# Conducting trial-based cost-effectiveness analyses:

## A tutorial

**Andrea Gabrio**

Department of Methodology and Statistics, FHML (UM)

a.gabrio@maastrichtuniversity.nl

https://github.com/AnGabrio

https://angabrio.github.io


HSR department seminar, 25 Oct 2025 - Maastricht

# Table of contents I



Maastricht University

Section 1

# Introduction & Modelling in HTA

# Something about me

- Phd in Statistics @UCL (University College London) with thesis on stat methods to handle missing data in HTA

Maastricht University

# Something about me

- Phd in Statistics @UCL (University College London) with thesis on stat methods to handle missing data in HTA
- Experience as trial stat/health economist @PRIMENT CTU (UCL)

Maastricht University

# Something about me

- Phd in Statistics @UCL (University College London) with thesis on stat methods to handle missing data in HTA
- Experience as trial stat/health economist @PRIMENT CTU (UCL)
- Collaborated on methodological/applied HTA projects with collegaues in both UK and NL
  - Prof. Baio, Gomes, Leurent (Stat Sci @UCL) & Prof. Manca (Centre for HE @York)
  - Prof. Evers, Hiligsmann, Mastrigt (HSR @UM) & Prof. Joore (@KEMTA)
  - Prof. Bosmans (Health Sciences @VU) & El Alili (@ZIN)

Maastricht University

# Something about me

- Phd in Statistics @UCL (University College London) with thesis on stat methods to handle missing data in HTA
- Experience as trial stat/health economist @PRIMENT CTU (UCL)
- Collaborated on methodological/applied HTA projects with collegaues in both UK and NL
  - Prof. Baio, Gomes, Leurent (Stat Sci @UCL) & Prof. Manca (Centre for HE @York)
  - Prof. Evers, Hiligsmann, Mastrigt (HSR @UM) & Prof. Joore (@KEMTA)
  - Prof. Bosmans (Health Sciences @VU) & El Alili (@ZIN)
- **Personal goal** to facilitate uptake of methods in HTA via software:
  - I am a bit of a `R` enthusiast
  - I am a bit of a `Excel` hater (for statistical analyses)

Maastricht University

# Disclaimer ...

Before I begin:

- **Statistics** is an all-encompassing discipline: *Statisticians are unified not by the subject matter they work in, but the methodology used to address problems that arise in diverse fields*

# Disclaimer ...

Before I begin:

- **Statistics** is an all-encompassing discipline: *Statisticians are unified not by the subject matter they work in, but the methodology used to address problems that arise in diverse fields*

- The **success** of Statistics relies on developing skills to make it easier for colleagues in other disciplines to appreciate our contribution

# Disclaimer ...

Before I begin:

- **Statistics** is an all-encompassing discipline: *Statisticians are unified not by the subject matter they work in, but the methodology used to address problems that arise in diverse fields*

- The **success** of Statistics relies on developing skills to make it easier for colleagues in other disciplines to appreciate our contribution

- My **personal** view of the world:
  *Statisticians should be in charge of everything.*

Maastricht University

# Disclaimer ...

Before I begin:

- **Statistics** is an all-encompassing discipline: *Statisticians are unified not by the subject matter they work in, but the methodology used to address problems that arise in diverse fields*
- The **success** of Statistics relies on developing skills to make it easier for colleagues in other disciplines to appreciate our contribution
- My **personal** view of the world:
  *Statisticians should be in charge of everything.*
- So I probably will be very annoying throughout the presentation[1]

[1] But luckily no non-Statistician has been harmed in the making of these slides

# Health technology assessment (HTA)

- **Objective**: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

# Health technology assessment (HTA)

- **Objective**: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

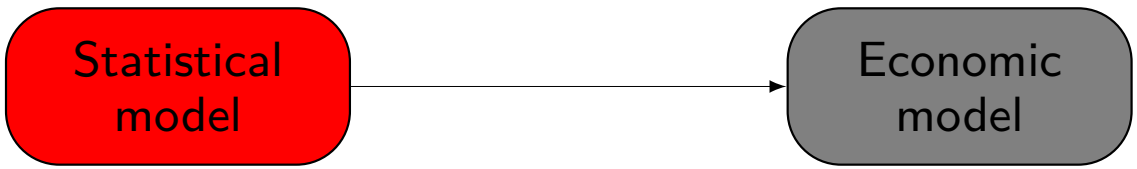
Statistical model

- Estimates relevant parameters $\theta$
- Varies with the type of data (& statistical approach!)

# Health technology assessment (HTA)

- **Objective**: Combine costs & benefits of a given intervention into a rational scheme for allocating resources

$$\Delta_e = f_e(\theta)$$
$$\Delta_c = f_c(\theta)$$



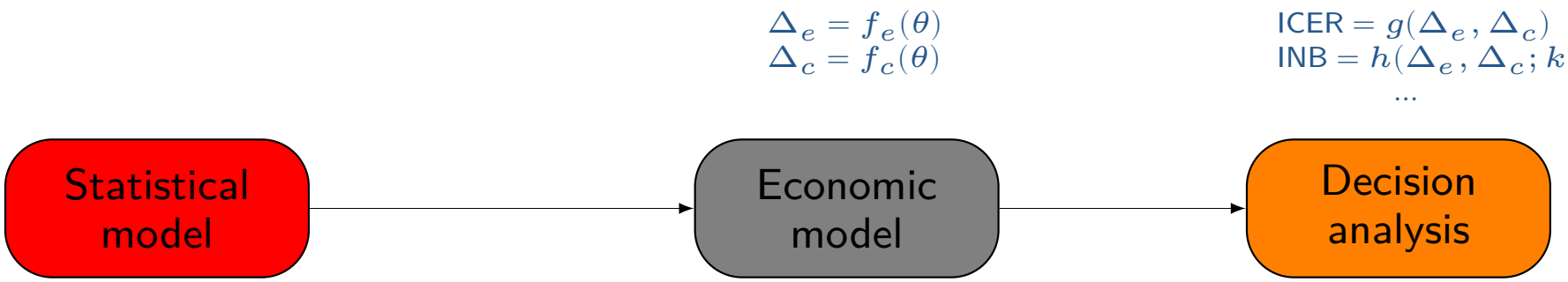

- Estimates relevant parameters $\theta$
- Varies with the type of data (& statistical approach!)

- Combines parameters to obtain average measure for costs and effects
- Varies with the type of data & statistical model used

Maastricht University

# Health technology assessment (HTA)

- **Objective**: Combine costs & benefits of a given intervention into a rational scheme for allocating resources



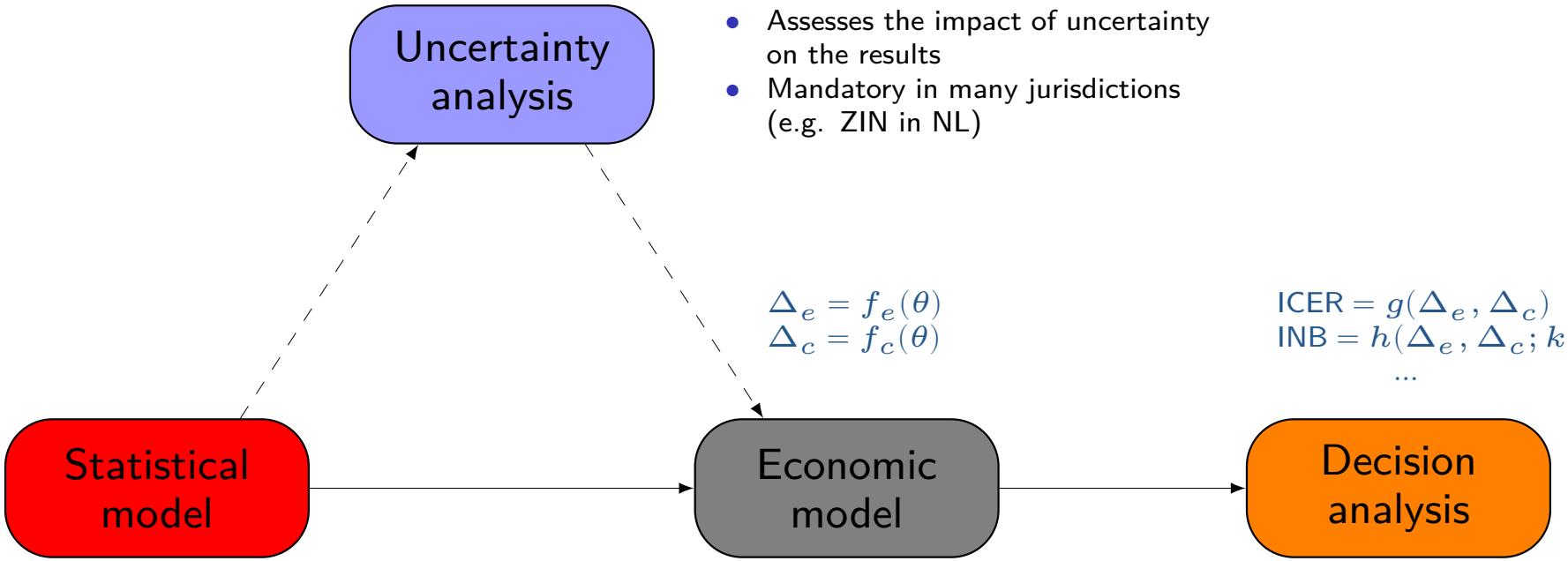

- Estimates relevant parameters $\theta$
- Varies with the type of data (& statistical approach!)

- Combines parameters to obtain average measure for costs and effects
- Varies with the type of data & statistical model used

- Computes suitable measures of "cost-effectiveness"
- Dictates the best course of actions
- Standardised process

# Health technology assessment (HTA)

- **Objective**: Combine costs & benefits of a given intervention into a rational scheme for allocating resources




- Assesses the impact of uncertainty on the results
- Mandatory in many jurisdictions (e.g. ZIN in NL)

- Estimates relevant parameters $\theta$
- Varies with the type of data (& statistical approach!)

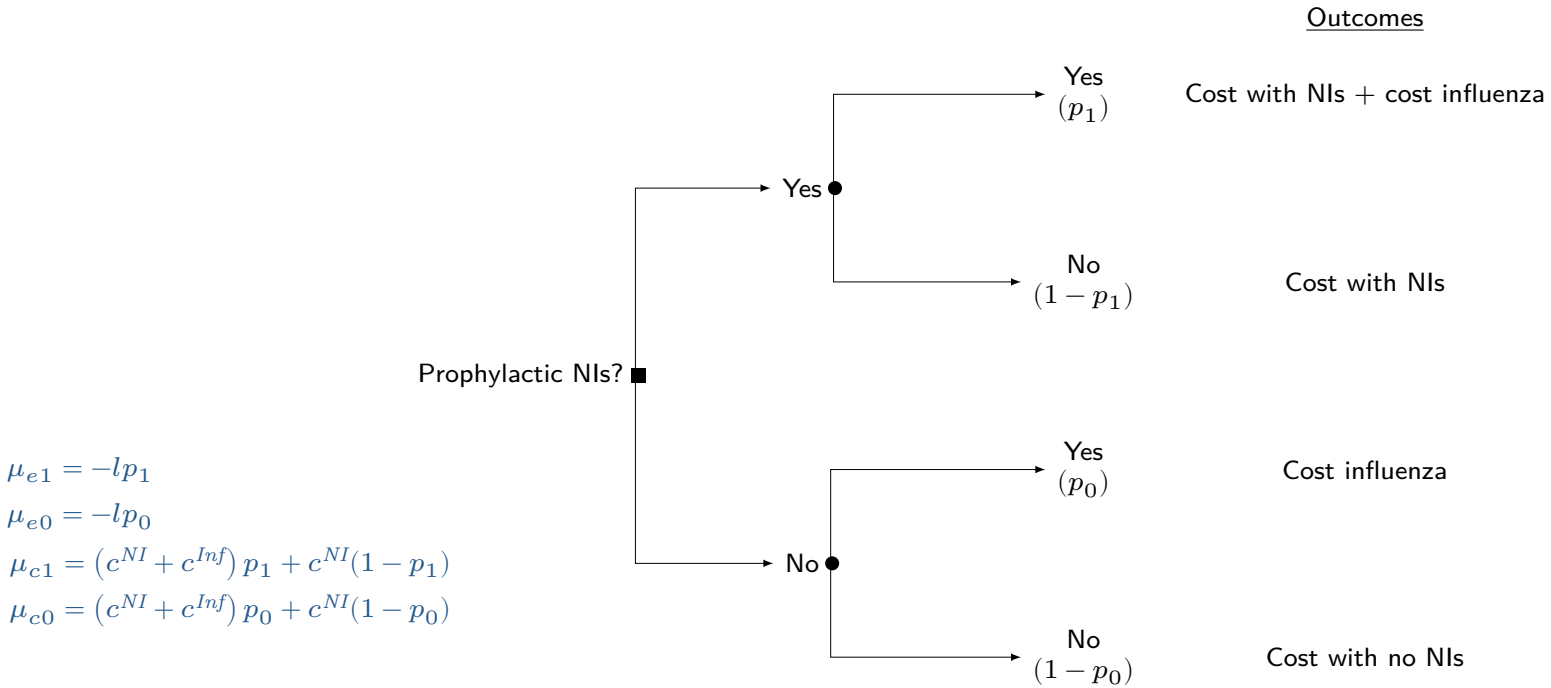
- Combines parameters to obtain average measure for costs and effects
- Varies with the type of data & statistical model used

- Computes suitable measures of "cost-effectiveness"
- Dictates the best course of actions
- Standardised process

Maastricht University

# Statistical modelling for model-based analyses

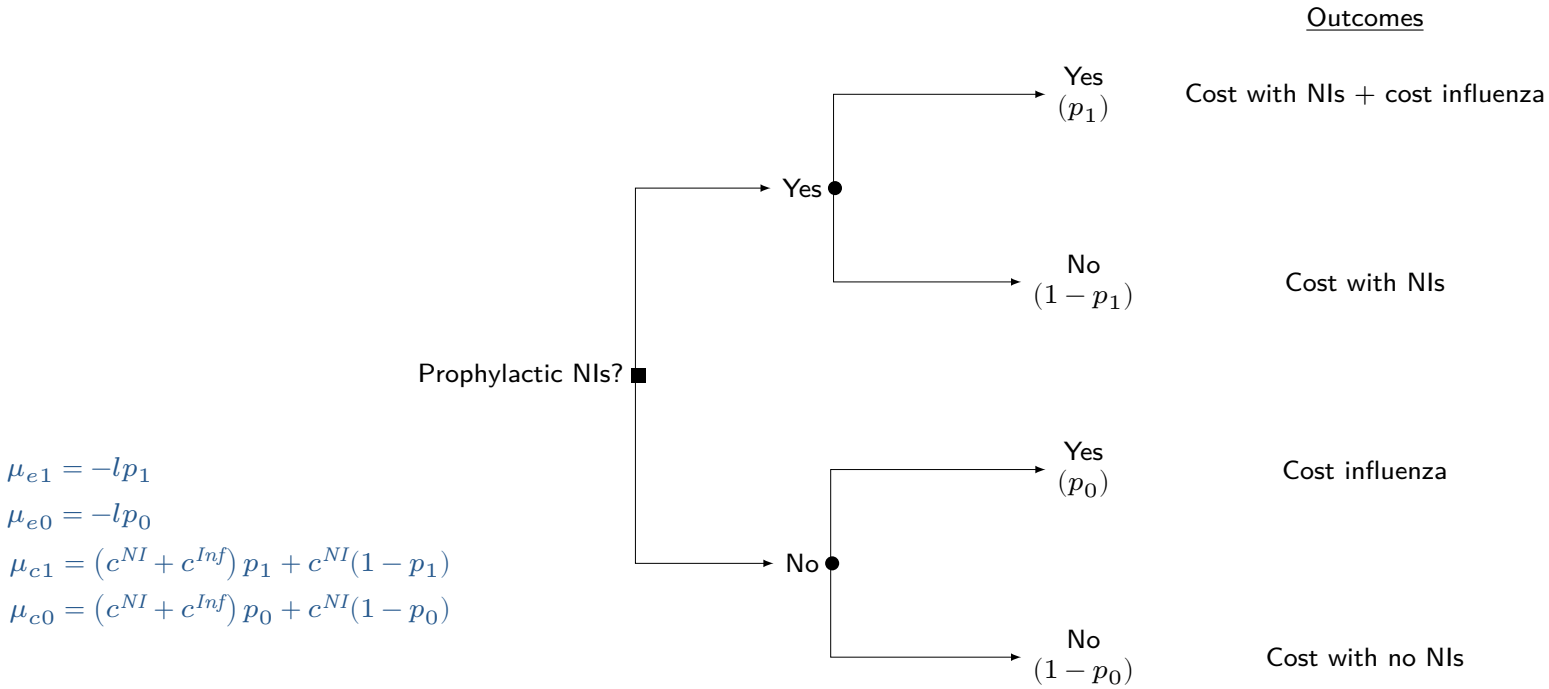
1. Build a **population level model** (eg decision tree/Markov model)




**NB**: "data" are represented by summary statistics for $\theta = (p_0, p_1, l, ...)$

# Statistical modelling for model-based analyses

1. Build a **population level model** (eg decision tree/Markov model)

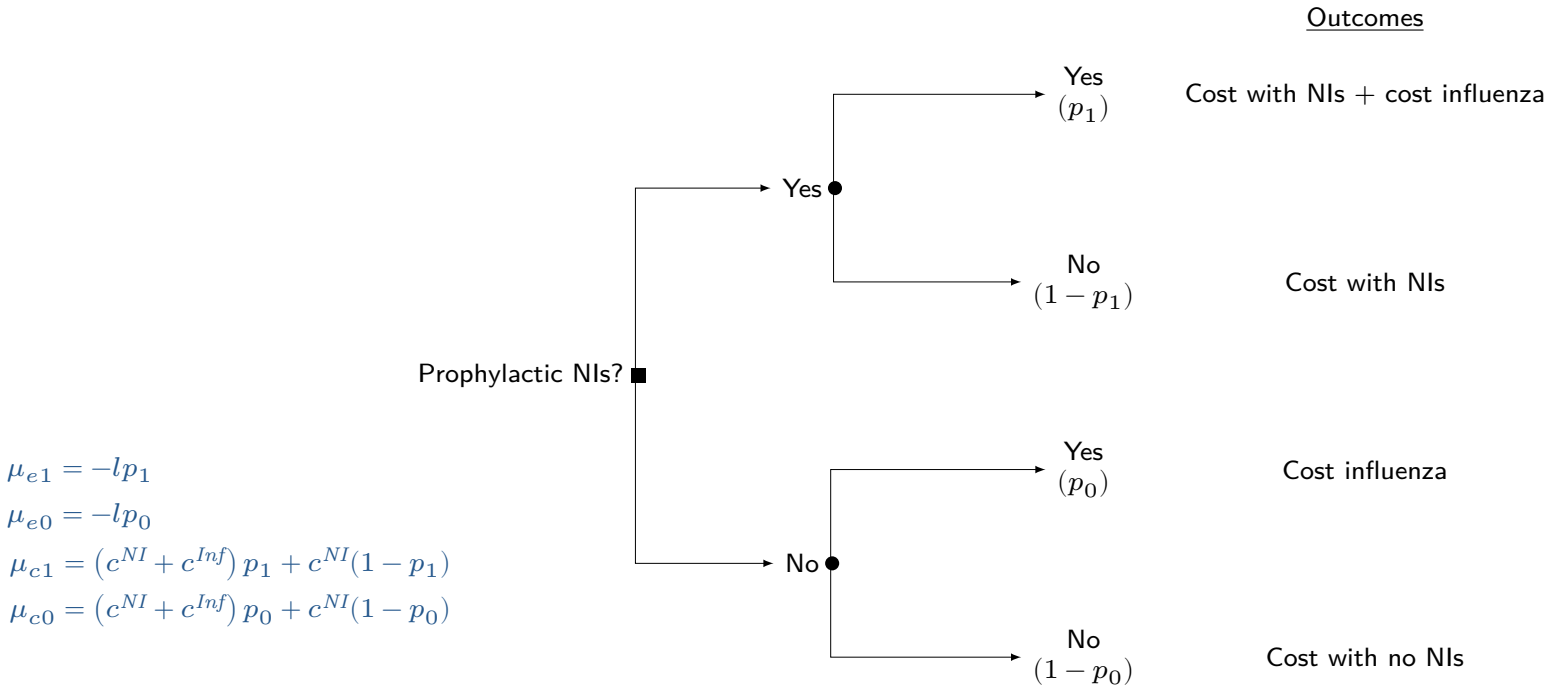



**NB**: "data" are represented by summary statistics for $\theta = (p_0, p_1, l, ...)$

2. Use point estimates for the parameters to build the "base-case" (average) evaluation

# Statistical modelling for model-based analyses

1. Build a **population level model** (eg decision tree/Markov model)




**NB**: "data" are represented by summary statistics for $\theta = (p_0, p_1, l, ...)$

2. Use point estimates for the parameters to build the "base-case" (average) evaluation
3. Use resampling methods (eg bootstrap) to propage uncertainty in the point estimates and perform uncertainty analysis

# Statistical modelling for empirical analyses

| | | Demographics | | | HRQL data | | | | Resource use data | | | | Clinical outcome | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ID** | **Trt** | Sex | Age | ... | $u_0$ | $u_1$ | ... | $u_J$ | $c_0$ | $c_1$ | ... | $c_J$ | $y_0$ | $y_1$ | ... | $y_J$ |
| 1 | 1 | M | 23 | ... | 0.32 | 0.66 | ... | 0.44 | 103 | 241 | ... | 80 | $y_{10}$ | $y_{11}$ | ... | $y_{1J}$ |
| 2 | 1 | M | 21 | ... | 0.12 | 0.16 | ... | 0.38 | 1 204 | 1 808 | ... | 877 | $y_{20}$ | $y_{21}$ | ... | $y_{2J}$ |
| 3 | 2 | F | 19 | ... | 0.49 | 0.55 | ... | 0.88 | 16 | 12 | ... | 22 | $y_{30}$ | $y_{31}$ | ... | $y_{3J}$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

$y_{ij}$ = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
$u_{ij}$ = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
$c_{ij}$ = Use of resources (drugs, hospital, GP appointments, ...)

