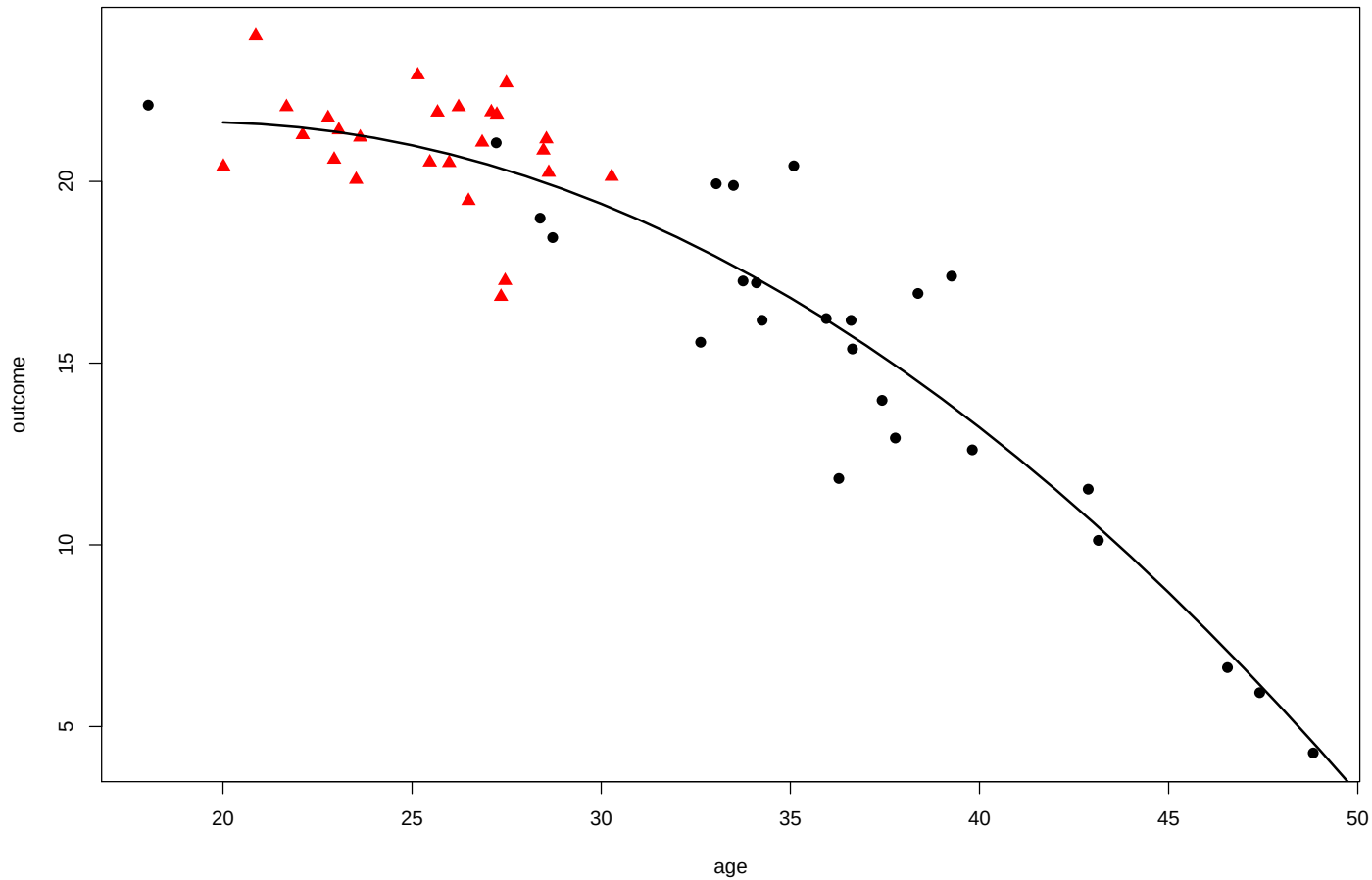
A linear model extrapolates, resulting in bias.