Maastricht University

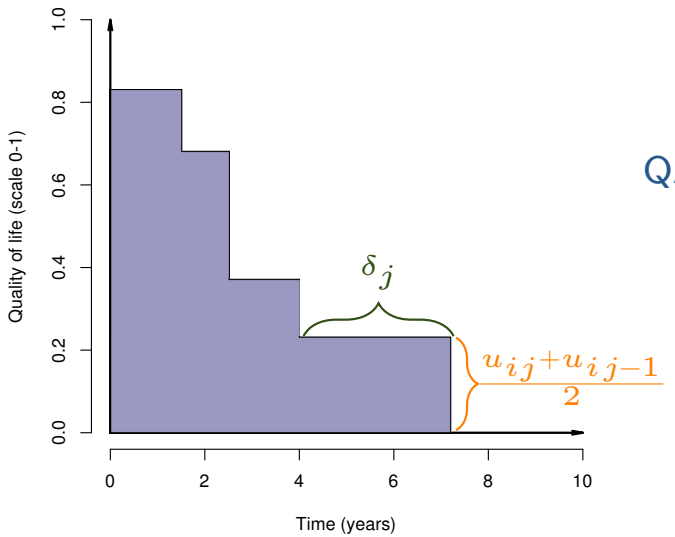
# Statistical modelling for empirical analyses

| | | Demographics | | | HRQL data | | | | Resource use data | | | | Clinical outcome | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ID** | **Trt** | Sex | Age | ... | $u_0$ | $u_1$ | ... | $u_J$ | $c_0$ | $c_1$ | ... | $c_J$ | $y_0$ | $y_1$ | ... | $y_J$ |
| 1 | 1 | M | 23 | ... | 0.32 | 0.66 | ... | 0.44 | 103 | 241 | ... | 80 | $y_{10}$ | $y_{11}$ | ... | $y_{1J}$ |
| 2 | 1 | M | 21 | ... | 0.12 | 0.16 | ... | 0.38 | 1 204 | 1 808 | ... | 877 | $y_{20}$ | $y_{21}$ | ... | $y_{2J}$ |
| 3 | 2 | F | 19 | ... | 0.49 | 0.55 | ... | 0.88 | 16 | 12 | ... | 22 | $y_{30}$ | $y_{31}$ | ... | $y_{3J}$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

$y_{ij}$ = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
$u_{ij}$ = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
$c_{ij}$ = Use of resources (drugs, hospital, GP appointments, ...)

1. Compute QALYs and total costs as

$$e_i = \sum_{j=1}^{J} \left(u_{ij} + u_{ij-1}\right) \frac{\delta_j}{2} \quad \text{and} \quad c_i = \sum_{j=1}^{J} c_{ij}, \qquad \left[\text{with: } \delta_j = \frac{\text{Time}_j - \text{Time}_{j-1}}{\text{Unit of time}}\right]$$




$\text{QALY}_i$ = "Area under the curve"

# Statistical modelling for empirical analyses

| | | Demographics | | | HRQL data | | | | Resource use data | | | | Clinical outcome | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ID** | **Trt** | Sex | Age | ... | $u_0$ | $u_1$ | ... | $u_J$ | $c_0$ | $c_1$ | ... | $c_J$ | $y_0$ | $y_1$ | ... | $y_J$ |
| 1 | 1 | M | 23 | ... | 0.32 | 0.66 | ... | 0.44 | 103 | 241 | ... | 80 | $y_{10}$ | $y_{11}$ | ... | $y_{1J}$ |
| 2 | 1 | M | 21 | ... | 0.12 | 0.16 | ... | 0.38 | 1 204 | 1 808 | ... | 877 | $y_{20}$ | $y_{21}$ | ... | $y_{2J}$ |
| 3 | 2 | F | 19 | ... | 0.49 | 0.55 | ... | 0.88 | 16 | 12 | ... | 22 | $y_{30}$ | $y_{31}$ | ... | $y_{3J}$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

$y_{ij}$ = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
$u_{ij}$ = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
$c_{ij}$ = Use of resources (drugs, hospital, GP appointments, ...)

1. Compute QALYs and total costs as

$$e_i = \sum_{j=1}^{J} \left(u_{ij} + u_{ij-1}\right) \frac{\delta_j}{2} \quad \text{and} \quad c_i = \sum_{j=1}^{J} c_{ij}, \qquad \left[\text{with: } \delta_j = \frac{\text{Time}_j - \text{Time}_{j-1}}{\text{Unit of time}}\right]$$

2. (Often implicitly) assume normality and linearity and model **independently** individual QALYs and total costs by controlling for baseline values

$$\begin{aligned} e_i &= \alpha_{e0} + \alpha_{e1} u_{0i} + \alpha_{e2} \text{Trt}_i + \varepsilon_{ei} \,[+ \ldots], & \varepsilon_{ei} &\sim \text{Normal}(0, \sigma_e) \\ c_i &= \alpha_{c0} + \alpha_{c1} c_{0i} + \alpha_{c2} \text{Trt}_i + \varepsilon_{ci} \,[+ \ldots], & \varepsilon_{ci} &\sim \text{Normal}(0, \sigma_c) \end{aligned}$$

Maastricht University

# Statistical modelling for empirical analyses

| | | Demographics | | | HRQL data | | | | Resource use data | | | | Clinical outcome | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ID** | **Trt** | Sex | Age | ... | $u_0$ | $u_1$ | ... | $u_J$ | $c_0$ | $c_1$ | ... | $c_J$ | $y_0$ | $y_1$ | ... | $y_J$ |
| 1 | 1 | M | 23 | ... | 0.32 | 0.66 | ... | 0.44 | 103 | 241 | ... | 80 | $y_{10}$ | $y_{11}$ | ... | $y_{1J}$ |
| 2 | 1 | M | 21 | ... | 0.12 | 0.16 | ... | 0.38 | 1 204 | 1 808 | ... | 877 | $y_{20}$ | $y_{21}$ | ... | $y_{2J}$ |
| 3 | 2 | F | 19 | ... | 0.49 | 0.55 | ... | 0.88 | 16 | 12 | ... | 22 | $y_{30}$ | $y_{31}$ | ... | $y_{3J}$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

$y_{ij}$ = Survival time, event indicator (eg CVD), number of events, continuous measurement (eg blood pressure), ...
$u_{ij}$ = Utility-based score to value health (eg EQ-5D, SF-36, Hospital Anxiety & Depression Scale, ...)
$c_{ij}$ = Use of resources (drugs, hospital, GP appointments, ...)

1. Compute QALYs and total costs as

$$e_i = \sum_{j=1}^{J} \left(u_{ij} + u_{ij-1}\right) \frac{\delta_j}{2} \quad \text{and} \quad c_i = \sum_{j=1}^{J} c_{ij}, \qquad \left[\text{with: } \delta_j = \frac{\text{Time}_j - \text{Time}_{j-1}}{\text{Unit of time}}\right]$$

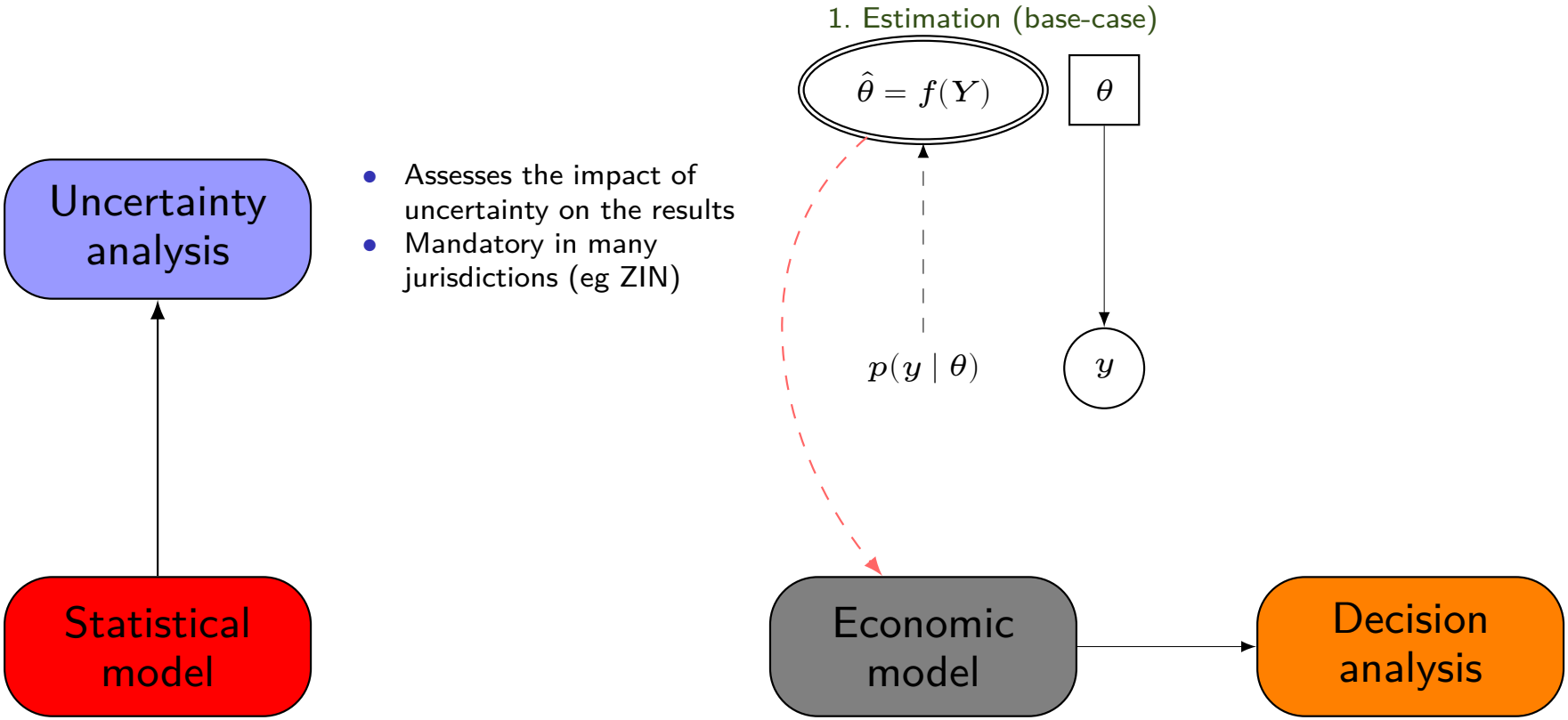
2. (Often implicitly) assume normality and linearity and model **independently** individual QALYs and total costs by controlling for baseline values

$$\begin{aligned} e_i &= \alpha_{e0} + \alpha_{e1} u_{0i} + \alpha_{e2} \text{Trt}_i + \varepsilon_{ei} \,[+ \ldots], & \varepsilon_{ei} &\sim \text{Normal}(0, \sigma_e) \\ c_i &= \alpha_{c0} + \alpha_{c1} c_{0i} + \alpha_{c2} \text{Trt}_i + \varepsilon_{ci} \,[+ \ldots], & \varepsilon_{ci} &\sim \text{Normal}(0, \sigma_c) \end{aligned}$$

3. Estimate population average cost and effectiveness differentials and use **bootstrap** to quantify uncertainty

Maastricht University

# "Two-stage" approach to HTA



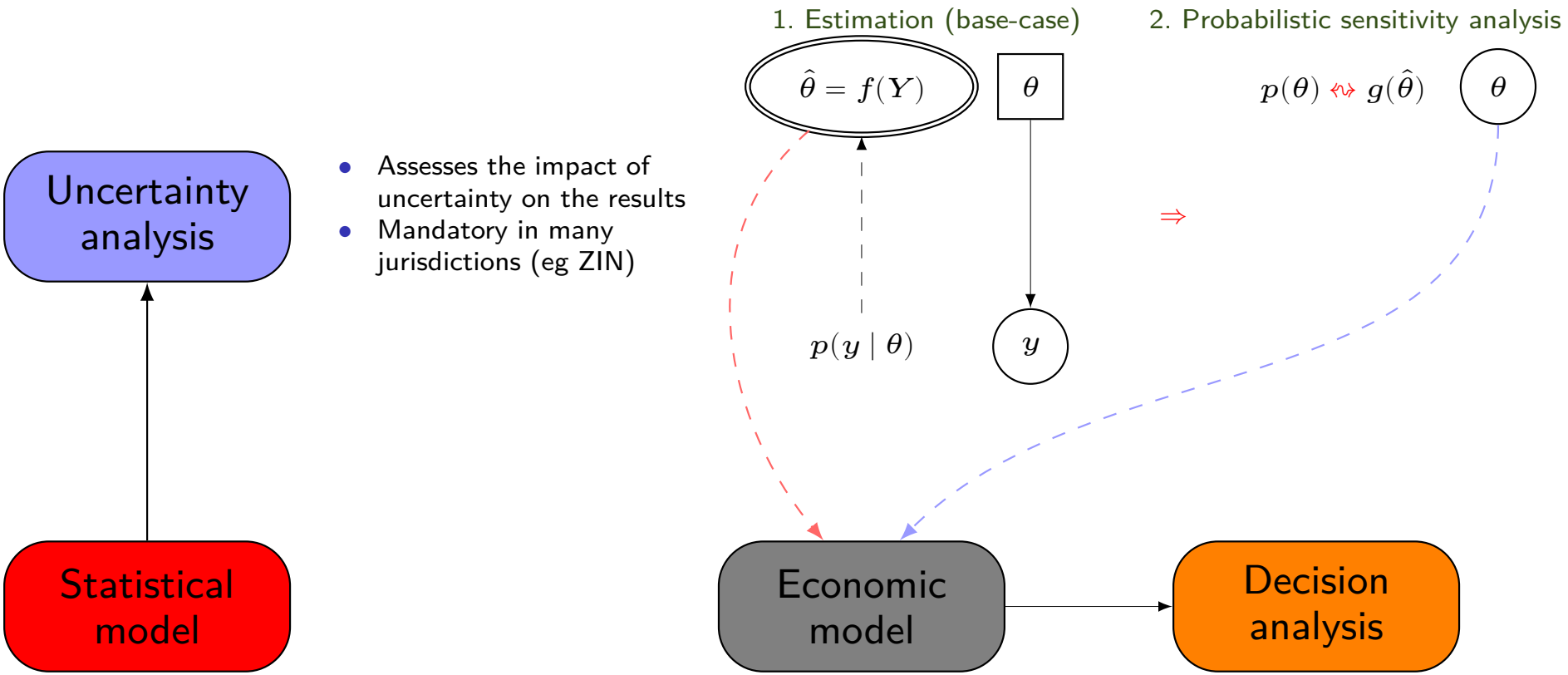

- Assesses the impact of uncertainty on the results
- Mandatory in many jurisdictions (eg ZIN)

- Estimates relevant parameters $\theta$
- Varies with the type of data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and effects
- Varies with the type of data & statistical model used

- Computes suitable measures of "cost-effectiveness"
- Dictates the best course of actions
- Standardised process

# "Two-stage" approach to HTA




- Assesses the impact of uncertainty on the results
- Mandatory in many jurisdictions (eg ZIN)

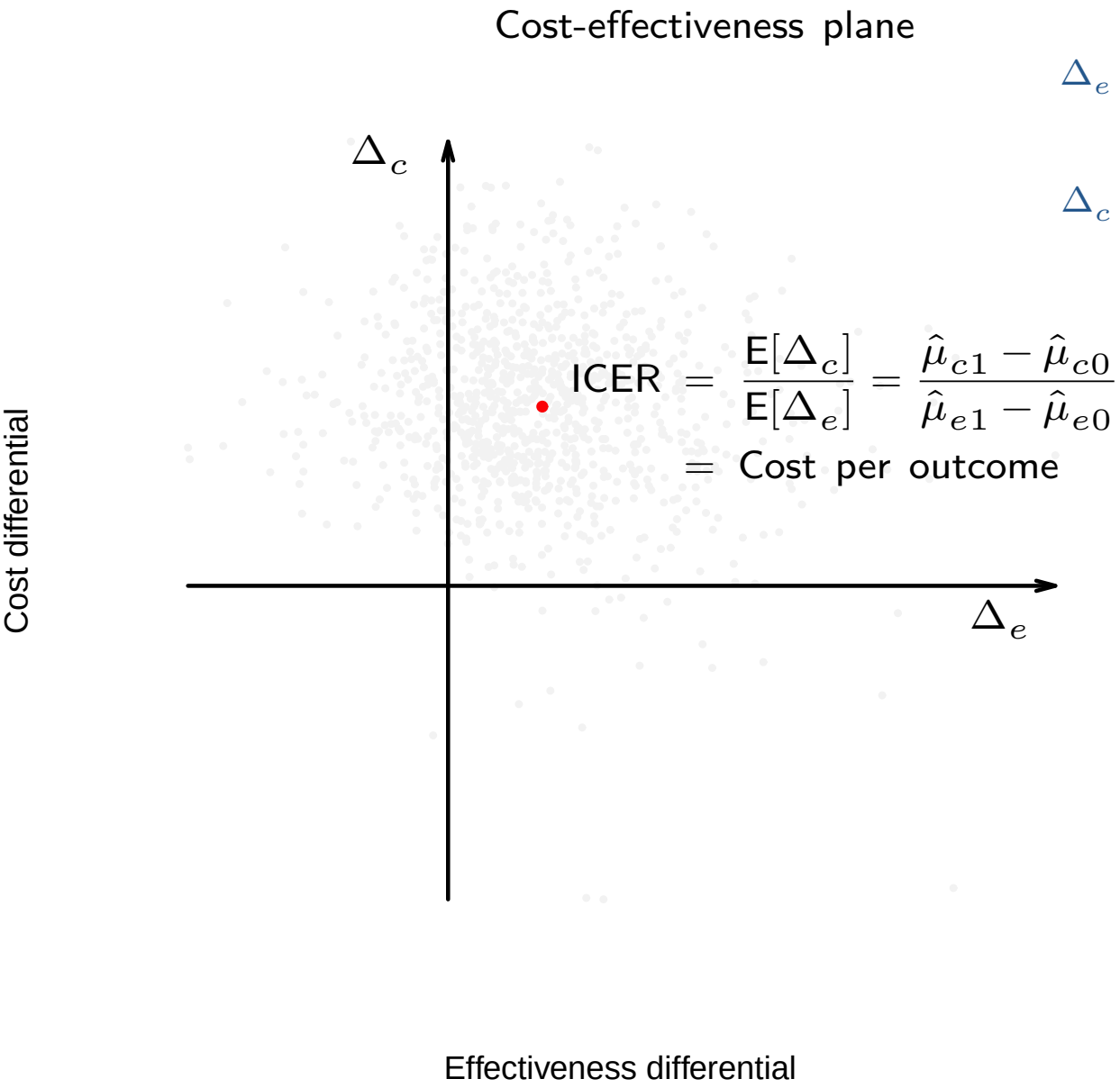
- Estimates relevant parameters $\theta$
- Varies with the type of data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and effects
- Varies with the type of data & statistical model used

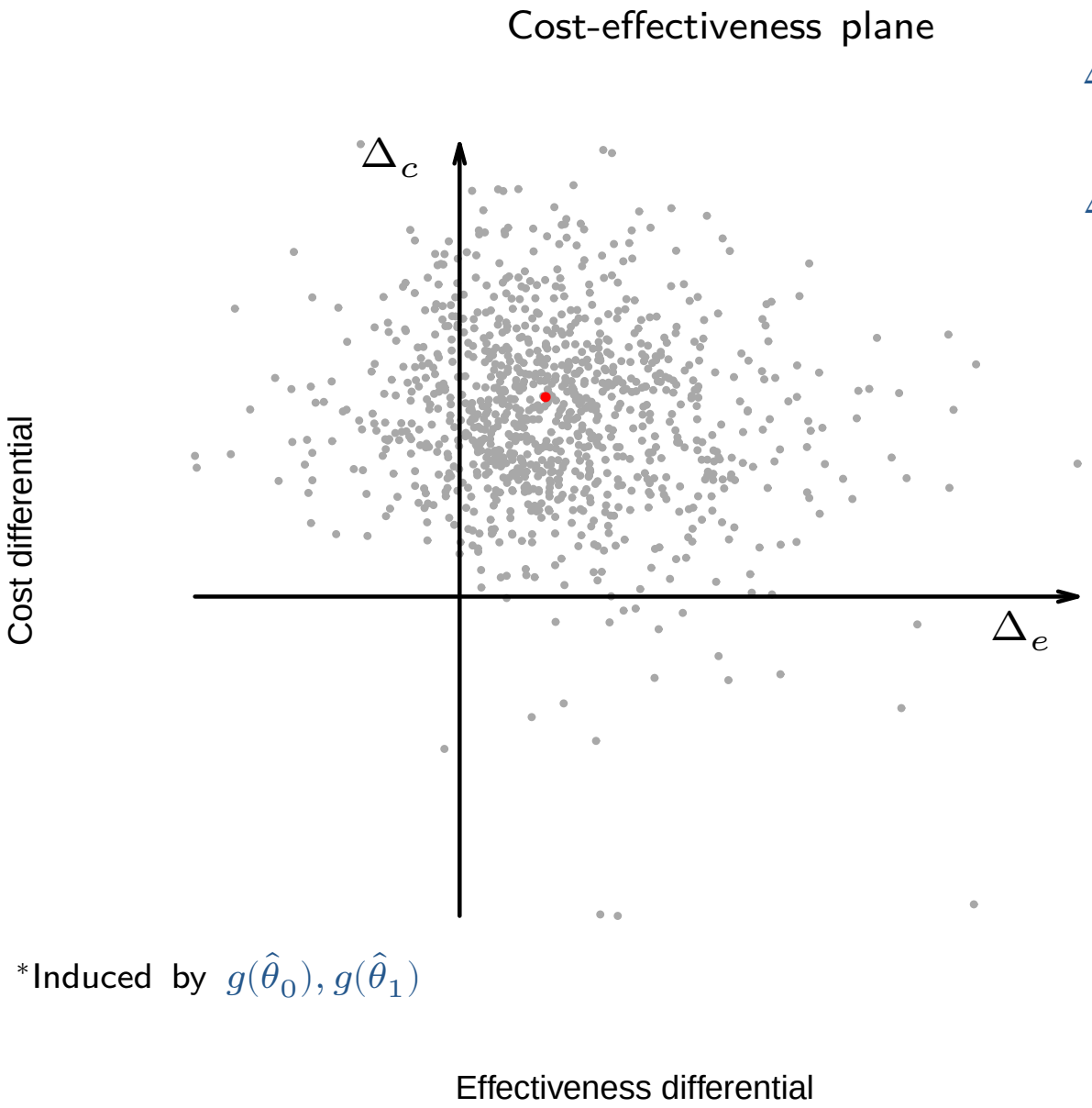
- Computes suitable measures of "cost-effectiveness"
- Dictates the best course of actions
- Standardised process

*"Two-stage approach" (Spiegelhalter, Abrams & Myles, 2004)*

# Decision + Uncertainty* analysis



# Decision + Uncertainty* analysis




$$\Delta_e = \underbrace{\mathsf{E}[e \mid \theta_1]}_{\mu_{e1}} - \underbrace{\mathsf{E}[e \mid \theta_0]}_{\mu_{e0}}$$

$$\Delta_c = \underbrace{\mathsf{E}[c \mid \theta_1]}_{\mu_{c1}} - \underbrace{\mathsf{E}[c \mid \theta_0]}_{\mu_{c0}}$$

*Induced by $g(\hat{\theta}_0), g(\hat{\theta}_1)$

Section 2

# Zorginstituut Nederland (ZIN) 2024 guidelines

# ZIN 2024 guidelines: summary

- Cost-effectiveness in the Netherlands has become more and more important in reimbursement decisions of the *National Health Care Institute* over the years
- **Standardise** analyses to improve comparability and enhance quality
- Revision of elements for the "**Reference Case**" to which all economic evaluations *have to comply with*

Maastricht University

# ZIN 2024 guidelines: summary

- Cost-effectiveness in the Netherlands has become more and more important in reimbursement decisions of the *National Health Care Institute* over the years
- **Standardise** analyses to improve comparability and enhance quality
- Revision of elements for the "**Reference Case**" to which all economic evaluations *have to comply with*
  1. **Perspective** of the analysis
     - Default: **Societal** - *all societal costs and benefits irrespective of who are the beneficiaries/payers*
     - Alternative perspectives may be presented as **scenario analyses**

# ZIN 2024 guidelines: summary

- Cost-effectiveness in the Netherlands has become more and more important in reimbursement decisions of the *National Health Care Institute* over the years
- **Standardise** analyses to improve comparability and enhance quality
- Revision of elements for the "**Reference Case**" to which all economic evaluations *have to comply with*
  1. **Perspective** of the analysis
  2. **PICOTS** criteria
     - Patient, Intervention, Comparison & Setting: **Dutch practice** - *relevant to current clinical practice at the time*
     - Outcomes: **Costs** (healthcare, P&F, other sectors) & **QALYs** (EQ-5D-5L)
     - Time horizon: **lifelong** (if possible)

Maastricht University

# ZIN 2024 guidelines: summary

- Cost-effectiveness in the Netherlands has become more and more important in reimbursement decisions of the *National Health Care Institute* over the years
- **Standardise** analyses to improve comparability and enhance quality
- Revision of elements for the "**Reference Case**" to which all economic evaluations *have to comply with*
  1. **Perspective** of the analysis
  2. **PICOTS** criteria
  3. **Type** of evaluation
     - Default: **CUA** - *allows comparison across populations and inteventions*
     - A **CEA** may also be conducted in addition

Maastricht University

# ZIN 2024 guidelines: summary

- Cost-effectiveness in the Netherlands has become more and more important in reimbursement decisions of the *National Health Care Institute* over the years
- **Standardise** analyses to improve comparability and enhance quality
- Revision of elements for the "**Reference Case**" to which all economic evaluations *have to comply with*
  1. **Perspective** of the analysis
  2. **PICOTS** criteria
  3. **Type** of evaluation
  4. **Data** (effectiveness, costs and QoL)
     - Litertaure data assessed via **systematic literature review**
     - Costs = volume (unit) $\times$ price (per unit) –> **Costing Manual**
     - QoL as "utilities" via EQ-5D-5L –> **QALY and QoL Manual**

Maastricht University

# ZIN 2024 guidelines: summary

- Cost-effectiveness in the Netherlands has become more and more important in reimbursement decisions of the *National Health Care Institute* over the years
- **Standardise** analyses to improve comparability and enhance quality
- Revision of elements for the "**Reference Case**" to which all economic evaluations *have to comply with*
  1. **Perspective** of the analysis
  2. **PICOTS** criteria
  3. **Type** of evaluation
  4. **Data** (effectiveness, costs and QoL)
  5. **Methods**
     - MB: Discount, Extrapolation, Subgroup, Uncertainty, Validation
     - **EMP**: Missingness, Adjustment, Uncertainty

Maastricht University

# ZIN 2024 guidelines: summary

- Cost-effectiveness in the Netherlands has become more and more important in reimbursement decisions of the *National Health Care Institute* over the years
- **Standardise** analyses to improve comparability and enhance quality
- Revision of elements for the "**Reference Case**" to which all economic evaluations *have to comply with*
  1. **Perspective** of the analysis
  2. **PICOTS** criteria
  3. **Type** of evaluation
  4. **Data** (effectiveness, costs and QoL)
  5. **Methods**
  6. **Reporting**
     - **Data**, **Methods**, **Results**

Maastricht University

# Analysis Methods in HTA

- **Study Design**:
  - **Empirical** - *costs & effects at patient level from a controlled study*
  - Model-based - *expected costs & effects estimated via simulation*

# Analysis Methods in HTA

- **Study Design**:
  - **Empirical** - *costs & effects at patient level from a controlled study*
  - Model-based - *expected costs & effects estimated via simulation*
- **Choice of methods** depends on the study design:
  - Account for patient-level data **complexities** (e.g. *imbalance*, *missingnes*, *skewness*, *correlation*, *clustering*)
  - In simulation, choice of model type and inputs should be based on the *research question* and *nature of the disease*

Maastricht University

# Analysis Methods in HTA

- **Study Design**:
  - **Empirical** - *costs & effects at patient level from a controlled study*
  - Model-based - *expected costs & effects estimated via simulation*
- **Choice of methods** depends on the study design:
  - Account for patient-level data **complexities** (e.g. *imbalance*, *missingnes*, *skewness*, *correlation*, *clustering*)
  - In simulation, choice of model type and inputs should be based on the *research question* and *nature of the disease*
- Focus on **recommended methods** in the context of a *RCT with a 1-year follow-up and homogeneous population*:
  - No *selection bias* due to lack of randomisation
  - No *discounting* for costs & effects
  - No *extrapolation* of results beyond end of trial
  - No *subgroup* or *Value of Information* analysis
  - No *validation* of source data

Maastricht University

# Statistical "issues" in analyses (El Alili et al. 2022)

- **Imbalances** in mean baseline effects/costs between arms, if not accounted for, may distort CE results (Manca et al. 2005, Asselt van et al. 2009)

# Statistical "issues" in analyses (El Alili et al. 2022)

- **Imbalances** in mean baseline effects/costs between arms, if not accounted for, may distort CE results (Manca et al. 2005, Asselt van et al. 2009)

- Ignoring the **correlation** between effects & costs is inappropriate and will lead to a loss in efficiency for the estimates (Willan et al. 2004)

Maastricht University

# Statistical "issues" in analyses (El Alili et al. 2022)

- **Imbalances** in mean baseline effects/costs between arms, if not accounted for, may distort CE results (Manca et al. 2005, Asselt van et al. 2009)
- Ignoring the **correlation** between effects & costs is inappropriate and will lead to a loss in efficiency for the estimates (Willan et al. 2004)
- **Skewness** in effects/costs in small samples undermines the validity of asymptotic tests and may lead to incorrect results (Barber et al. 2004)

Maastricht University

# Statistical "issues" in analyses (El Alili et al. 2022)

- **Imbalances** in mean baseline effects/costs between arms, if not accounted for, may distort CE results (Manca et al. 2005, Asselt van et al. 2009)
- Ignoring the **correlation** between effects & costs is inappropriate and will lead to a loss in efficiency for the estimates (Willan et al. 2004)
- **Skewness** in effects/costs in small samples undermines the validity of asymptotic tests and may lead to incorrect results (Barber et al. 2004)
- **Clustering** of data linked to treatment (eg cluster RCTs) invalidates the assumptions of standard methods (eg OLS), underestimates variance and possibly bias results (Gomes et al. 2012)

Maastricht University

# Statistical "issues" in analyses (El Alili et al. 2022)

- **Imbalances** in mean baseline effects/costs between arms, if not accounted for, may distort CE results (Manca et al. 2005, Asselt van et al. 2009)
- Ignoring the **correlation** between effects & costs is inappropriate and will lead to a loss in efficiency for the estimates (Willan et al. 2004)
- **Skewness** in effects/costs in small samples undermines the validity of asymptotic tests and may lead to incorrect results (Barber et al. 2004)
- **Clustering** of data linked to treatment (eg cluster RCTs) invalidates the assumptions of standard methods (eg OLS), underestimates variance and possibly bias results (Gomes et al. 2012)
- **Missing** effects/costs during follow-up are common and may introduce bias. Methods that appropriately account for missing data uncertainty and assess the sensitivity of results are needed (Gabrio et al. 2017, Leurent et a. 2018)

# How do we deal with all of this?

Section 3

# Methods to handle statistical issues in CEA

# Baseline imbalances

- Randomisation ensures that baseline variables are *balanced* between arms but **SOME** differences will inevitably occur

# Baseline imbalances

- Randomisation ensures that baseline variables are *balanced* between arms but **SOME** differences will inevitably occur
- This is particularly an issue for **baseline utilities** because:
  - they are used in the *computation of QALYs* through AUC method
  - they are likely to be *predictive* of utilities during follow-up

Maastricht University

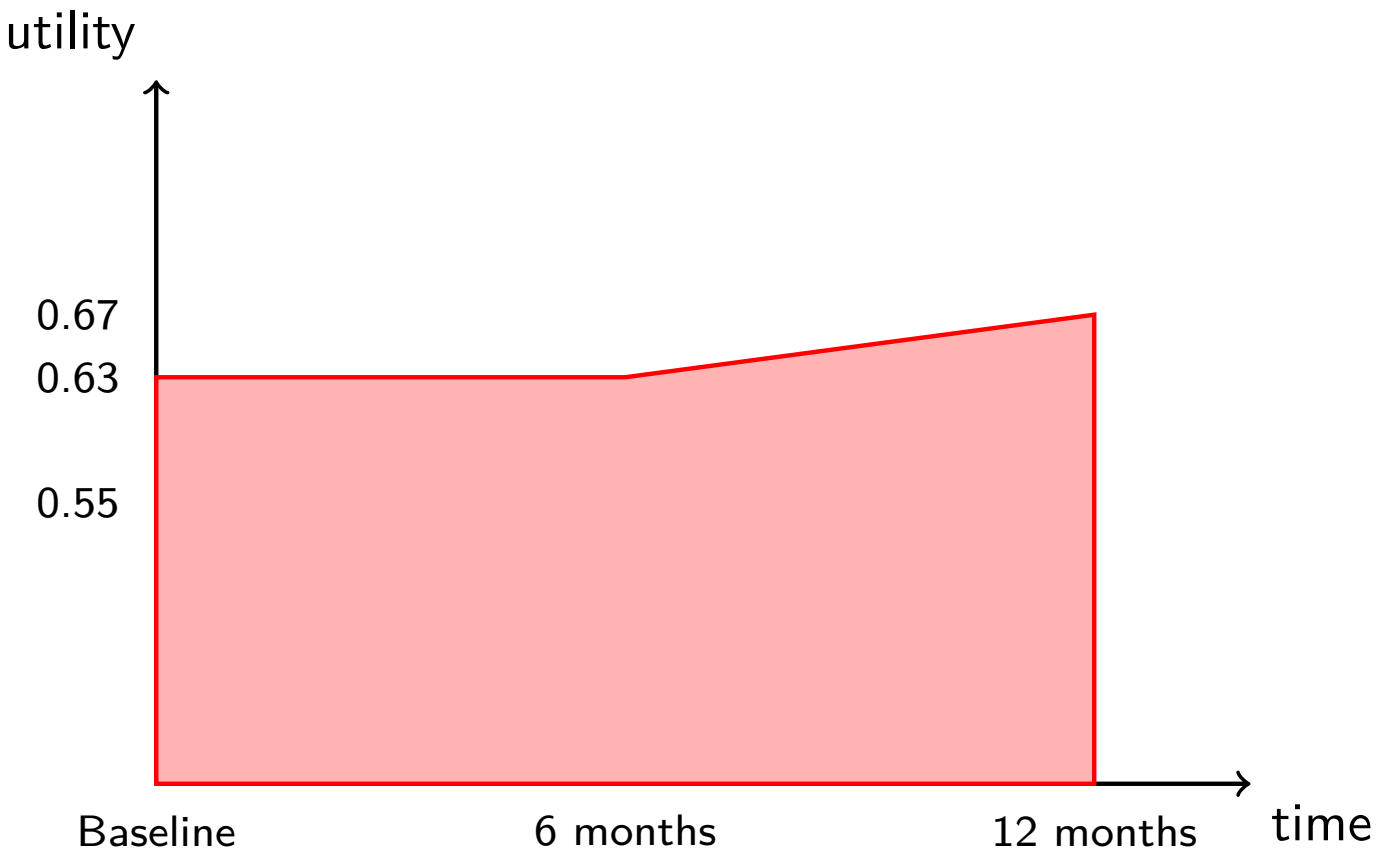
# Baseline imbalances

- Randomisation ensures that baseline variables are *balanced* between arms but **SOME** differences will inevitably occur
- This is particularly an issue for **baseline utilities** because:
  - they are used in the *computation of QALYs* through AUC method
  - they are likely to be *predictive* of utilities during follow-up
- Even small changes in $\Delta_e$ have consequences on *ICER* and *CE results*
- This is true **regardless** of whether the difference in baseline values is statistically significant or not
- This is also true for any baseline variable that is **strongly predictive** of follow-up outcome values (eg **baseline costs**)

Maastricht University

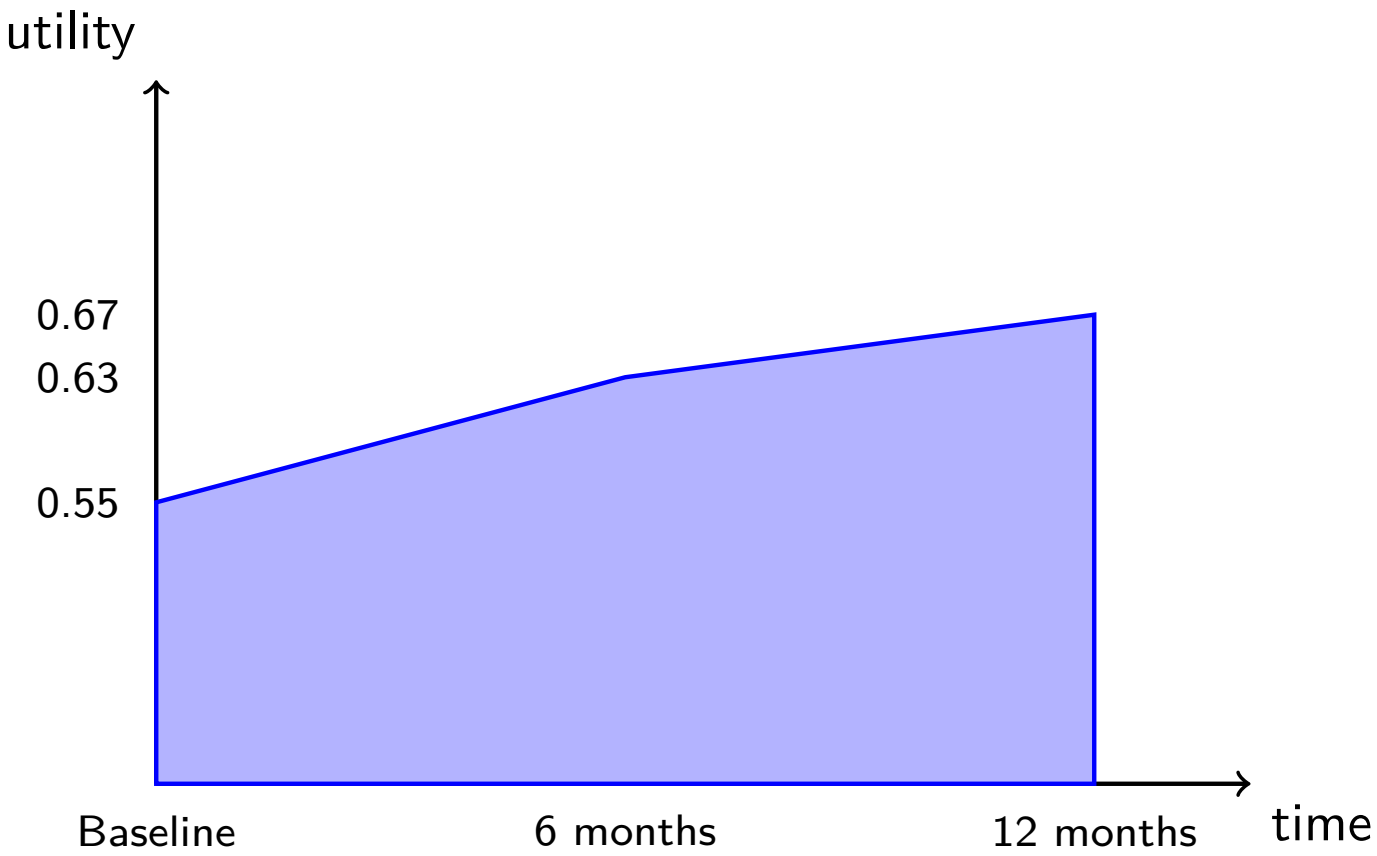
# Baseline imbalances - an example

- Old : $\text{QALY}_{\text{old}} = \left[\frac{0.63+0.63}{2} \times \frac{6}{12} + \frac{0.63+0.67}{2} \times \frac{6}{12}\right] = \mathbf{0.64}$



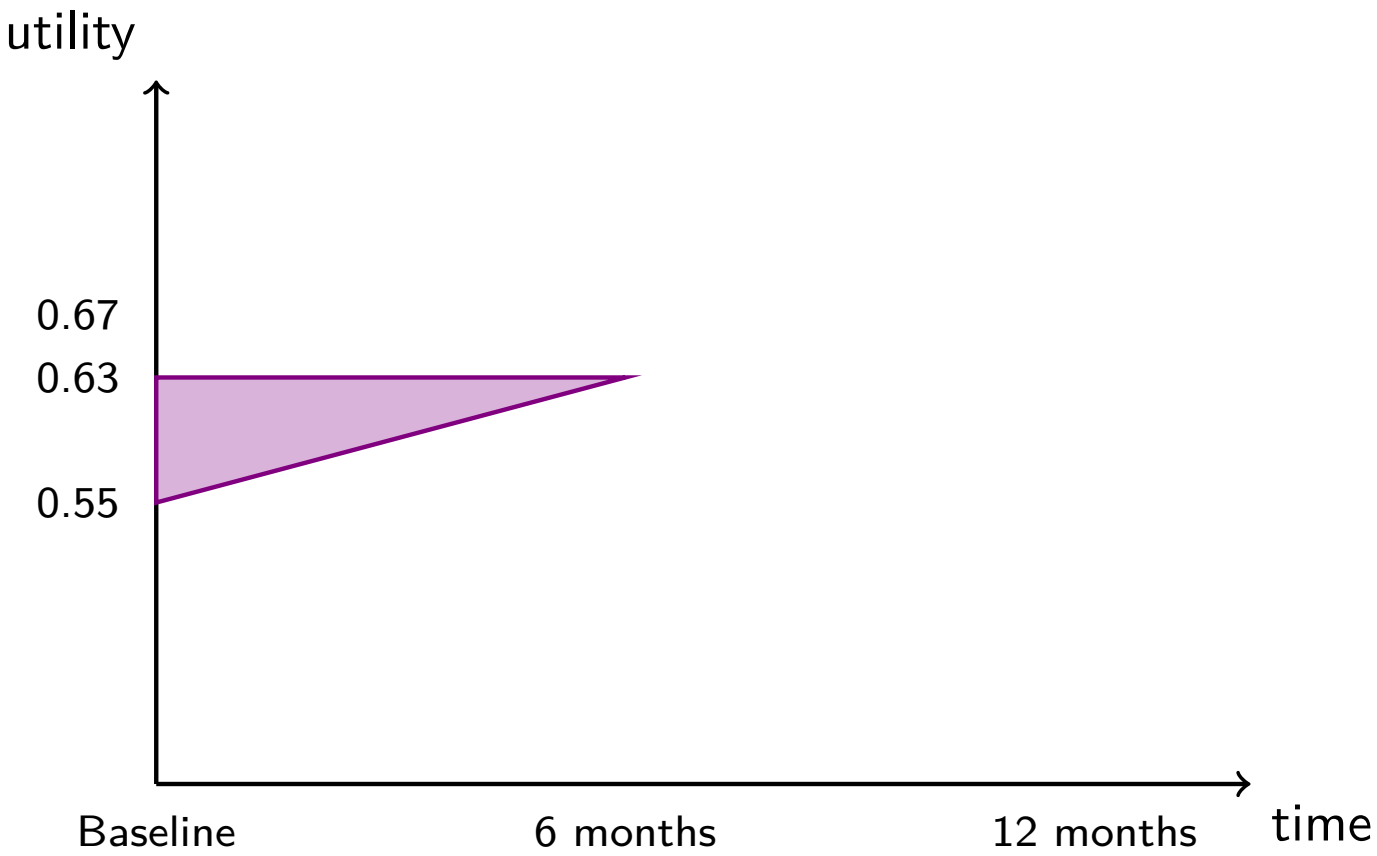
# Baseline imbalances - an example

- New : $\text{QALY}_{\text{new}} = \left[\frac{0.55+0.63}{2} \times \frac{6}{12} + \frac{0.63+0.67}{2} \times \frac{6}{12}\right] = \mathbf{0.62}$



# Baseline imbalances - an example

- $\Delta_e$ = New - Old : $\text{QALY}_{\text{new}} - \text{QALY}_{\text{old}} = \mathbf{0.62 - 0.64 = -0.02}$



# Baseline imbalances - unadjusted estimates

- **Unadjusted mean difference** can be obtained from the estimated *slope* ($\hat{\beta}_1$) of the linear regression

$$\text{QALY}_i = \beta_0 + \beta_1 \times \text{arm}_i + \varepsilon_i$$

- **Unadjusted means** in each arm can be obtained as *functions* of the estimated regression coefficients

$$\text{E}[\text{QALY} \mid \text{arm} = \text{old}] = \hat{\beta}_0$$
$$\text{E}[\text{QALY} \mid \text{arm} = \text{new}] = \hat{\beta}_0 + \hat{\beta}_1$$

Maastricht University

# Baseline imbalances - adjusted estimates

- Recommended approach is to *include the imbalanced baseline variable into the regression*

$$\text{QALY}_i = \beta_0 + \beta_1 \times \text{arm}_i + \beta_2 \times \text{u}_{i0} + \varepsilon_i$$