Linear standardisation (2)



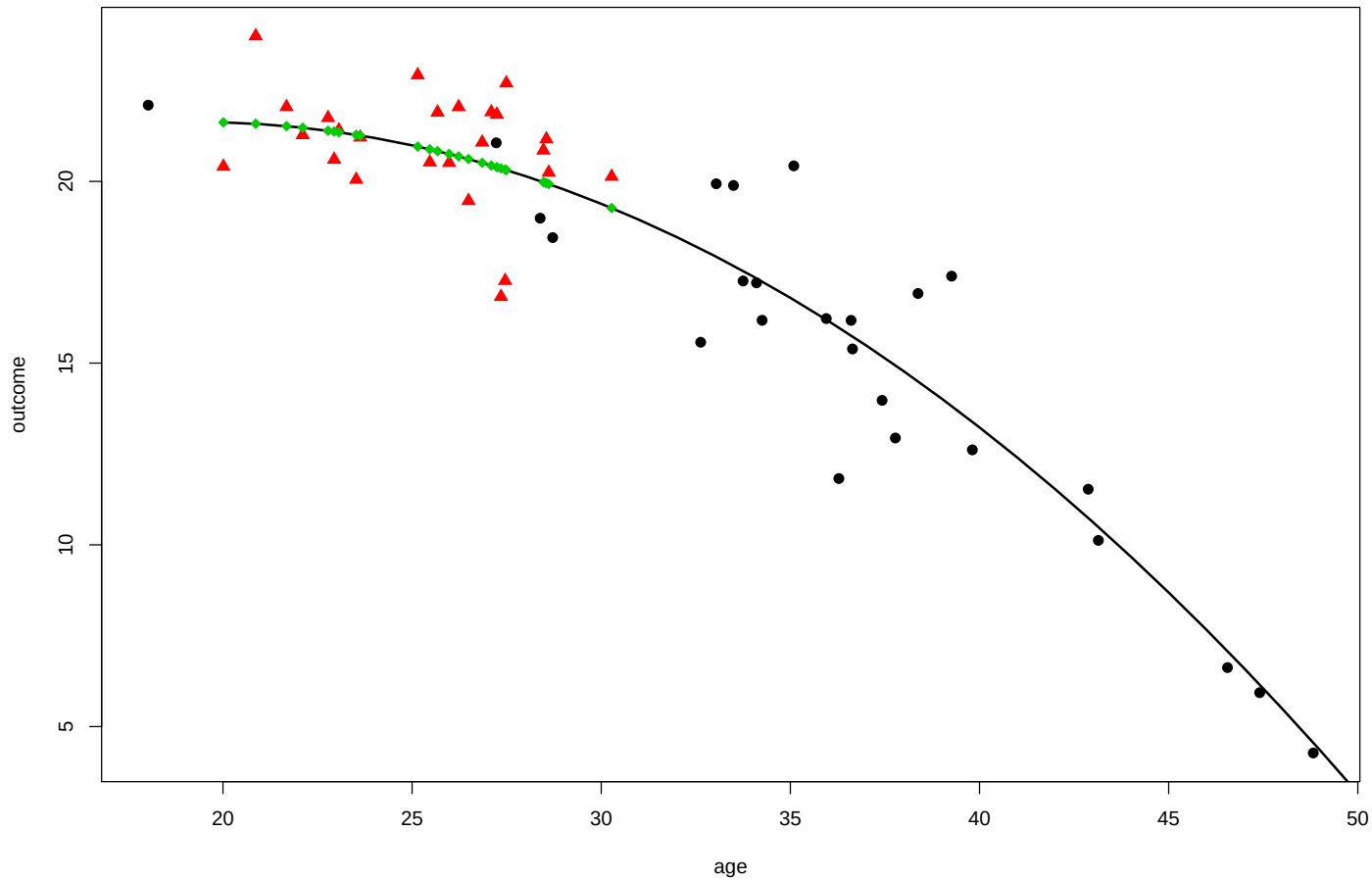
A linear model extrapolates, resulting in bias.

Quadratic standardisation (1)



A quadratic model is hard to distinguish from a linear model.