- **Adjusted** estimates for $\Delta_e$ and mean QALYs in each arm can be obtained exactly as in the unadjusted case **BUT**with the *inclusion* of $\text{u}_0$ into the model

Maastricht University

# Baseline imbalances - adjusted estimates

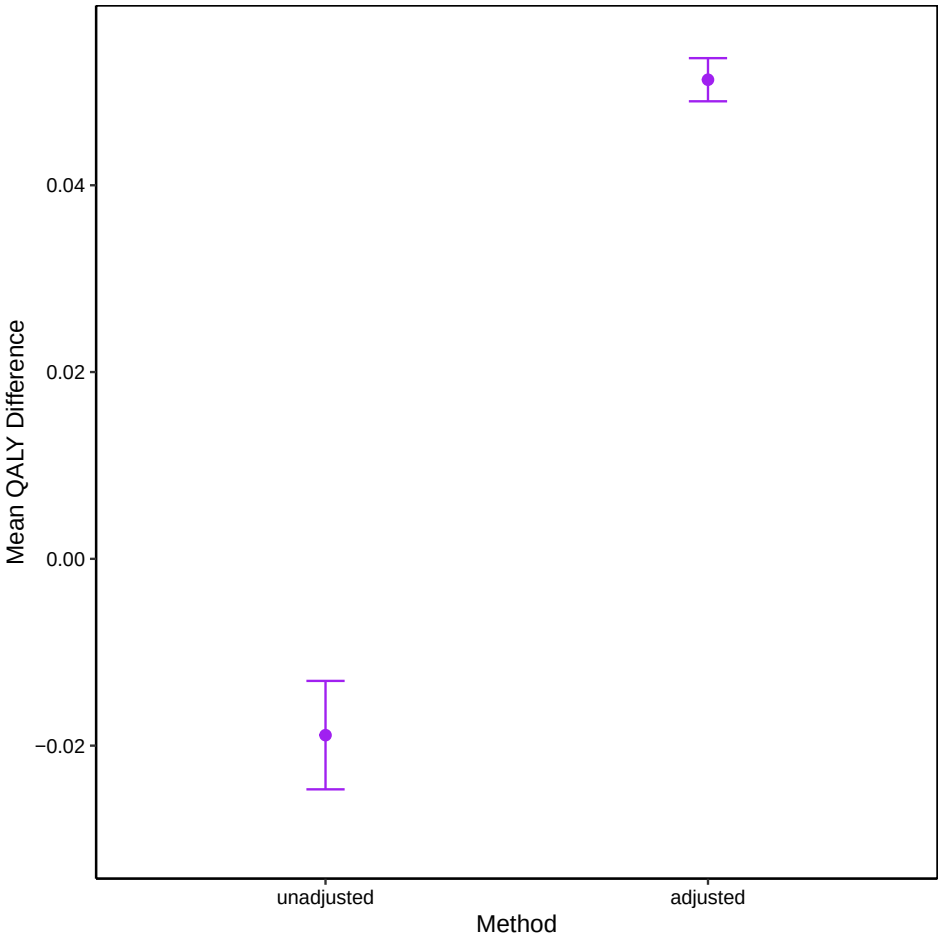
- Recommended approach is to *include the imbalanced baseline variable into the regression*

$$\text{QALY}_i = \beta_0 + \beta_1 \times \text{arm}_i + \beta_2 \times \text{u}_{i0} + \varepsilon_i$$

- **Adjusted** estimates for $\Delta_e$ and mean QALYs in each arm can be obtained exactly as in the unadjusted case **BUT**with the *inclusion* of $\text{u}_0$ into the model
- In the presence of baseline imbalances, the **adjusted model** allows to:
  - Retrieve through $\widehat{\beta}_1$ an **unbiased estimate** of $\Delta_e$ and E[QALY | arm]
  - Increase the **precision** of these estimates (ie *lower SEs & narrower CIs*)

Maastricht University

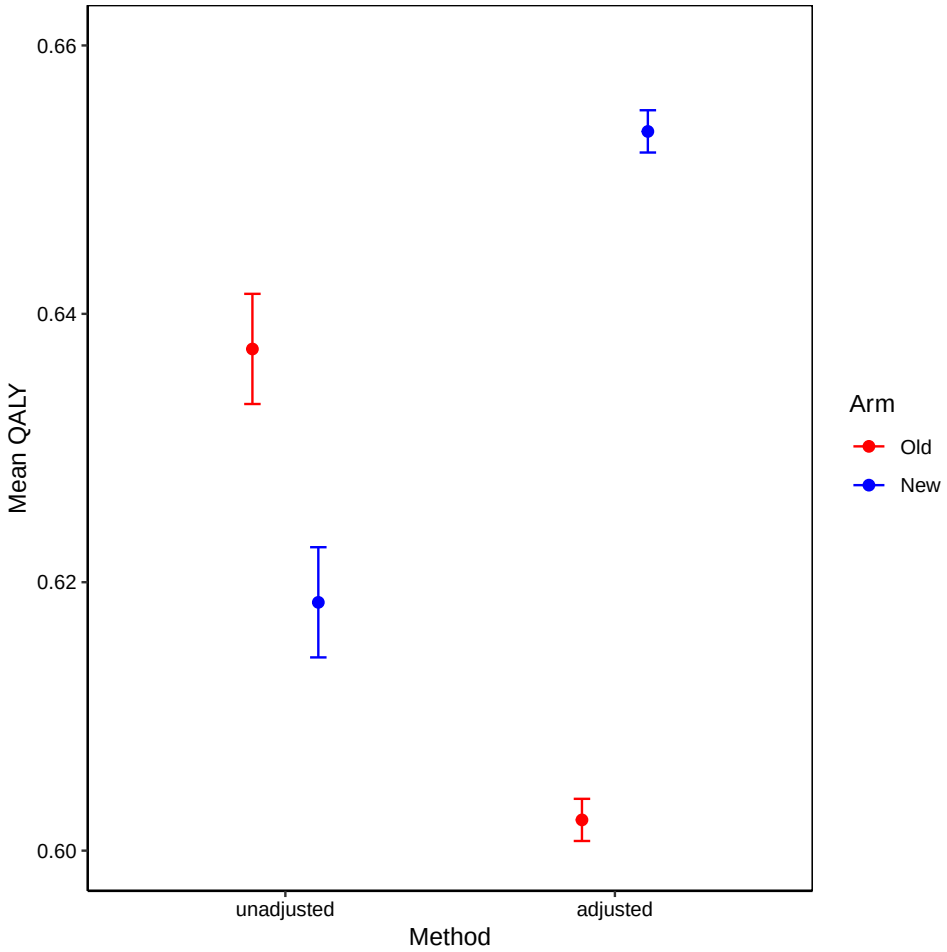
# Baseline imbalances - unadjusted vs adjusted estimates




- **Mean QALY difference $\Delta_e$**

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***-0.019*** | *-0.025* | *-0.013* | *unadjusted* |
| ***0.051*** | *0.049* | *0.054* | *adjusted* |

# Baseline imbalances - unadjusted vs adjusted estimates




- **Mean QALY** for arm = "Old"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.637*** | *0.633* | *0.641* | *unadjusted* |
| ***0.602*** | *0.601* | *0.604* | *adjusted* |

- **Mean QALY** for arm = "New"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.619*** | *0.614* | *0.623* | *unadjusted* |
| ***0.654*** | *0.652* | *0.655* | *adjusted* |

# Baseline imbalances - regression adjustment

- **Advantages**:
  - Easy to implement
  - Allow to control for multiple variables
  - Assess impact of specific variables on marginal CE outcomes (eg via interaction terms)

# Baseline imbalances - regression adjustment

- **Advantages**:
  - Easy to implement
  - Allow to control for multiple variables
  - Assess impact of specific variables on marginal CE outcomes (eg via interaction terms)
- **Drawbacks**:
  - Assume variables' distribution is the same across arms
  - Adjusting for many variables may result in "overfitting"

Maastricht University

# Baseline imbalances - regression adjustment

- **Advantages**:
  - Easy to implement
  - Allow to control for multiple variables
  - Assess impact of specific variables on marginal CE outcomes (eg via interaction terms)
- **Drawbacks**:
  - Assume variables' distribution is the same across arms
  - Adjusting for many variables may result in "overfitting"
- **Other approaches**:
  - *Propensity score adjustment* (Indurkhya et al. 2006)
  - *Propensity score matching* (Sekhon et al. 2012)
  - *Genetic matching* (Sekhon et al. 2012)

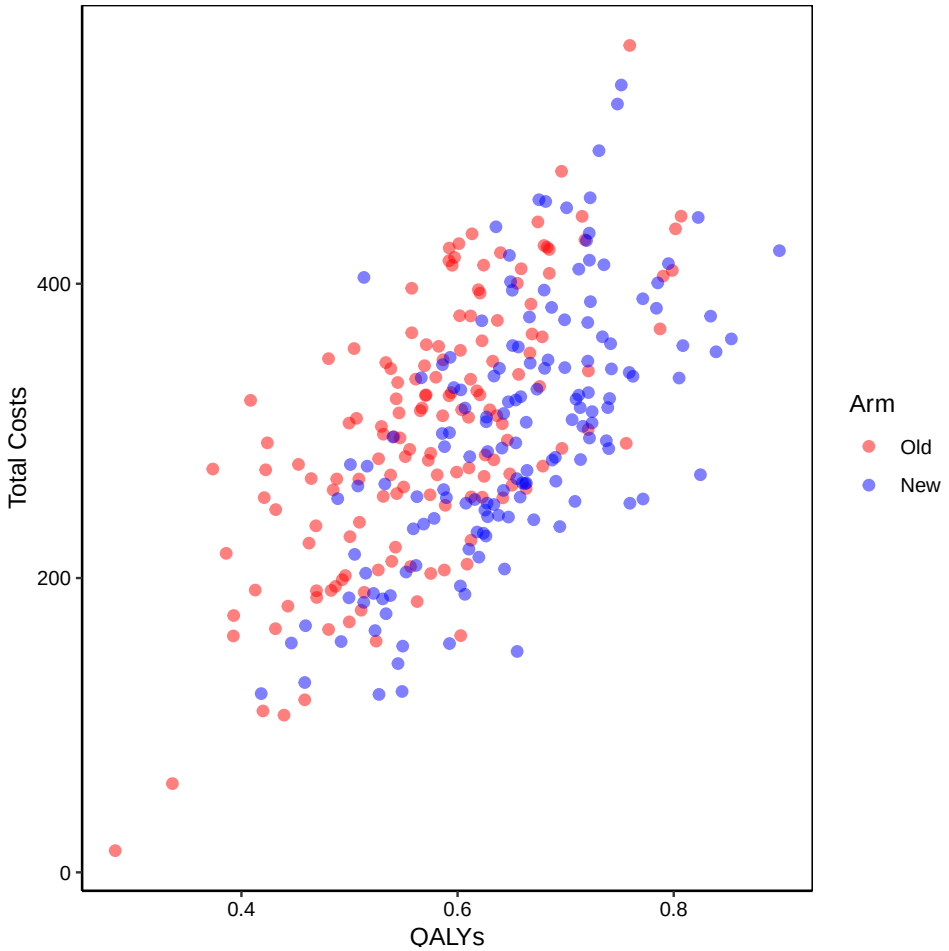
Maastricht University

# Correlation between CE outcomes

- Possible types of associations:
  - Treatments come from intensive research and are **(+)** associated with higher unit costs
  - Treatments **(-)** reduce care pathway costs (eg fewer hospitalisations, side effects, etc.)
- If the outcomes are **sufficiently correlated**:
  - **Joint modelling** of costs & effects is needed to properly characterise uncertainty around parameter estimates and CE results
  - There is additional benefit from *borrowing information* across outcomes to estimate variance components and standard errors *more efficiently* than in separate univariate analyses

Maastricht University

# Correlation between CE outcomes - an example




- Means and (Pearson's) **correlations** between costs & effects

| QALY | TC | corr | arm |
|---|---|---|---|
| *0.57* | *298* | ***0.694*** | *Old* |
| *0.65* | *299* | ***0.648*** | *New* |

# CE correlation - independent analsyes

- We could run **separate analyses** for each outcome using the models

$$\text{QALY}_i = \beta_0 + \beta_1 \times \text{arm}_i + \beta_2 \times u_{i0} + \varepsilon_{ie}$$
$$\text{TC}_i = \alpha_0 + \alpha_1 \times \text{arm}_i + \alpha_2 \times c_{i0} + \varepsilon_{ic}$$

- Get the estimates $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ and $\hat{\alpha} = (\hat{\alpha}_0, \hat{\alpha}_1)$, then derive:
  - **Incremental**$(\Delta_e, \Delta_c)$ and **marginal**(E[QALY | arm], E[TC | arm]) quantities of interest
  - Associated *measures of uncertainty* (eg SEs or CIs)

Maastricht University

# CE correlation - independent analsyes

- We could run **separate analyses** for each outcome using the models

$$\text{QALY}_i = \beta_0 + \beta_1 \times \text{arm}_i + \beta_2 \times u_{i0} + \varepsilon_{ie}$$
$$\text{TC}_i = \alpha_0 + \alpha_1 \times \text{arm}_i + \alpha_2 \times c_{i0} + \varepsilon_{ic}$$

- Get the estimates $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ and $\hat{\alpha} = (\hat{\alpha}_0, \hat{\alpha}_1)$, then derive:
  - **Incremental**$(\Delta_e, \Delta_c)$ and **marginal**(E[QALY | arm], E[TC | arm]) quantities of interest
  - Associated *measures of uncertainty* (eg SEs or CIs)
- **Important**: by doing so, we are assuming **independence** between the outcomes!

Maastricht University

# CE correlation - "joint" analsyes

- **Seemingly Unrelated Regression** (SUR) equations
  - Same models but with **error terms linked** through the parameter $\rho$:

$$\begin{pmatrix} \varepsilon_{ie} \\ \varepsilon_{ic} \end{pmatrix} \sim \text{Normal}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma_e^2 & \rho\sigma_e\sigma_c \\ \rho\sigma_c\sigma_e & \sigma_c^2 \end{pmatrix}\right]$$

Maastricht University

# CE correlation - "joint" analsyes

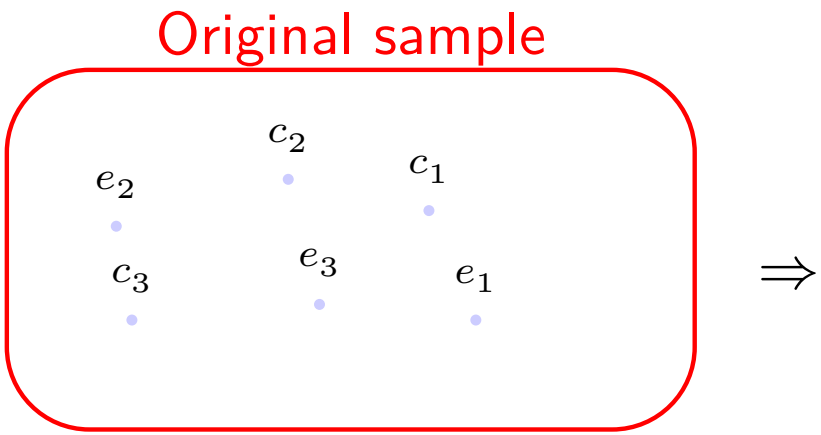
- **Seemingly Unrelated Regression** (SUR) equations
  - Same models but with **error terms linked** through the parameter $\rho$:

$$\begin{pmatrix} \varepsilon_{ie} \\ \varepsilon_{ic} \end{pmatrix} \sim \text{Normal} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \sigma_e^2 & \rho\sigma_e\sigma_c \\ \rho\sigma_c\sigma_e & \sigma_c^2 \end{pmatrix} \right]$$

- **Bootstrapping** costs & effects "in pairs"
  - Generate a "bootstrap" sample by **sampling with replacement** CE values for each individual from the observed data
  - Analyse CE outcomes and derive $(\hat{\beta}, \hat{\alpha})^b$ –> $(\Delta_e, \Delta_c)^b$
  - Iterate the process a *sufficiently large number of times* (B)
  - Use $(\Delta_e, \Delta_c)^b$ to approximate the sampling distribution of $(\Delta_e, \Delta_c)$
  - Use this distribution to quantify uncertainty (eg obtain CIs)

Maastricht University

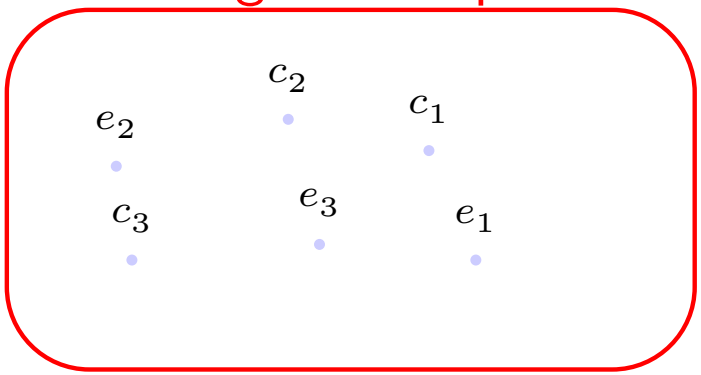
# What is bootstrapping?




- Original sample $\approx$ population
- Original sample size $n = 6$
- Mean $\bar{e} = 0.5$ & $\bar{c} = 150$
- Sd $s_e = 0.25$ & $s_c = 50$
- Coeff $\hat{\beta}_1 = 0.15$ & $\hat{\alpha}_1 = 35$

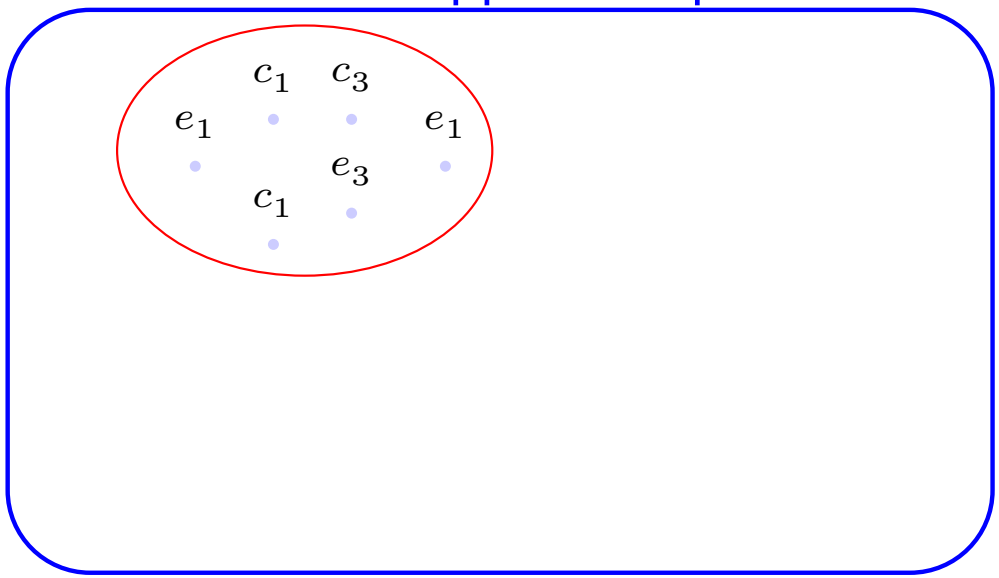
# What is bootstrapping?

## Original sample




$\Rightarrow$

## Bootstrapped samples

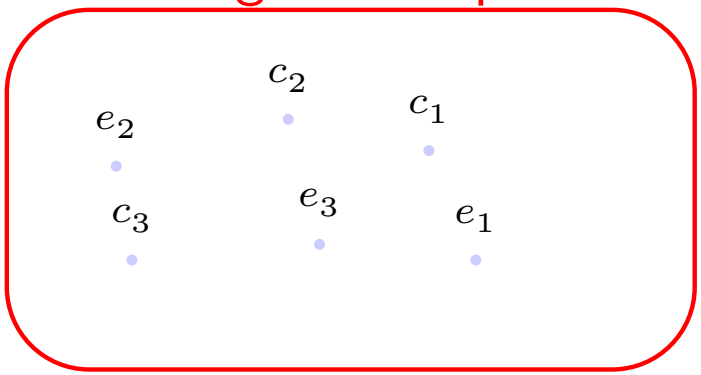



- Original sample $\approx$ population
- Original sample size $n = 6$
- Mean $\bar{e} = 0.5$ & $\bar{c} = 150$
- Sd $s_e = 0.25$ & $s_c = 50$
- Coeff $\hat{\beta}_1 = 0.15$ & $\hat{\alpha}_1 = 35$

- Bootstrap sample number $b = 1$
- Bootstrapped sample size $n^b = 6$
- Bootstrapped Mean $\bar{e}^b = 0.4$ & $\bar{c}^b = 160$
- Bootstrapped Sd $s_e^b = 0.2$ & $s_c^b = 50$
- Bootstrapped Coeff $\hat{\beta}_1^b = 0.2$ & $\hat{\alpha}_1^b = 40$

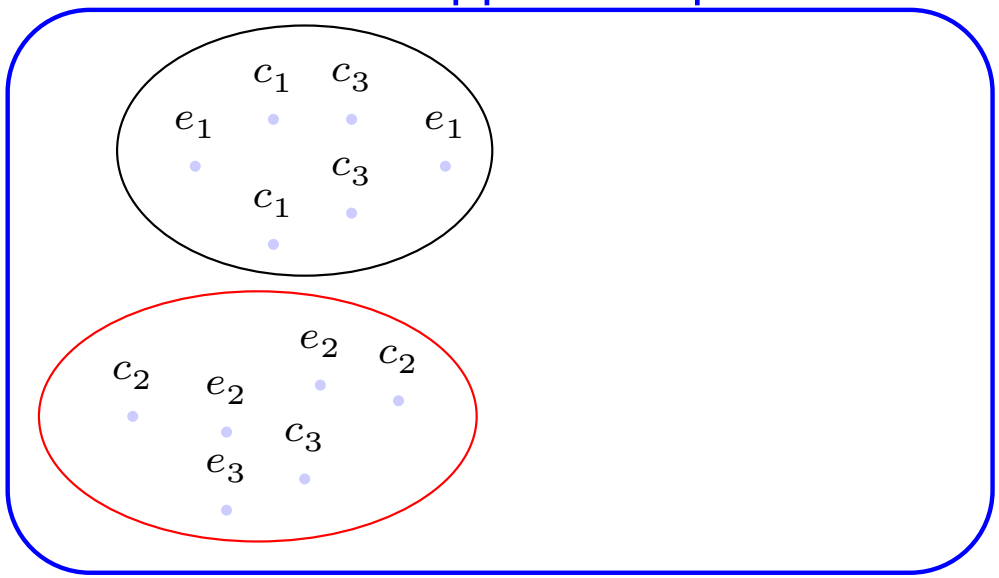
# What is bootstrapping?