Quadratic standardisation (2)



A quadratic model is hard to distinguish from a linear model.

Problems of regression adjustment (1)

- This problem of model misspecification and extrapolation is especially severe when there is little overlap between studies.

  (Rubin, *Ann Int Med* 97; Tan, *Stat Science* 08)

- This is because,
  to transport the results from one study to the other,
  we can only learn from subjects in different studies,
  with the same measured characteristics.

Problems of regression adjustment (1)

- This problem of model misspecification and extrapolation is especially severe when there is little overlap between studies.

  (Rubin, *Ann Int Med* 97; Tan, *Stat Science* 08)

- This is because,
  to transport the results from one study to the other,
  we can only learn from subjects in different studies,
  with the same measured characteristics.
- When it is difficult to find such subjects,
  this involves extrapolation.
- Even models that fit the observed data well,
  may then yield severe bias.

Problems of regression adjustment (2)

- This concern is ignored
  by the previous standardisation approach.
- It assumes the outcome model is correct,
  hence no extrapolation.
- It would even allow to transport the information from
  experiment 1 to participants of experiment 2,
  while giving apparently good results.

Problems of regression adjustment (2)

- This concern is ignored
  by the previous standardisation approach.
- It assumes the outcome model is correct,
  hence no extrapolation.
- It would even allow to transport the information from
  experiment 1 to participants of experiment 2,
  while giving apparently good results.
- This concern is also ignored by current approaches,
  which pool results regardless of the similarity of subjects
  between studies.

Propensity scores

- This is why propensity score methods are useful.

  (Rosenbaum and Rubin, *Bka* 83)

- Suppose our aim is again to use the data from experiment 2 to infer what the risk of adverse events for the participants of experiment 1 would be if all were exposed.

Propensity scores

- This is why propensity score methods are useful.

  (Rosenbaum and Rubin, *Bka* 83)

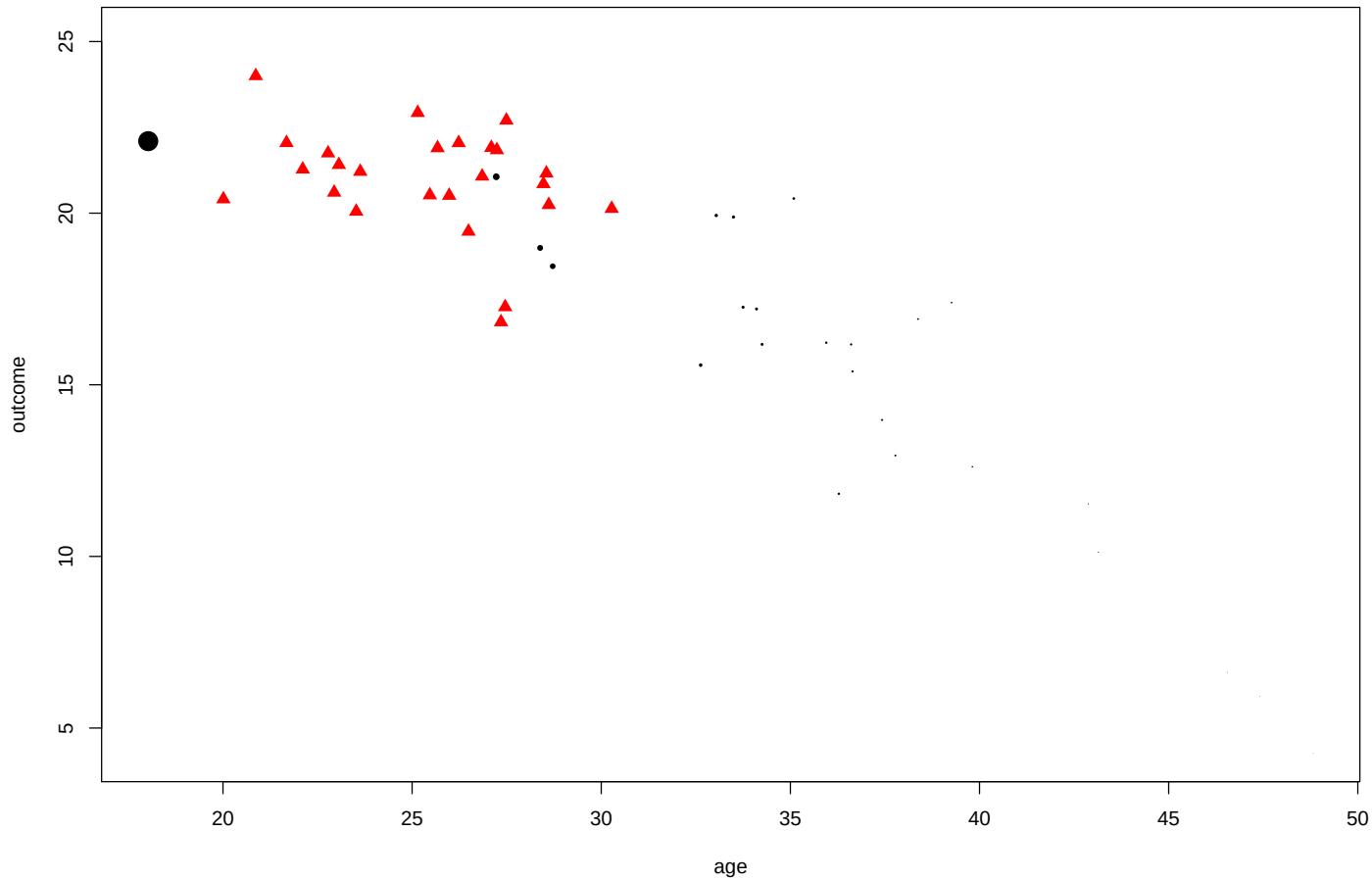
- Suppose our aim is again to use the data from experiment 2 to infer what the risk of adverse events for the participants of experiment 1 would be if all were exposed.
- Then we use the previous standardisation technique, except that the prediction model for the risk of adverse events must be a canonical GLM, fitted with weights