## Original sample




$\Rightarrow$

## Bootstrapped samples

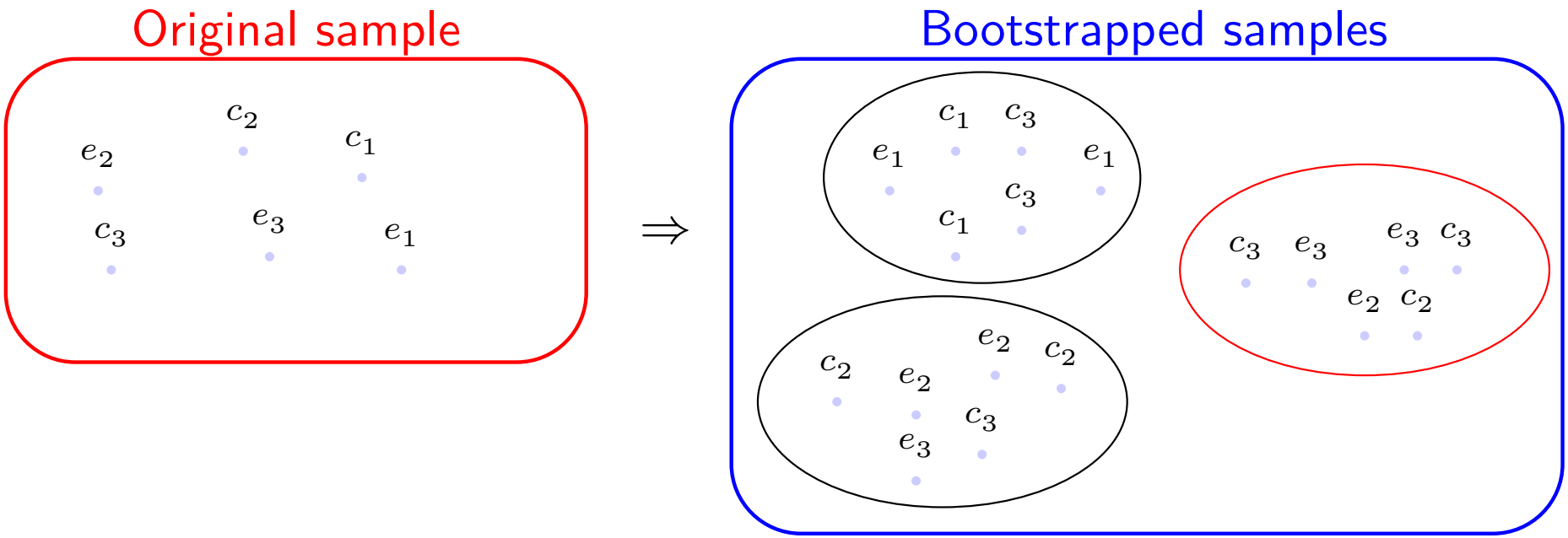



- Original sample $\approx$ population
- Original sample size $n = 6$
- Mean $\bar{e} = 0.5$ & $\bar{c} = 150$
- Sd $s_e = 0.25$ & $s_c = 50$
- Coeff $\hat{\beta}_1 = 0.15$ & $\hat{\alpha}_1 = 35$

- Bootstrap sample number $b = 2$
- Bootstrapped sample size $n^b = 6$
- Bootstrapped Mean $\bar{e}^b = 0.37$ & $\bar{c}^b = 130$
- Bootstrapped Sd $s_e^b = 0.13$ & $s_c^b = 42$
- Bootstrapped Coeff $\hat{\beta}_1^b = 0.18$ & $\hat{\alpha}_1^b = 25$

Maastricht University

# What is bootstrapping?

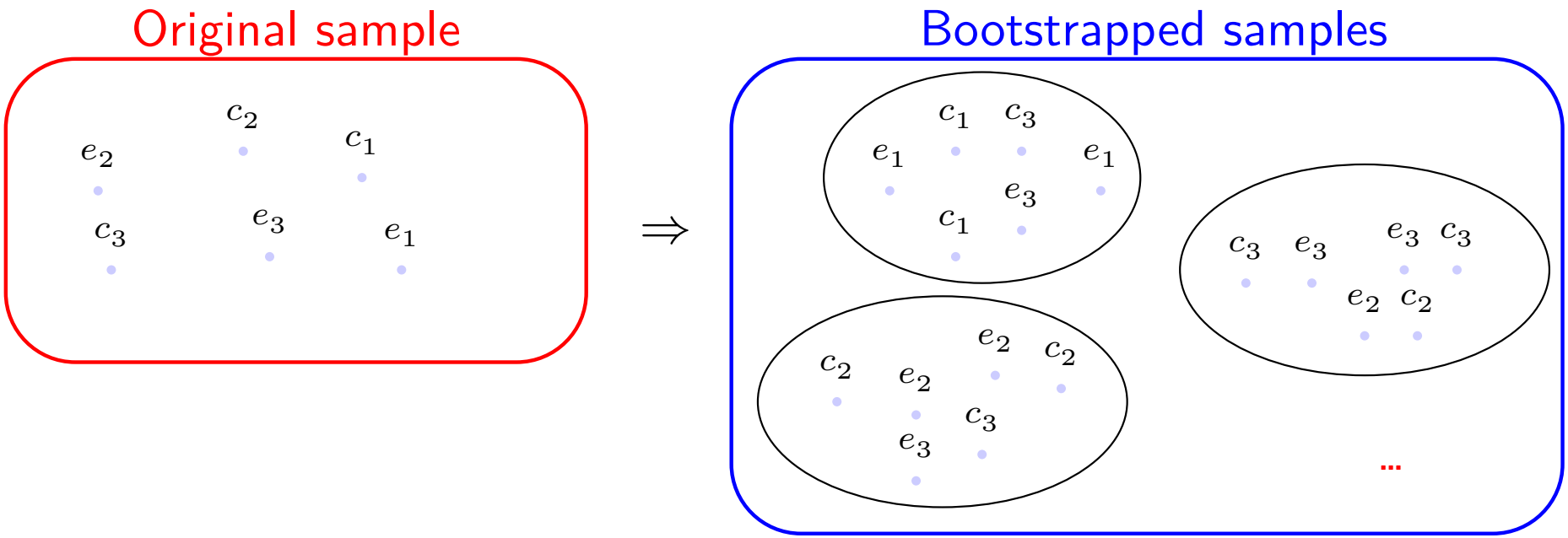



- Original sample $\approx$ population
- Original sample size $n = 6$
- Mean $\bar{e} = 0.5$ & $\bar{c} = 150$
- Sd $s_e = 0.25$ & $s_c = 50$
- Coeff $\hat{\beta}_1 = 0.15$ & $\hat{\alpha}_1 = 35$

- Bootstrap sample number $b = 3$
- Bootstrapped sample size $n^b = 6$
- Bootstrapped Mean $\bar{e}^b = 0.5$ & $\bar{c}^b = 155$
- Bootstrapped Sd $s_e^b = 0.14$ & $s_c^b = 32$
- Bootstrapped Coeff $\hat{\beta}_1^b = 0.15$ & $\hat{\alpha}_1^b = 33$

Maastricht University

# What is bootstrapping?

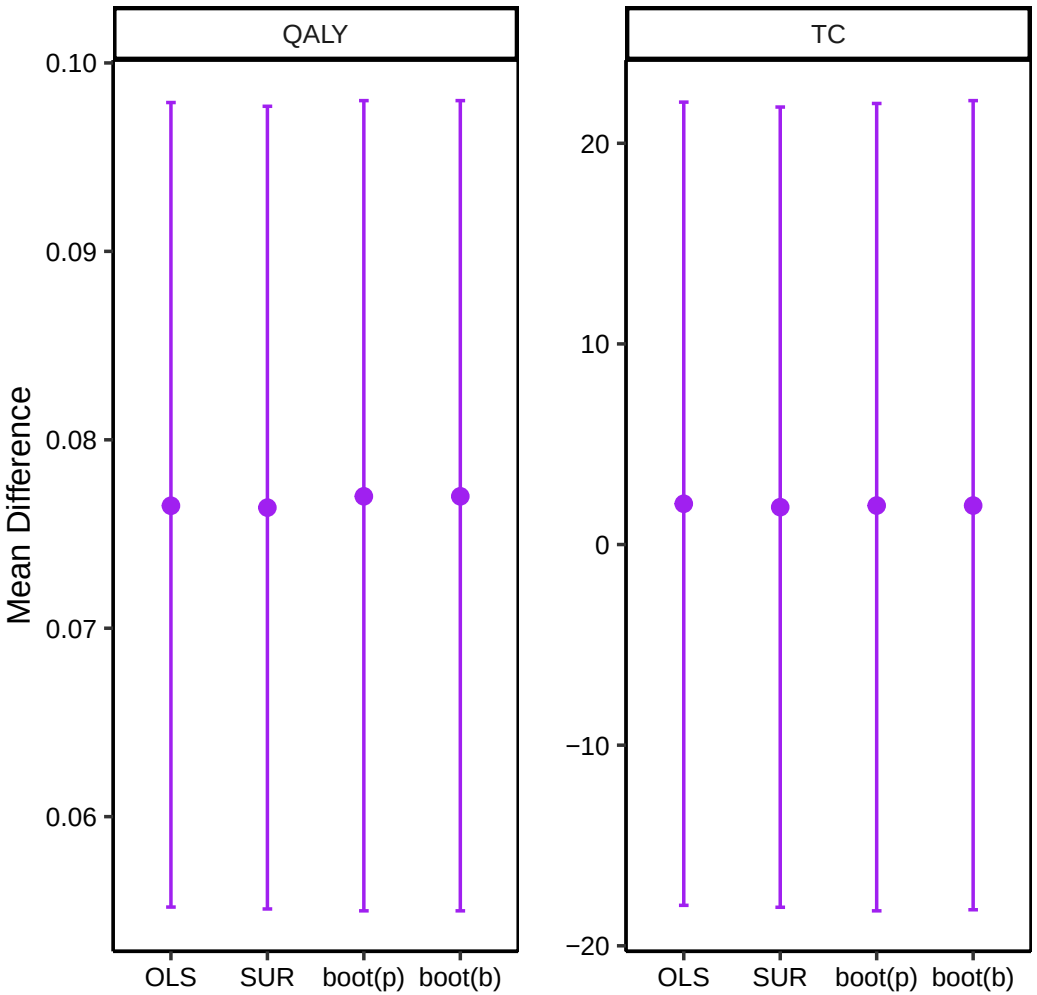



- Original sample $\approx$ population
- Original sample size $n = 6$
- Mean $\bar{e} = 0.5$ & $\bar{c} = 150$
- Sd $s_e = 0.25$ & $s_c = 50$
- Coeff $\hat{\beta}_1 = 0.15$ & $\hat{\alpha}_1 = 35$

- Iterate the process $b = 4, \ldots, B$ times
- Quantify uncertainty of the estimators (eg CIs) using all bootstrapped samples

Maastricht University

# Correlation - comparison of approaches




- **Mean QALY difference $\Delta_e$**

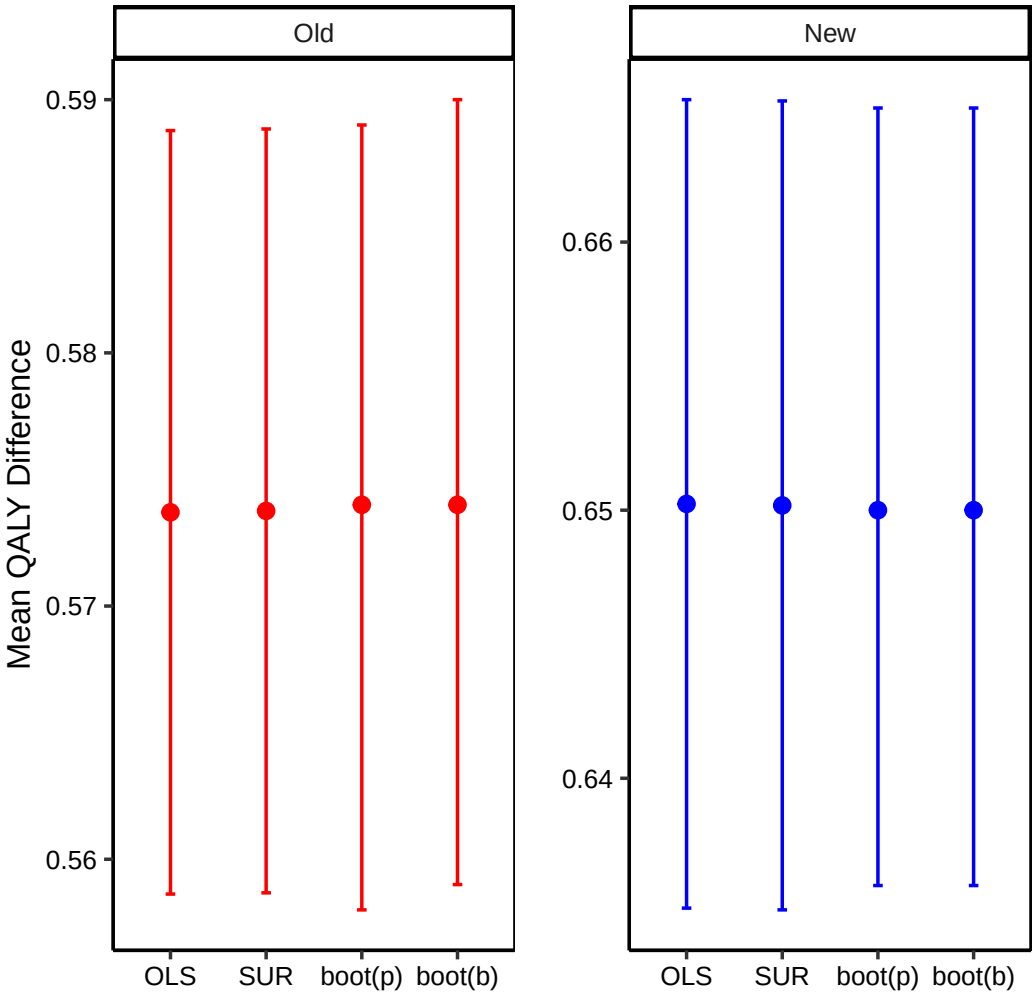
| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.0765*** | *0.0552* | *0.0979* | *OLS* |
| ***0.0764*** | *0.0551* | *0.0977* | *SUR* |
| ***0.0770*** | *0.0550* | *0.0980* | *boot(p)* |
| ***0.0770*** | *0.0550* | *0.0980* | *boot(b)* |

- **Mean TC difference $\Delta_c$**

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***2.033*** | *-17.984* | *22.051* | *OLS* |
| ***1.862*** | *-18.080* | *21.805* | *SUR* |
| ***1.943*** | *-18.258* | *21.983* | *boot(p)* |
| ***1.943*** | *-18.204* | *22.125* | *boot(b)* |

Maastricht University

# Correlation - comparison of approaches




- **Mean QALY** for arm = "Old"

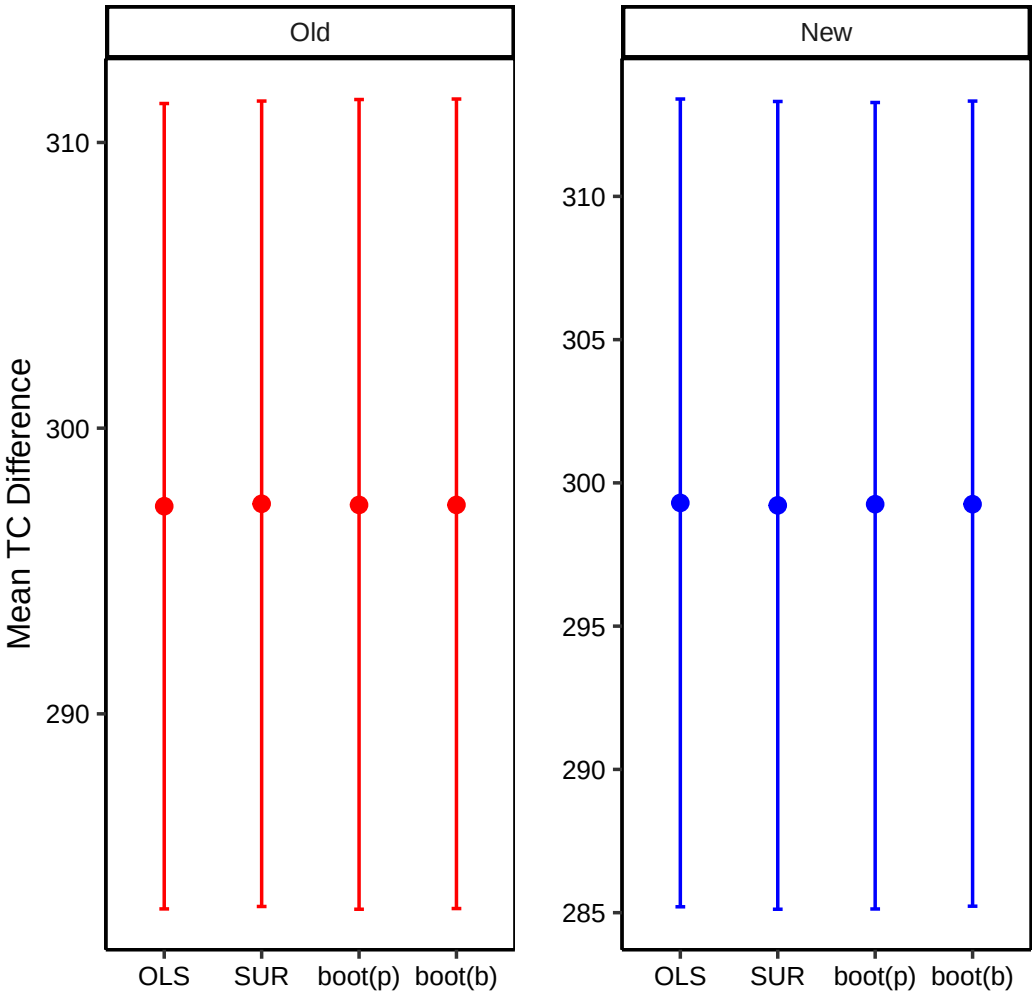
| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.5737*** | *0.5586* | *0.5888* | *OLS* |
| ***0.5738*** | *0.5587* | *0.5888* | *SUR* |
| ***0.5740*** | *0.5580* | *0.5890* | *boot(p)* |
| ***0.5740*** | *0.5590* | *0.5900* | *boot(b)* |

- **Mean QALY** for arm = "New"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.6502*** | *0.6352* | *0.6653* | *OLS* |
| ***0.6502*** | *0.6351* | *0.6653* | *SUR* |
| ***0.6500*** | *0.6360* | *0.6650* | *boot(p)* |
| ***0.6500*** | *0.6360* | *0.6650* | *boot(b)* |

Maastricht University

# Correlation - comparison of approaches




- **Mean TC** for arm = "Old"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***297.268*** | *283.172* | *311.364* | *OLS* |
| ***297.353*** | *283.257* | *311.449* | *SUR* |
| ***297.313*** | *283.160* | *311.503* | *boot(p)* |
| ***297.313*** | *283.184* | *311.520* | *boot(b)* |

- **Mean TC** for arm = "New"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***299.301*** | *285.205* | *313.397* | *OLS* |
| ***299.215*** | *285.119* | *313.312* | *SUR* |
| ***299.256*** | *285.129* | *313.276* | *boot(p)* |
| ***299.256*** | *285.226* | *313.325* | *boot(b)* |

# Correlation - SUR/bootstrapping

- **SUR**:
  - Can be more **efficient** than separate OLS analyses
  - Allow inclusion of outcome-specific variables into the regressions
  - Similar **limitations** of standard regression

# Correlation - SUR/bootstrapping

- **SUR**:
  - Can be more **efficient** than separate OLS analyses
  - Allow inclusion of outcome-specific variables into the regressions
  - Similar **limitations** of standard regression
- (non-parametric) **Bootstrapping**:
  - Paired re-sampling can **maintain** CE correlation
  - Performance wrt joint models has **not** been fully assessed
  - **Different ways** to compute CIs (percentile, bias-corrected and accelerated, ...)