$$\frac{P(S=1|X,Z)}{P(S=2|X,Z)}$$

- Here, the probability to belong to study 1, $P(S=1|X,Z)$, can be calculated under some prediction model, e.g. multinomial regression.
- This approach has been called double-robust standardisation.

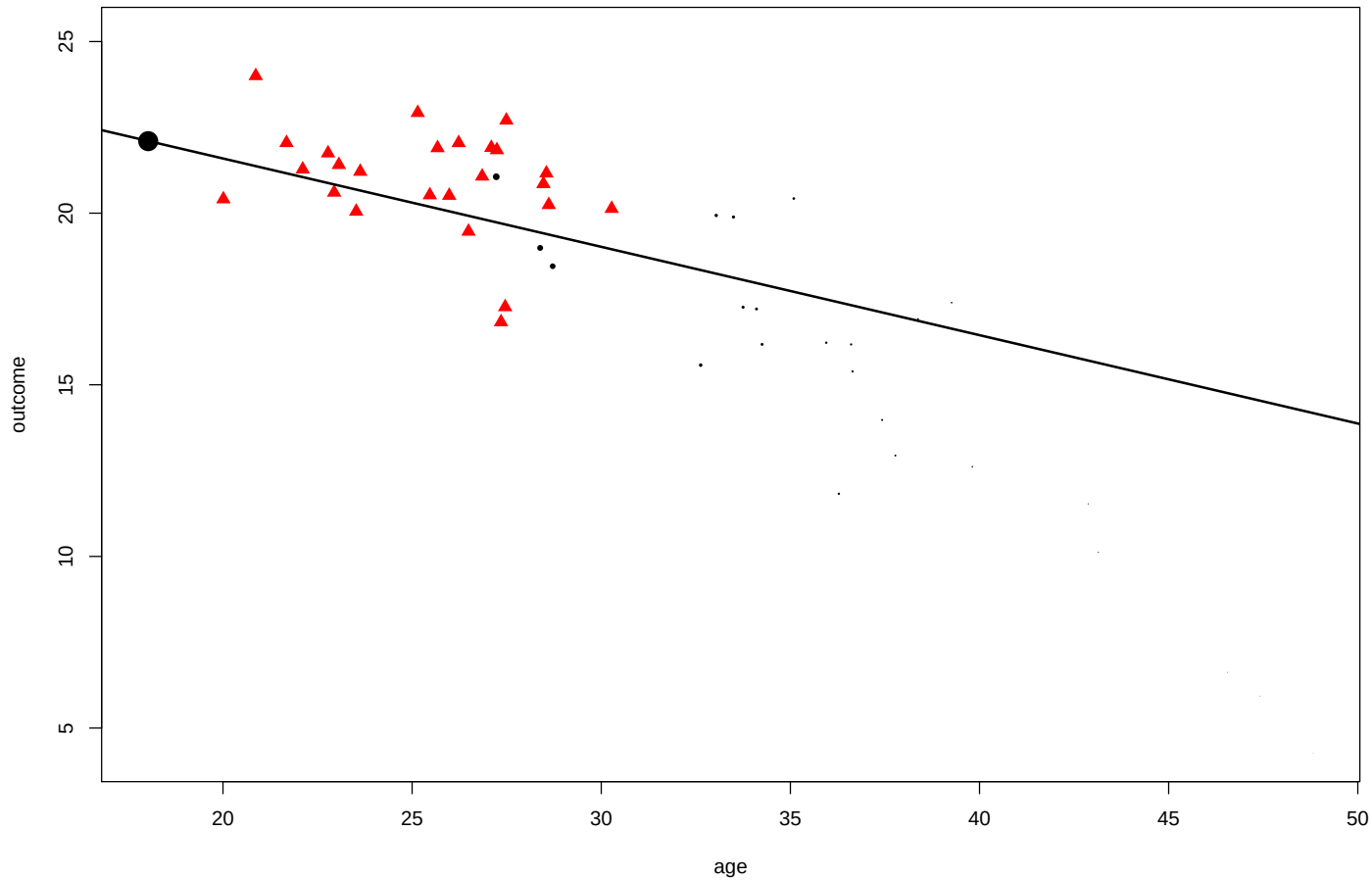
  (Vansteelandt and Keiding, 2012)

Double robust standardisation (1)



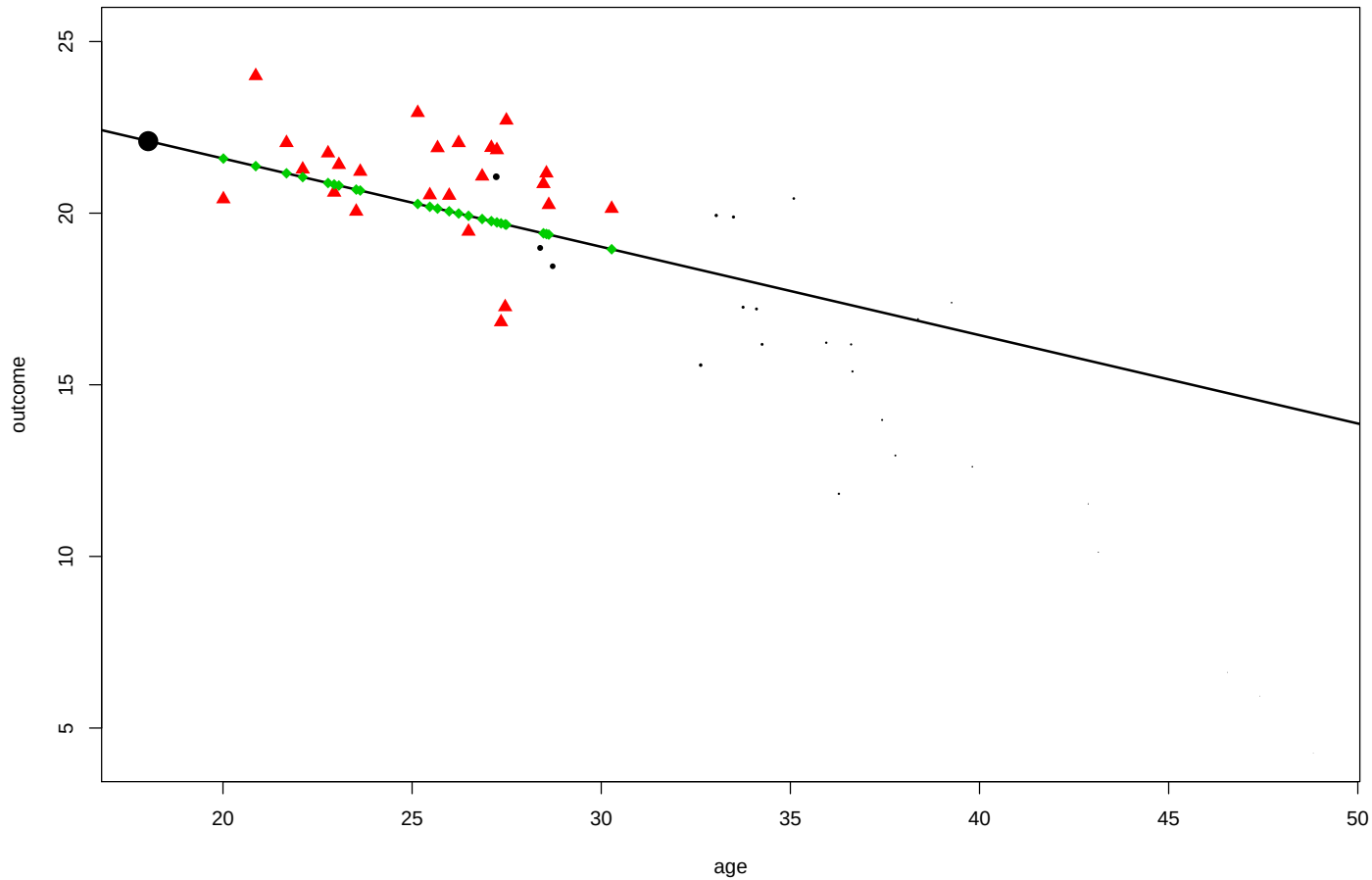
Weights accentuate where subjects from study 1 are.

Double robust standardisation (2)



Weights accentuate where subjects from study 1 are.

Double robust standardisation (3)



Weights accentuate where subjects from study 1 are.

Properties of double robust standardisation

- Resulting estimator is valid,
  even when the outcome prediction model is wrong.
- This approach thus avoids extrapolation
  by not relying on outcome regression.

Properties of double robust standardisation

- Resulting estimator is valid,
  even when the outcome prediction model is wrong.
- This approach thus avoids extrapolation
  by not relying on outcome regression.
- It typically results in larger standard errors,
  thus more honestly reflecting the limited information.

Synthesising data from human and animal studies (1)

- The concern for extrapolation becomes even more pronounced when synthesising data from human and animal studies.
- It requires
  the availability of characteristics for both animals and humans, such that animals and humans with the same characteristics have the same risk of adverse events.

Synthesising data from human and animal studies (1)

- The concern for extrapolation becomes even more pronounced when synthesising data from human and animal studies.
- It requires
  the availability of characteristics for both animals and humans, such that animals and humans with the same characteristics have the same risk of adverse events.
- One may alternatively rely on outcome regression,
  and assume that the dose-response odds ratio is transportable between animals and humans.
- This is more in line with current approaches.
- However, it is also a dangerous undertaking
  in view of effect modification and non-collapsibility.
- Safer may be to use the animal data
  only to inform a Bayesian prior, or to use weights of evidence.

Summary

- Pooling results from animal and human studies seems dangerous business.
- Existing approaches acknowledge heterogeneity between studies, but
  - it is unclear what they infer;
  - they ignore the dangers of extrapolation when transporting results from one study to another.

(Bareinboim and Pearl; Cole; Hernan; Stuart; ...)

- Regardless of the approach taken, there is a need for being more explicit what estimand is inferred and what assumptions are made when synthesising results from these different studies.