Maastricht University

# Correlation - SUR/bootstrapping

- **SUR**:
  - Can be more **efficient** than separate OLS analyses
  - Allow inclusion of outcome-specific variables into the regressions
  - Similar **limitations** of standard regression
- (non-parametric) **Bootstrapping**:
  - Paired re-sampling can **maintain** CE correlation
  - Performance wrt joint models has **not** been fully assessed
  - **Different ways** to compute CIs (percentile, bias-corrected and accelerated, ...)
- **Other approaches**:
  - *Bayesian joint models* (Gabrio et al. 2019)

Maastricht University

# Skewness in CE outcomes

- Costs are typically **right-skewed** while utilities tend to be **left-skewed**:
  - Costs are bound at $0$ with few participants with very high costs
  - Utilities are bound at $1$ with few participants with very low (also negative) values

# Skewness in CE outcomes

- Costs are typically **right-skewed** while utilities tend to be **left-skewed**:
  - Costs are bound at $0$ with few participants with very high costs
  - Utilities are bound at $1$ with few participants with very low (also negative) values
- Many methods rely on **Normality** assumptions that may be violated and lead to misleading inferences:
  - In "large" samples, **CLT** ensures that estimates are unbiased but *efficiency* may not be achieved
  - In "small" samples, *unbiasdness* is no longer guaranteed

Maastricht University

# Skewness in CE outcomes

- Costs are typically **right-skewed** while utilities tend to be **left-skewed**:
  - Costs are bound at $0$ with few participants with very high costs
  - Utilities are bound at $1$ with few participants with very low (also negative) values
- Many methods rely on **Normality** assumptions that may be violated and lead to misleading inferences:
  - In "large" samples, **CLT** ensures that estimates are unbiased but *efficiency* may not be achieved
  - In "small" samples, *unbiasdness* is no longer guaranteed
- **No general consensus** at which point:
  - The degree of skewness is "too high"
  - The size of the sample is "too small"

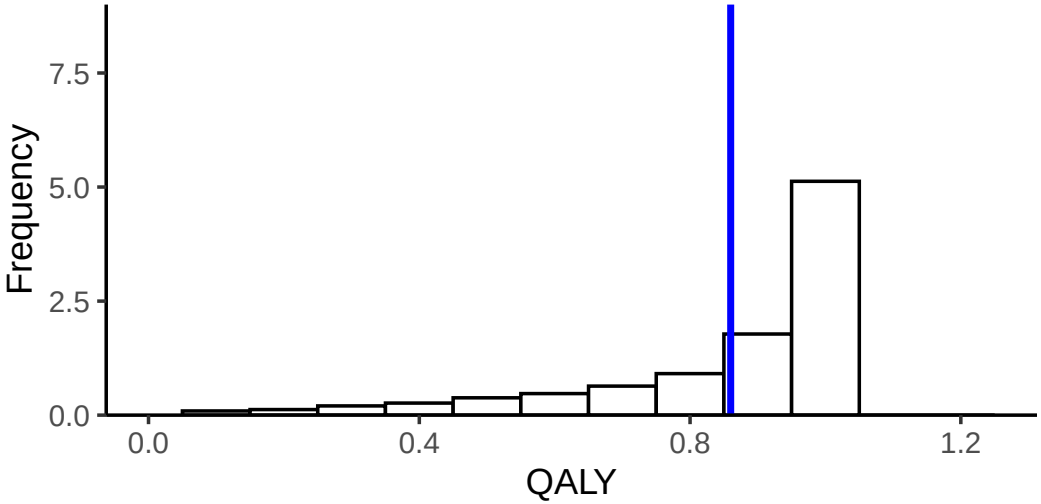
Maastricht University

# Skewness in CE outcomes

- Costs are typically **right-skewed** while utilities tend to be **left-skewed**:
  - Costs are bound at $0$ with few participants with very high costs
  - Utilities are bound at $1$ with few participants with very low (also negative) values
- Many methods rely on **Normality** assumptions that may be violated and lead to misleading inferences:
  - In "large" samples, **CLT** ensures that estimates are unbiased but *efficiency* may not be achieved
  - In "small" samples, *unbiasdness* is no longer guaranteed
- **No general consensus** at which point:
  - The degree of skewness is "too high"
  - The size of the sample is "too small"
- Data *transformations* (eg log) are **not appropriate** as they do not provide inferences about population means of interest

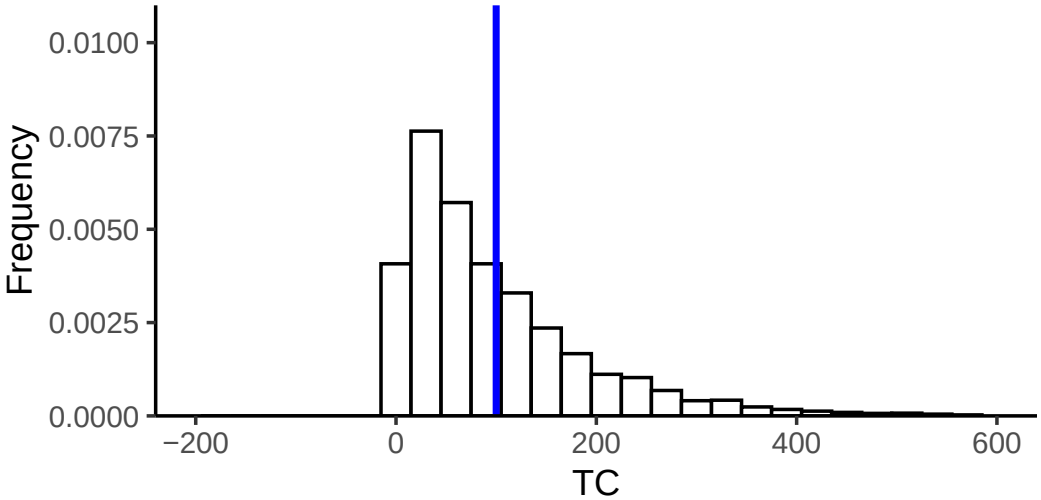
Maastricht University

# Skewness in CE outcomes - an example




- QALYs in **large** samples

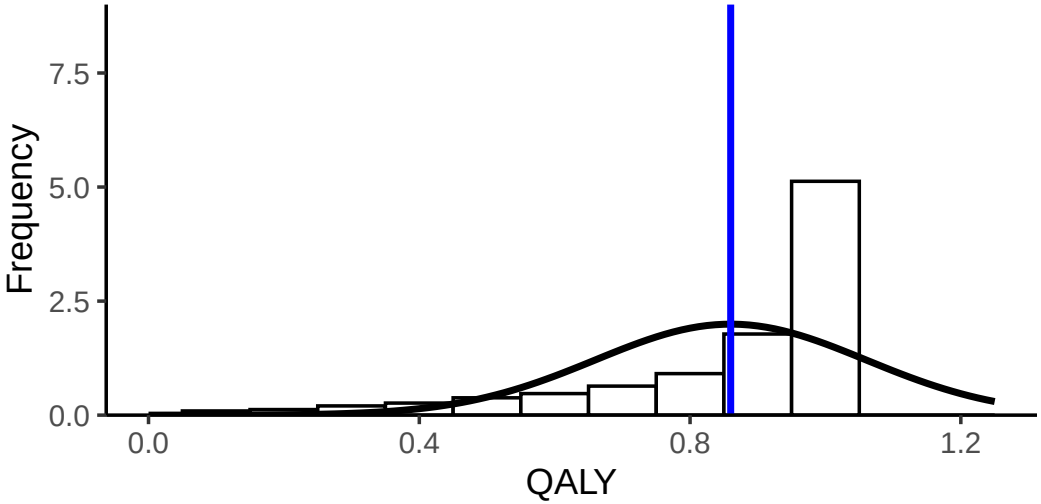
| Mean | Sd | CI(low) | CI(high) | size |
|---|---|---|---|---|
| ***0.86*** | *0.2* | *0.266* | *1* | *5000* |




- TC in **large** samples

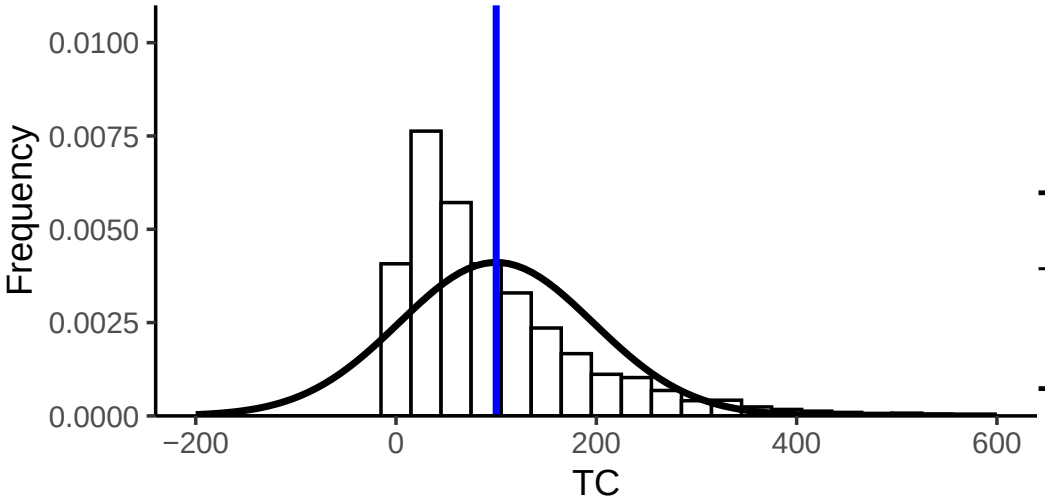
| Mean | Sd | CI(low) | CI(high) | size |
|---|---|---|---|---|
| ***100*** | *97* | *2.67* | *350.153* | *5000* |

# Skewness in CE outcomes - an example




- **CLT** in **large** samples

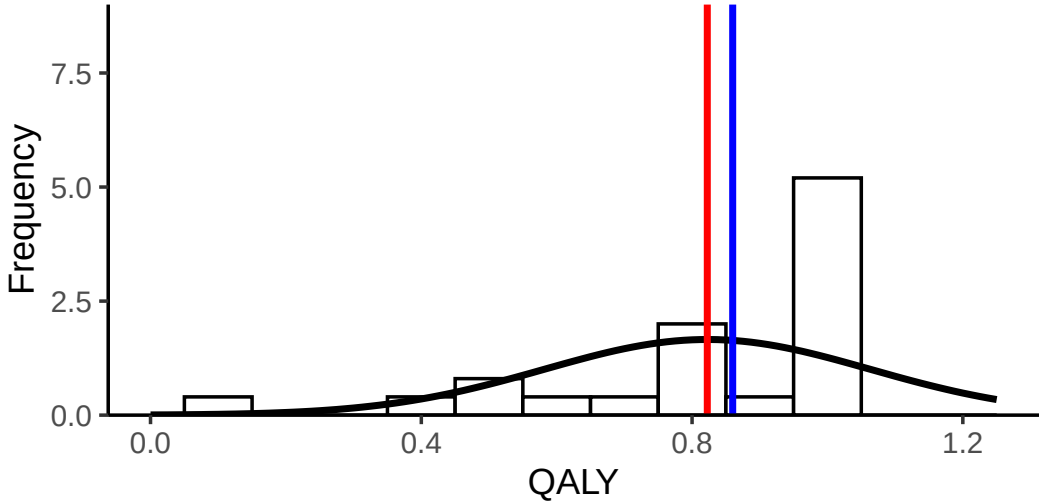
| Mean | Sd | CI(low) | CI(high) | size |
|---|---|---|---|---|
| ***0.86*** | *0.200* | *0.266* | *1.00* | *5000* |
| **0.86** | 0.003 | 0.860 | 0.86 | 5000 |




- **CLT** **large** samples

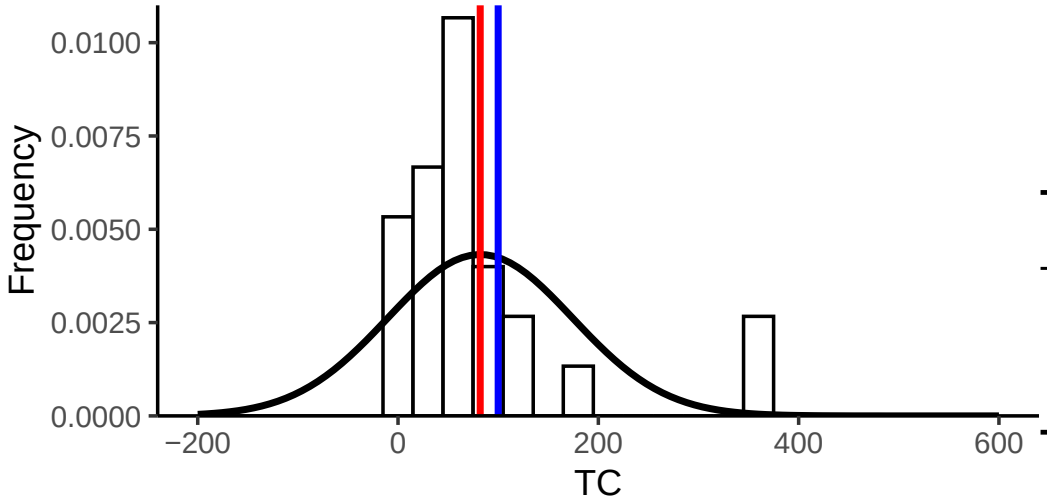
| Mean | Sd | CI(low) | CI(high) | size |
|---|---|---|---|---|
| ***100.000*** | *97.000* | *2.67* | *350.153* | *5000* |
| **99.999** | 1.372 | 99.96 | 100.034 | 5000 |

# Skewness in CE outcomes - an example




- **CLT** in **small** samples

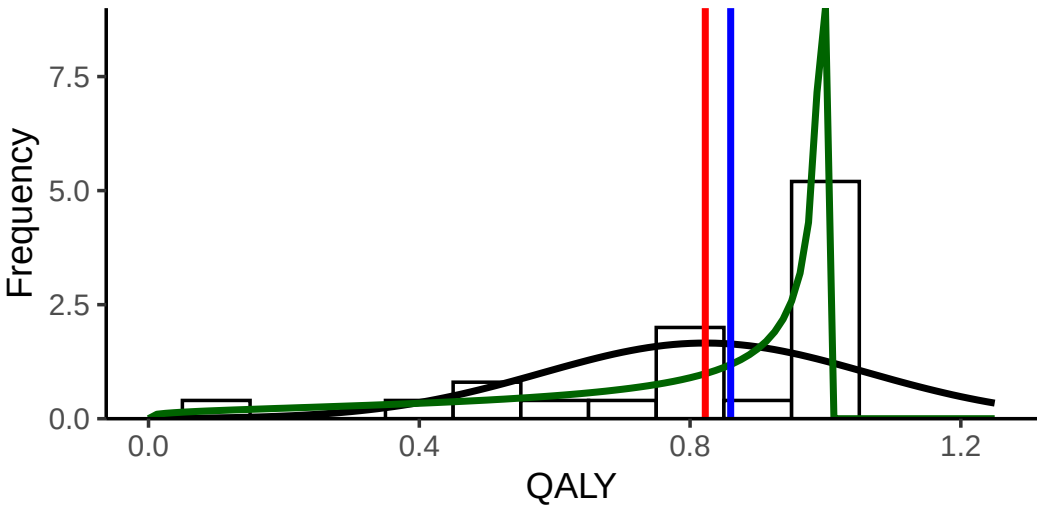
| Mean | Sd | CI(low) | CI(high) | size |
|---|---|---|---|---|
| ***0.860*** | *0.200* | *0.266* | *1.000* | *5000* |
| ***0.822*** | *0.241* | *0.266* | *1.000* | *25* |
| **0.822** | 0.040 | 0.822 | 0.823 | 25 |




- **CLT** in **small** samples

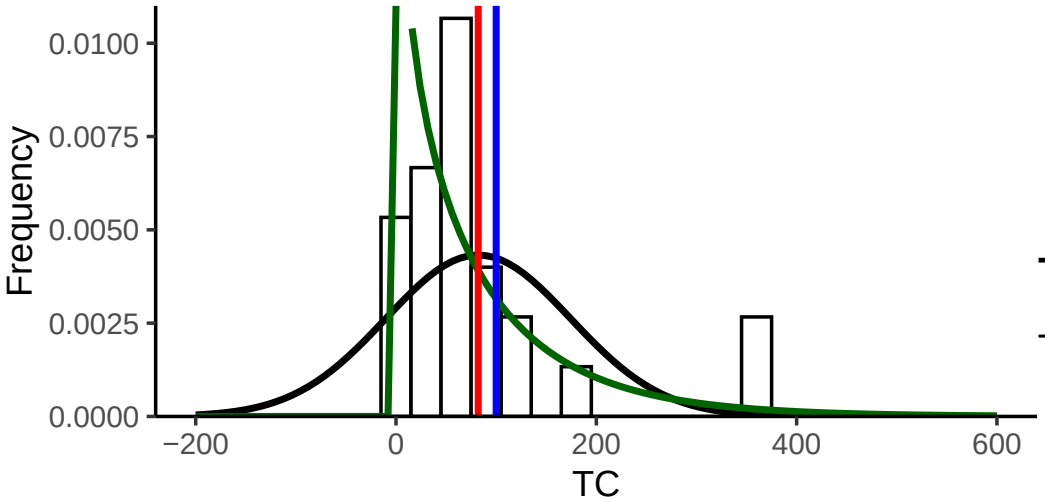
| Mean | Sd | CI(low) | CI(high) | size |
|---|---|---|---|---|
| ***100.000*** | *97.000* | *2.670* | *350.153* | *5000* |
| ***82.062*** | *92.265* | *6.047* | *356.448* | *25* |
| **82.063** | 19.400 | 82.025 | 82.098 | 25 |

# Skewness in CE outcomes - an example




- Other distribution (**Beta**) in **small** samples

| Mean | Sd | CI(low) | CI(high) | size |
|---|---|---|---|---|
| ***0.860*** | *0.200* | *0.266* | *1.000* | *5000* |
| ***0.822*** | *0.241* | *0.266* | *1.000* | *25* |
| **0.813** | 0.238 | 0.362 | 1.295 | 25 |
| ***0.818*** | *0.245* | *0.157* | *1.000* | *25* |




- Other distribution (**Gamma**) in **small** samples

| Mean | Sd | CI(low) | CI(high) | size |
|---|---|---|---|---|
| ***100.000*** | *97.000* | *2.670* | *350.153* | *5000* |
| ***82.062*** | *92.265* | *6.047* | *356.448* | *25* |
| **84.503** | 89.313 | -89.825 | 263.089 | 25 |
| ***84.592*** | *98.106* | *0.886* | *351.883* | *25* |

# Skewness - alternative methods

- OLS/SUR *ignoring skewness* and rely on **CLT/bootstrapping**:

# Skewness - alternative methods

- OLS/SUR *ignoring skewness* and rely on **CLT/bootstrapping**:
- **Generalised Linear Model** (GLM):

$$\mathsf{E}[\mathsf{QALY}_i] = g_e(\beta_0 + \beta_1 \times \mathsf{arm}_i + \beta_2 \times u_{i0})^{-1}$$
$$\mathsf{E}[\mathsf{TC}_i] = g_c(\alpha_0 + \alpha_1 \times \mathsf{arm}_i + \alpha_2 \times c_{i0})^{-1}$$

Maastricht University

# Skewness - alternative methods

- OLS/SUR *ignoring skewness* and rely on **CLT/bootstrapping**:
- **Generalised Linear Model** (GLM):

$$\mathsf{E}[\mathsf{QALY}_i] = g_e(\beta_0 + \beta_1 \times \mathsf{arm}_i + \beta_2 \times u_{i0})^{-1}$$
$$\mathsf{E}[\mathsf{TC}_i] = g_c(\alpha_0 + \alpha_1 \times \mathsf{arm}_i + \alpha_2 \times c_{i0})^{-1}$$

- Choose **link functions** to model the *expected values*:
  - $g_e(\cdot)^{-1} = \mathsf{logit}()$ & $g_c(\cdot)^{-1} = \mathsf{log}()$

Maastricht University

# Skewness - alternative methods

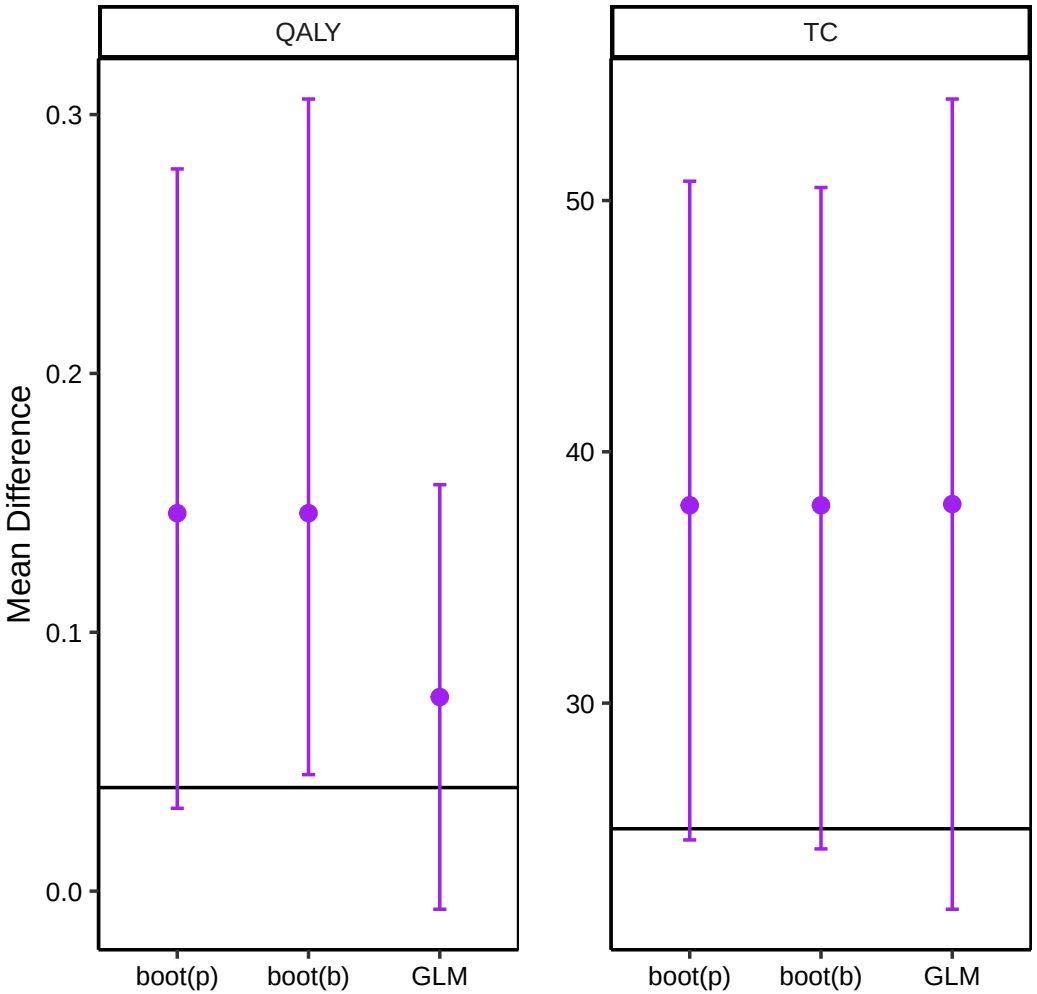
- OLS/SUR *ignoring skewness* and rely on **CLT/bootstrapping**:
- **Generalised Linear Model** (GLM):

$$\mathsf{E}[\mathsf{QALY}_i] = g_e(\beta_0 + \beta_1 \times \mathsf{arm}_i + \beta_2 \times u_{i0})^{-1}$$
$$\mathsf{E}[\mathsf{TC}_i] = g_c(\alpha_0 + \alpha_1 \times \mathsf{arm}_i + \alpha_2 \times c_{i0})^{-1}$$

- Choose **link functions** to model the *expected values*:
  - $g_e(\cdot)^{-1} = \mathsf{logit}()$ & $g_c(\cdot)^{-1} = \mathsf{log}()$
- Choose **alternative distributions** for the *error terms*:
  - $\varepsilon_{ie} \sim \mathsf{Beta}()$ & $\varepsilon_{ic} \sim \mathsf{Gamma}()$

Maastricht University

# Skewness - alternative methods



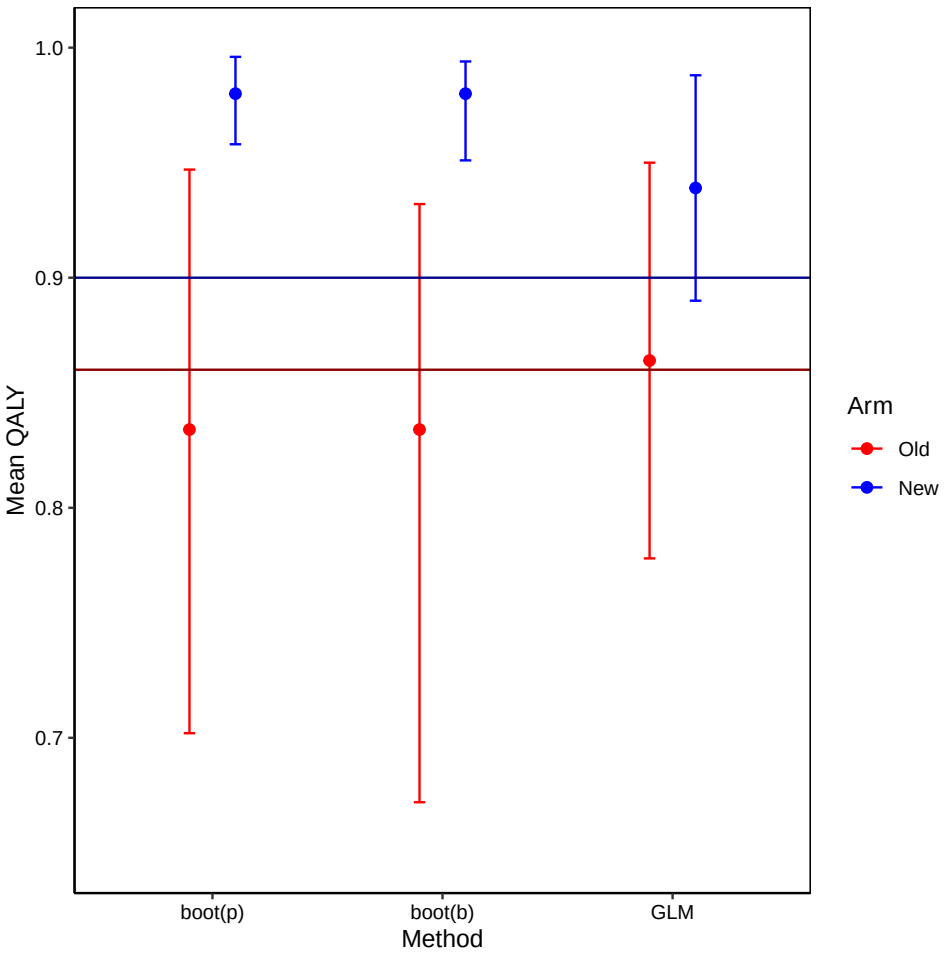

- **Mean QALY difference $\Delta_e$**

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.146*** | *0.032* | *0.279* | *boot(p)* |
| ***0.146*** | *0.045* | *0.306* | *boot(b)* |
| ***0.075*** | *-0.007* | *0.157* | *GLM* |

- **Mean TC difference $\Delta_c$**

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***37.872*** | *24.555* | *50.768* | *boot(p)* |
| ***37.872*** | *24.198* | *50.517* | *boot(b)* |
| ***37.918*** | *21.797* | *54.038* | *GLM* |

# Skewness - alternative methods



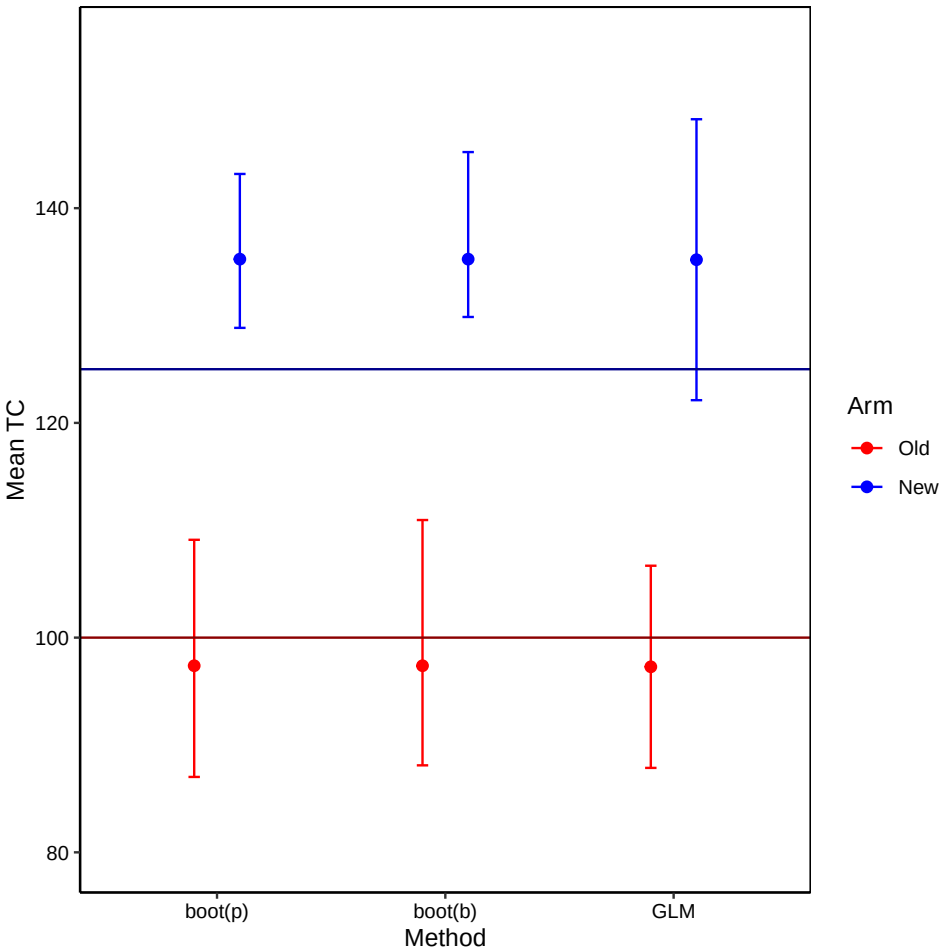

- **Mean QALY** for arm = "Old"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.834*** | *0.702* | *0.947* | *boot(p)* |
| ***0.834*** | *0.672* | *0.932* | *boot(b)* |
| ***0.864*** | *0.778* | *0.950* | *GLM* |

- **Mean QALY** for arm = "New"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.980*** | *0.958* | *0.996* | *boot(p)* |
| ***0.980*** | *0.951* | *0.994* | *boot(b)* |
| ***0.939*** | *0.890* | *0.988* | *GLM* |

# Skewness - alternative methods




- **Mean TC** for arm = "Old"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***97.381*** | *87.023* | *109.112* | *boot(p)* |
| ***97.381*** | *88.104* | *110.953* | *boot(b)* |
| ***97.282*** | *87.866* | *106.697* | *GLM* |

- **Mean TC** for arm = "New"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***135.253*** | *128.852* | *143.187* | *boot(p)* |
| ***135.253*** | *129.866* | *145.226* | *boot(b)* |
| ***135.199*** | *122.114* | *148.284* | *GLM* |

Maastricht University

# Skewness - GLM/bootstrapping

- **GLM**:
  - Allow comparison of means on **natural scale**
  - Can handle **skewness** also in "small" samples
  - The choice of **distribution** and **link function** crucial (not always easy)

# Skewness - GLM/bootstrapping

- **GLM**:
  - Allow comparison of means on **natural scale**
  - Can handle **skewness** also in "small" samples
  - The choice of **distribution** and **link function** crucial (not always easy)
- (non-parametric) **Bootstrapping**:
  - Allow comparison of means on **natural scale**
  - Does not assume specific **distributions**
  - Good performance in **"large"** samples but **poor** in **"small"** samples
  - **Different ways** to compute CIs (percentile, bias-corrected and accelerated, ...)

Maastricht University

# Skewness - GLM/bootstrapping

- **GLM**:
  - Allow comparison of means on **natural scale**
  - Can handle **skewness** also in "small" samples
  - The choice of **distribution** and **link function** crucial (not always easy)
- (non-parametric) **Bootstrapping**:
  - Allow comparison of means on **natural scale**
  - Does not assume specific **distributions**
  - Good performance in **"large"** samples but **poor** in **"small"** samples
  - **Different ways** to compute CIs (percentile, bias-corrected and accelerated, ...)
- **Other approaches**:
  - *Bayesian joint models* (Gabrio et al. 2019, )
  - *Hurdle models* (Gabrio et al. 2019, Lambert et al. 2008)

Maastricht University

# Clustering of data

- Trial participants may be **clustered** (eg practices/hospitals) and *clusters may be randomised instead of individuals*
  - Typical of **cluster RCTs** where patients in a cluster are randomised to different treatments
  - Individual effects & costs in the same cluster tend to be more *homogeneous* than those in different clusters

Maastricht University

# Clustering of data

- Trial participants may be **clustered** (eg practices/hospitals) and *clusters may be randomised instead of individuals*
  - Typical of **cluster RCTs** where patients in a cluster are randomised to different treatments
  - Individual effects & costs in the same cluster tend to be more *homogeneous* than those in different clusters
- Standard methods (eg OLS, SUR & GLM) assume **independence** of observations but participants in same cluster are likely to be *associated with each other*
  - Ignoring clustering **underestimates** statistical uncertainty (eg variance)
  - may lead to **incorrect** inferences and **bias** the results

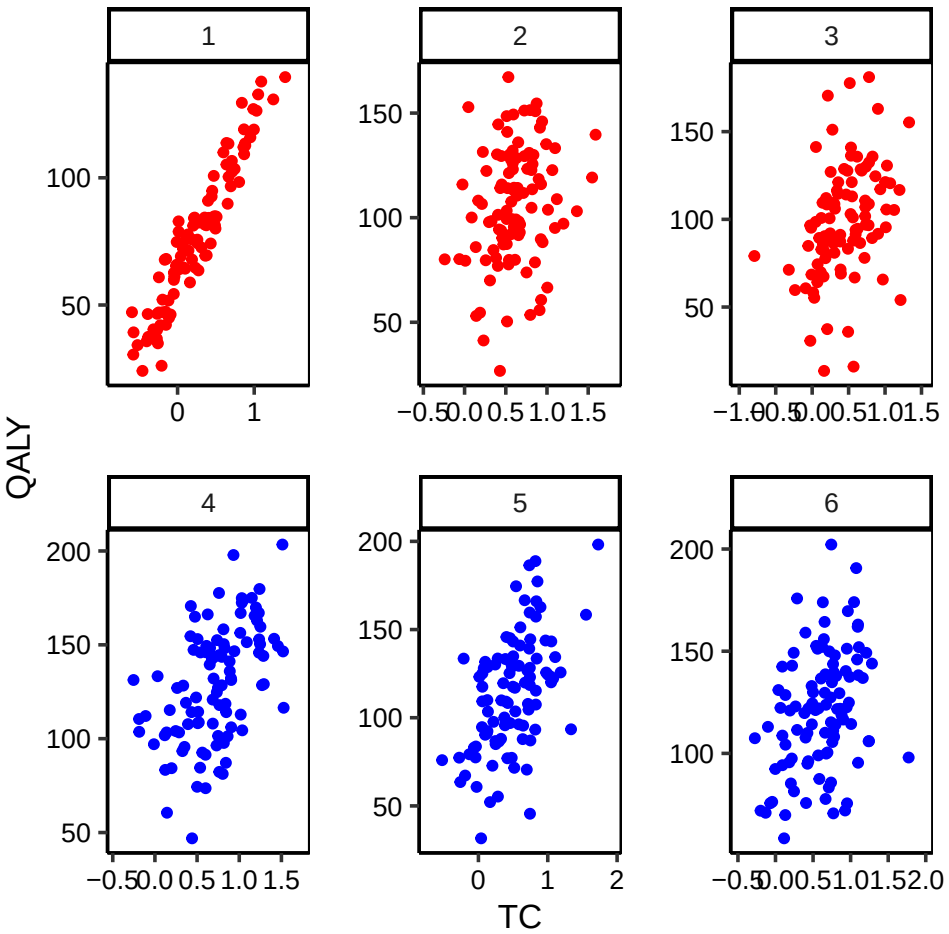
Maastricht University

# Clustering of data

- Trial participants may be **clustered** (eg practices/hospitals) and *clusters may be randomised instead of individuals*
  - Typical of **cluster RCTs** where patients in a cluster are randomised to different treatments
  - Individual effects & costs in the same cluster tend to be more *homogeneous* than those in different clusters
- Standard methods (eg OLS, SUR & GLM) assume **independence** of observations but participants in same cluster are likely to be *associated with each other*
  - Ignoring clustering **underestimates** statistical uncertainty (eg variance)
  - may lead to **incorrect** inferences and **bias** the results
- In CE studies, **additional challenges** need to be addressed:
  - Account for **correlation** between outcomes
  - Deal with **skewness** in both outcomes

Maastricht University

# Clustering of data - an example




- Costs & effects across **clusters**

| mean(e) | sd(e) | mean(c) | sd(c) | arm |
|---|---|---|---|---|
| *0.433* | *0.410* | *93.959* | *31.518* | *Old* |
| *0.604* | *0.407* | *121.437* | *31.525* | *New* |

- Costs & effects by **cluster**

| mean(e) | sd(e) | mean(c) | sd(c) | arm |
|---|---|---|---|---|
| *0.252* | *0.450* | *77.095* | *27.742* | *Old* |
| *0.608* | *0.326* | *106.259* | *28.088* | *Old* |
| *0.438* | *0.369* | *98.524* | *31.314* | *Old* |
| *0.718* | *0.402* | *129.361* | *30.897* | *New* |
| *0.489* | *0.401* | *114.054* | *32.540* | *New* |
| *0.606* | *0.387* | *120.896* | *29.484* | *New* |

Maastricht University

# Clustering - alternative methods

- **Multilevel Model** (MLM):

1. Add **"random" terms** $(u_{je}, u_{jc})$ to OLS/SUR model to capture differences between $j$-th cluster QALY/TC means from overall means in each arm

$$\text{QALY}_{ij} = \beta_0 + \beta_1 \times \text{arm}_j + u_{je} + \varepsilon_{ije}$$
$$\text{TC}_{ij} = \alpha_0 + \alpha_1 \times \text{arm}_j + u_{jc} + \varepsilon_{ijc}$$

Maastricht University

# Clustering - alternative methods

- **Multilevel Model** (MLM):

1. Add **"random" terms** $(u_{je}, u_{jc})$ to OLS/SUR model to capture differences between $j$-th cluster QALY/TC means from overall means in each arm

$$\begin{aligned} \text{QALY}_{ij} &= \beta_0 + \beta_1 \times \text{arm}_j + u_{je} + \varepsilon_{ije} \\ \text{TC}_{ij} &= \alpha_0 + \alpha_1 \times \text{arm}_j + u_{jc} + \varepsilon_{ijc} \end{aligned}$$

2. Model cluster-specific error terms using some *distribution* (eg Normal) and **link them** through the parameter $\psi$ (often assumed $0$)

$$\begin{pmatrix} u_{je} \\ u_{jc} \end{pmatrix} \sim \text{Normal}\left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \tau_e^2 & \psi\tau_e\tau_c \\ \psi\tau_c\tau_e & \tau_c^2 \end{pmatrix} \right]$$

# Clustering - alternative methods

- (non-parametric) **Two-Stage Bootstrap** (TSB):

1 **Two-stage routine** to account for data clustering:

  - First **sample clusters** then **individuals** within each resampled cluster
  - Analyse CE outcomes and derive $(\hat{\beta}, \hat{\alpha})^b$ –> $(\Delta_e, \Delta_c)^b$
  - Iterate the process a *sufficiently* large number of times (B)
  - Use the estimates $(\Delta_e, \Delta_c)^b$ to quantify uncertainty (eg obtain CIs)

Maastricht University

# Clustering - alternative methods

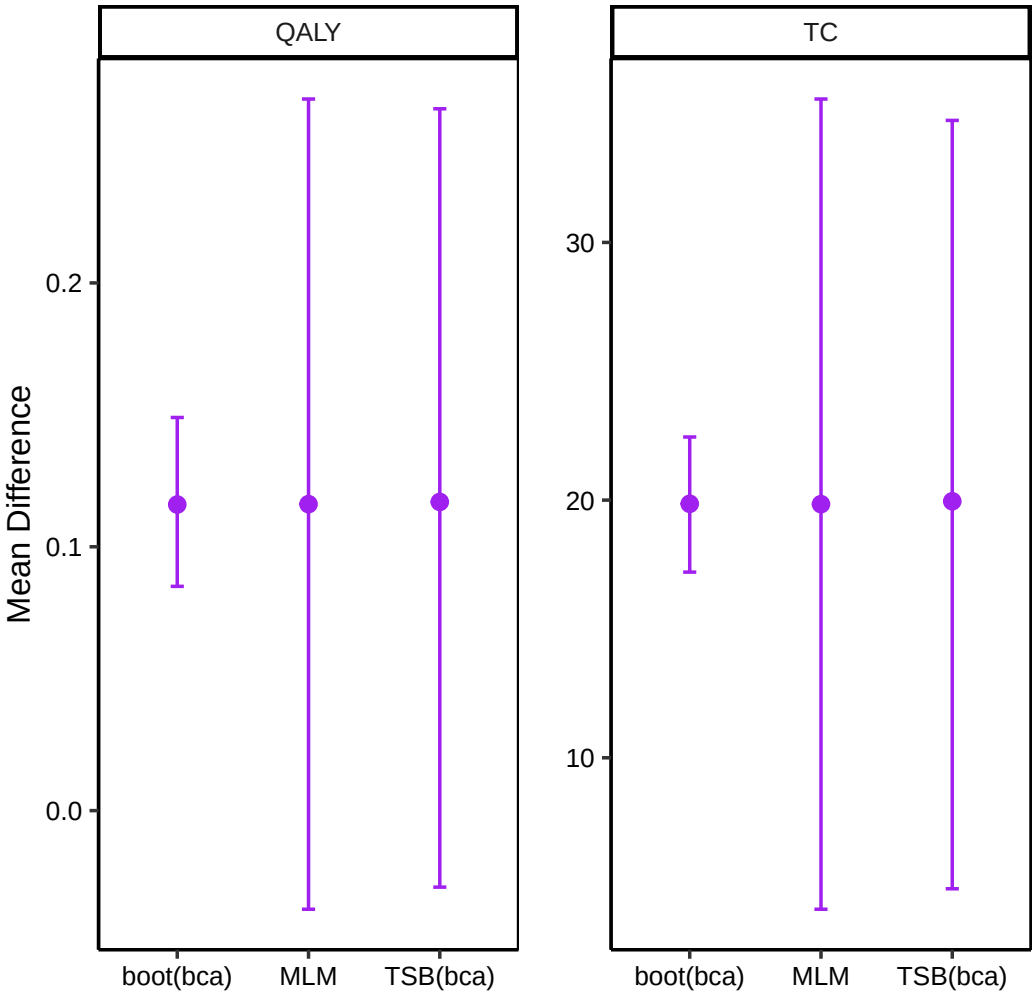
- (non-parametric) **Two-Stage Bootstrap** (TSB):

1 **Two-stage routine** to account for data clustering:
   - First **sample clusters** then **individuals** within each resampled cluster
   - Analyse CE outcomes and derive $(\hat{\beta}, \hat{\alpha})^b$ –> $(\Delta_e, \Delta_c)^b$
   - Iterate the process a *sufficiently* large number of times (B)
   - Use the estimates $(\Delta_e, \Delta_c)^b$ to quantify uncertainty (eg obtain CIs)

2 Apply **"shrinkage"** correction to avoid **overestimation** of variance
   - Before any resampling, **cluster means** are *shrunk* and **individual residuals** estimated from cluster means
   - Sample **shrunken cluster means** then **individual residuals**
   - Combine shrunken means and residuals to obtain bootstrap dataset

Maastricht University

# Clustering - alternative methods



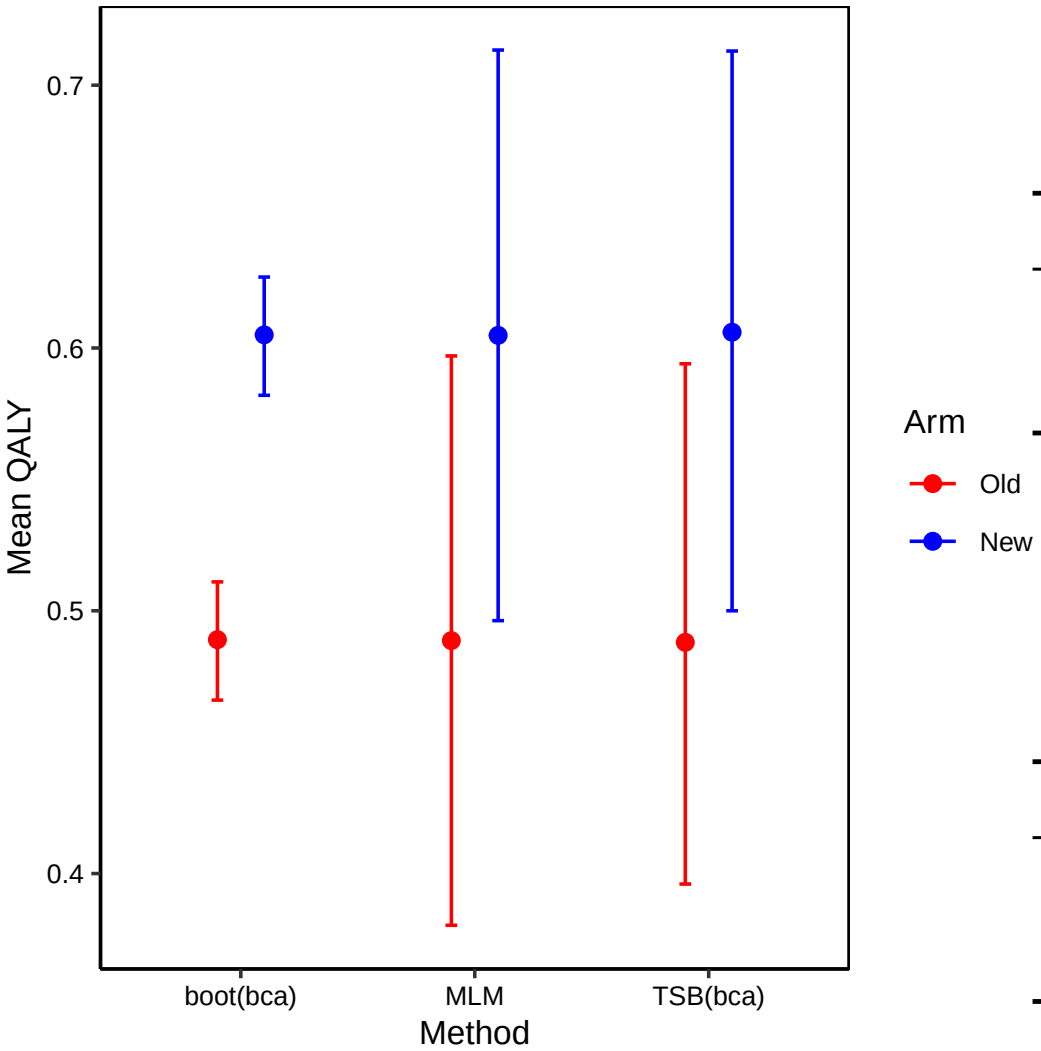

- **Mean QALY difference $\Delta_e$**

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.116*** | *0.085* | *0.149* | *boot(bca)* |
| ***0.116*** | *-0.037* | *0.270* | *MLM* |
| ***0.117*** | *-0.029* | *0.266* | *TSB(bca)* |

- **Mean TC difference $\Delta_c$**

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| *19.857* | ***17.205*** | ***22.449*** | *boot(bca)* |
| *19.843* | ***4.122*** | ***35.565*** | *MLM* |
| *19.955* | ***4.916*** | ***34.736*** | *TSB(bca)* |

# Clustering - alternative methods




- **Mean QALY** for arm = "Old"

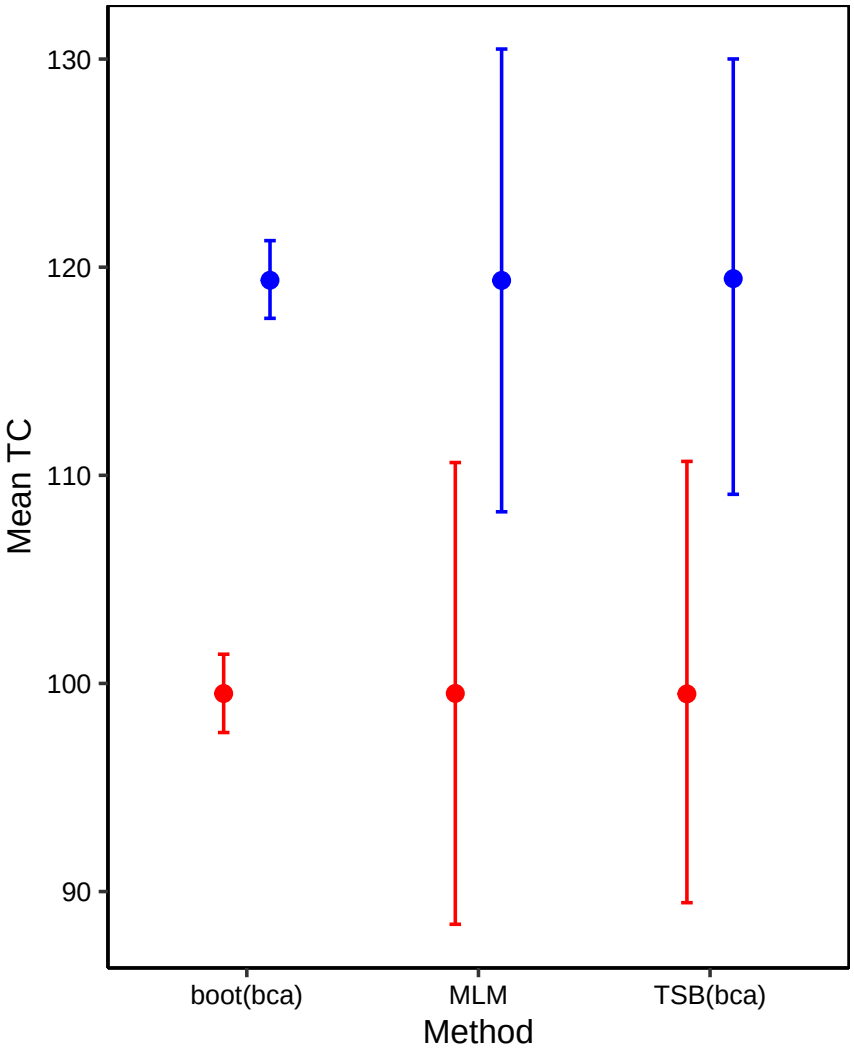
| Estimate | CI.low | CI.high | method | arm |
|---|---|---|---|---|
| *0.489* | ***0.466*** | ***0.511*** | *boot(bca)* | *Old* |
| *0.489* | ***0.380*** | ***0.597*** | *MLM* | *Old* |
| *0.488* | ***0.396*** | ***0.594*** | *TSB(bca)* | *Old* |

- **Mean QALY** for arm = "New"

| Estimate | CI.low | CI.high | method | arm |
|---|---|---|---|---|
| *0.605* | ***0.582*** | ***0.627*** | *boot(bca)* | *New* |
| *0.605* | ***0.496*** | ***0.713*** | *MLM* | *New* |
| *0.606* | ***0.500*** | ***0.713*** | *TSB(bca)* | *New* |

Maastricht University

# Clustering - alternative methods




- **Mean TC** for arm = "Old"

| Estimate | CI.low | CI.high | method | arm |
|---|---|---|---|---|
| *99.516* | ***97.642*** | ***101.403*** | *boot(bca)* | *Old* |
| *99.520* | ***88.427*** | ***110.614*** | *MLM* | *Old* |
| *99.497* | ***89.465*** | ***110.668*** | *TSB(bca)* | *Old* |

- **Mean TC** for arm = "New"

| Estimate | CI.low | CI.high | method | arm |
|---|---|---|---|---|
| *119.373* | ***117.542*** | ***121.276*** | *boot(bca)* | *New* |
| *119.364* | ***108.247*** | ***130.481*** | *MLM* | *New* |
| *119.451* | ***109.087*** | ***130.005*** | *TSB(bca)* | *New* |

# Clustering - MLM/TSB

- **MLM**:
  - Extends standard OLS/SUR framework to deal with **clustering**
  - Good performance except with **small samples** & **numbers of clusters**
  - Difficult to implement unless assuming **independence** between outcomes and Normality

# Clustering - MLM/TSB

- **MLM**:
  - Extends standard OLS/SUR framework to deal with **clustering**
  - Good performance except with **small samples** & **numbers of clusters**
  - Difficult to implement unless assuming **independence** between outcomes and Normality
- (non-parametric) **TSB**:
  - Standard bootstrap **ignores** clustering and **underestimates** uncertainty
  - **Two-stage procedure** needed to account for clustering
  - Can be combined with SUR to deal with CE correlation
  - **Shrinkage correction** needed to avoid overestimation

Maastricht University

# Clustering - MLM/TSB

- **MLM**:
  - Extends standard OLS/SUR framework to deal with **clustering**
  - Good performance except with **small samples** & **numbers of clusters**
  - Difficult to implement unless assuming **independence** between outcomes and Normality
- (non-parametric) **TSB**:
  - Standard bootstrap **ignores** clustering and **underestimates** uncertainty
  - **Two-stage procedure** needed to account for clustering
  - Can be combined with SUR to deal with CE correlation
  - **Shrinkage correction** needed to avoid overestimation
- **Other approaches**:
  - *Bayesian hierarchical models* (Grieve et al. 2010, Gomes et al. 2012)

Maastricht University

# Missing data

- Missing CE values almost **inevitably** occur:
  - A single missing CE *component* leads to a missing *QALY/TC* value
  - Need to accept **inherent uncertainty** in analysis
  - Any method is based on some **untestable assumptions**

# Missing data

- Missing CE values almost **inevitably** occur:
  - A single missing CE *component* leads to a missing *QALY/TC* value
  - Need to accept **inherent uncertainty** in analysis
  - Any method is based on some **untestable assumptions**

- Express assumptions as **missingness mechanism** (Little et al. 2019):
  - **Missing Completely At Random** (MCAR): occur *randomly*
  - **Missing At Random** (MAR): due to some *observed variables*
  - **Missing Not At Random** (MNAR): due to some *unobserved variables*

Maastricht University

# Missing data

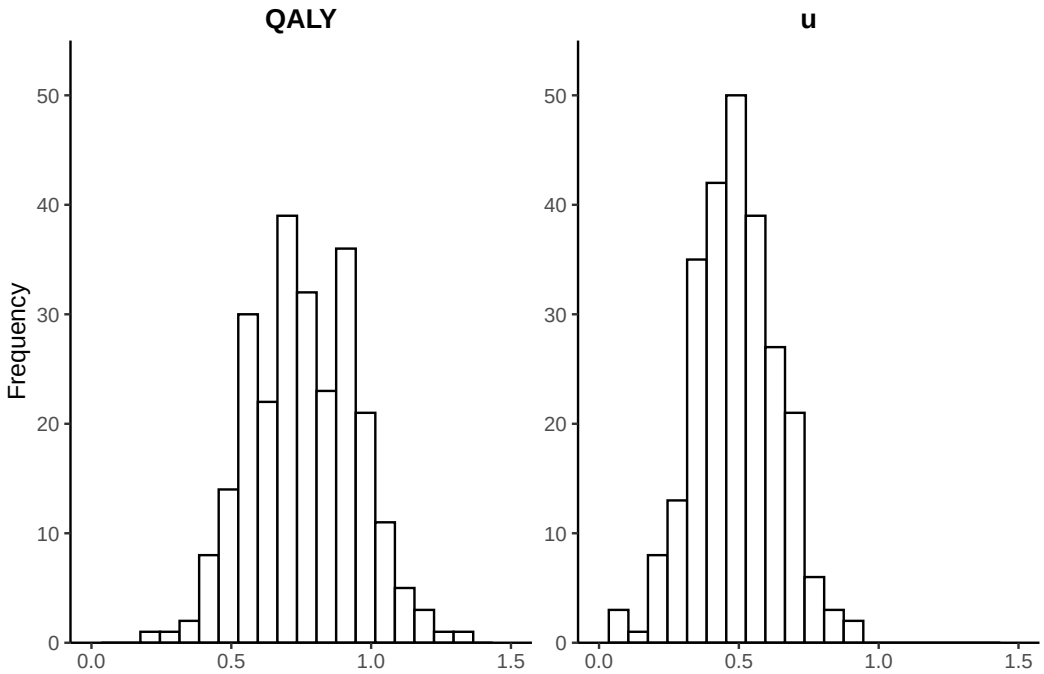
- Missing CE values almost **inevitably** occur:
  - A single missing CE *component* leads to a missing *QALY/TC* value
  - Need to accept **inherent uncertainty** in analysis
  - Any method is based on some **untestable assumptions**

- Express assumptions as **missingness mechanism** (Little et al. 2019):
  - **Missing Completely At Random** (MCAR): occur *randomly*
  - **Missing At Random** (MAR): due to some *observed variables*
  - **Missing Not At Random** (MNAR): due to some *unobserved variables*

- Method should be *aligned* with the **selected assumptions** :
  - Avoid methods with **unrealistic** assumptions (eg mean imputation)
  - Define a *base-case* scenario (eg MAR)
  - Assess **sensitivity** of results to some *departures*

Maastricht University

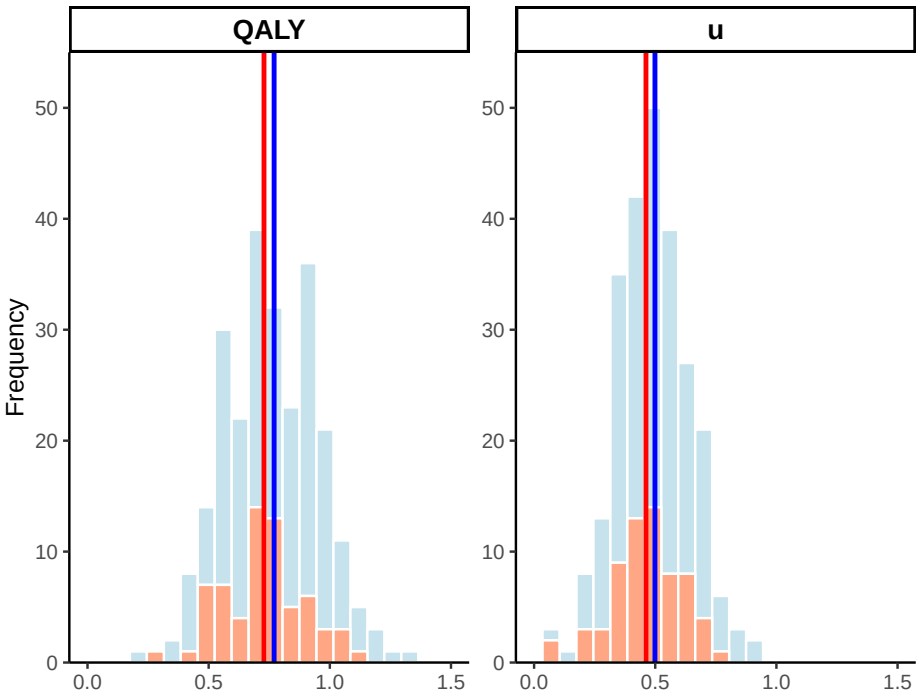
# Missing data - an example




- QALYs & utilities **fully observed**

| Mean(e) | Sd(e) | Mean(u) | Sd(u) | n |
|---|---|---|---|---|
| **0.759** | 0.189 | **0.489** | 0.147 | *250* |

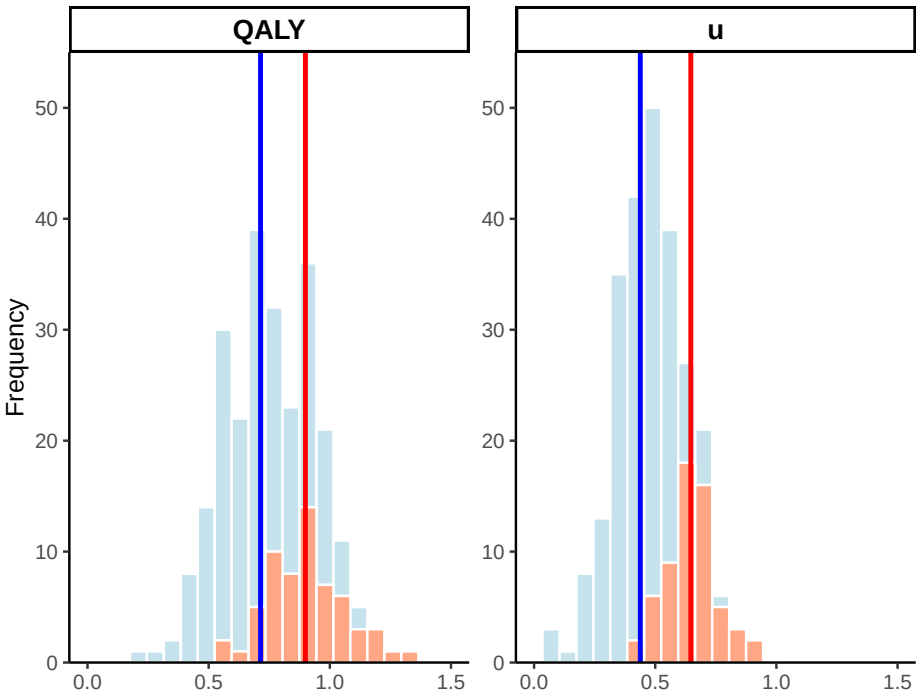
# Missing data - an example




- QALYs under **MCAR** & utilities

| Mean(e) | Sd(e) | Mean(u) | Sd(u) | n |
|---|---|---|---|---|
| **0.770** | 0.194 | **0.499** | 0.147 | *185* |
| **0.728** | 0.171 | **0.462** | 0.145 | *65* |

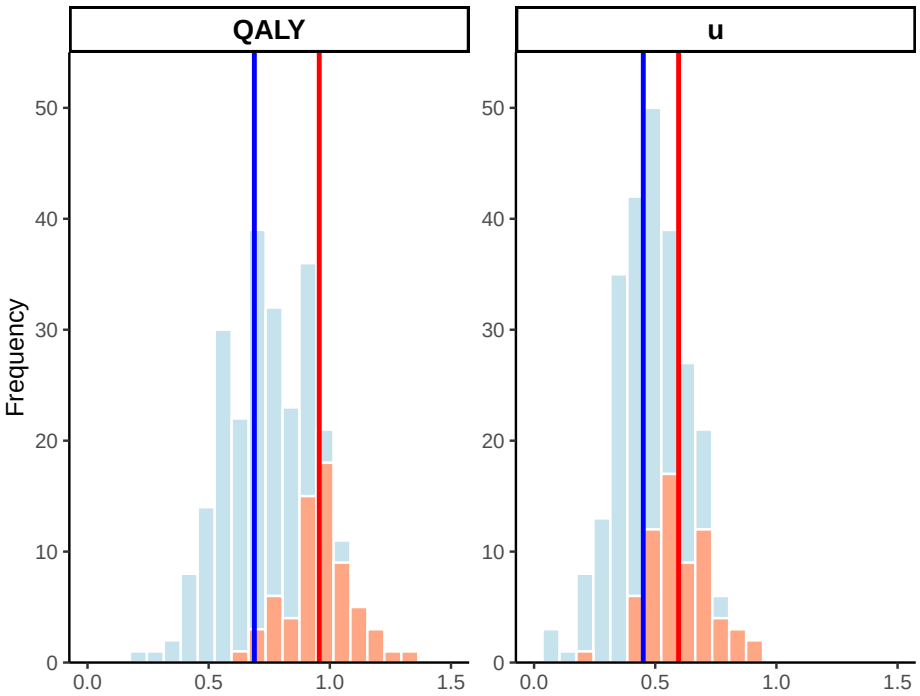
# Missing data - an example




- QALYs under **MAR** & utilities

| Mean(e) | Sd(e) | Mean(u) | Sd(u) | n |
|---|---|---|---|---|
| **0.714** | 0.176 | **0.438** | 0.121 | *189* |
| **0.899** | 0.157 | **0.647** | 0.103 | *61* |

# Missing data - an example




- QALYs under **MNAR** & utilities

| Mean(e) | Sd(e) | Mean(u) | Sd(u) | n |
|---|---|---|---|---|
| **0.689** | 0.154 | **0.451** | 0.134 | *184* |
| **0.956** | 0.129 | **0.596** | 0.129 | *66* |

# Missing data - alternative methods

- **Complete Case Analysis** (CCA):
  - **Eliminate** partially-observed cases $y^{\text{mis}}$
  - Reduce **efficiency** and may **bias** estimates

# Missing data - alternative methods

- **Complete Case Analysis** (CCA):
  - **Eliminate** partially-observed cases $y^{\text{mis}}$
  - Reduce **efficiency** and may **bias** estimates
- **Single Imputation** (SI):
  - Impute $y^{\text{mis}}$ with a **single value** $y^{\text{mis}} = y^{\star}$ (eg mean, zero, etc.)
  - **Ignores** uncertainty with *restrictive* assumptions

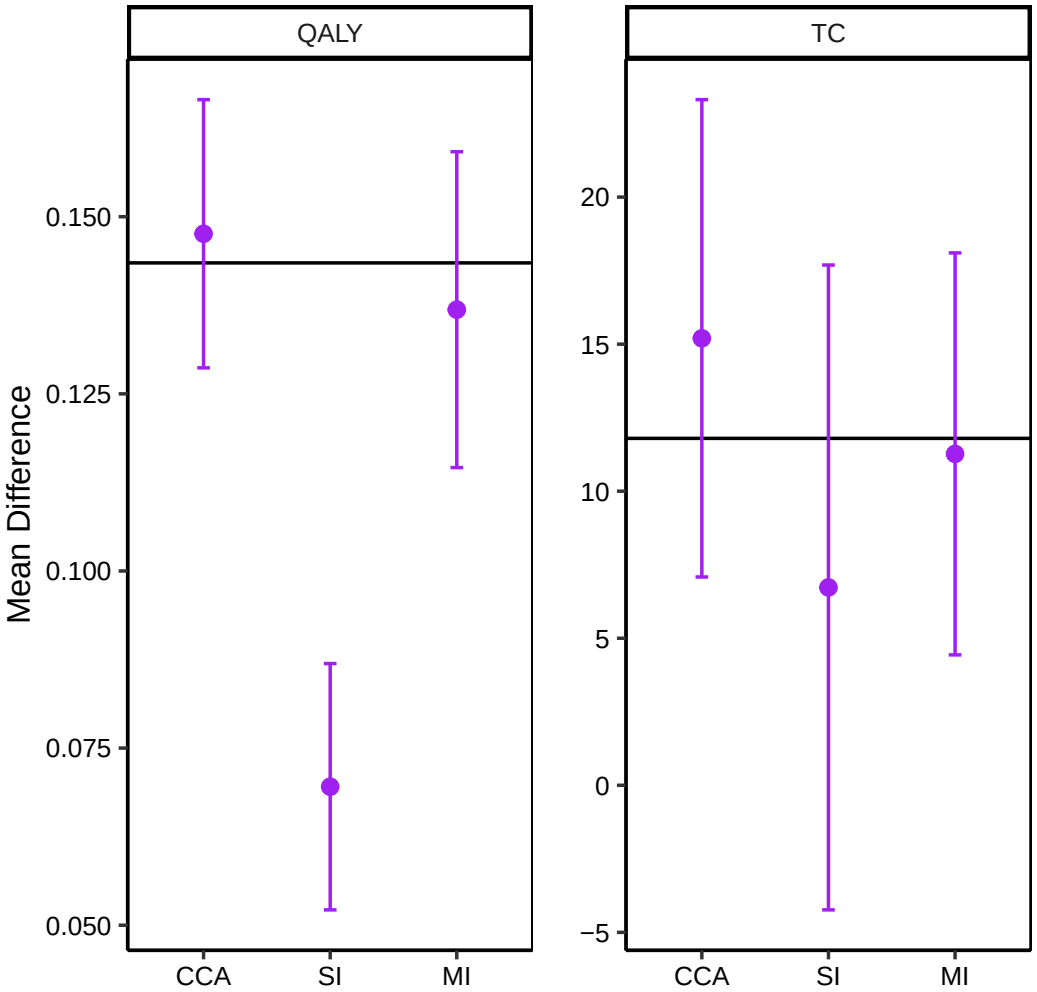
Maastricht University

# Missing data - alternative methods

- **Complete Case Analysis** (CCA):
  - **Eliminate** partially-observed cases $y^{\text{mis}}$
  - Reduce **efficiency** and may **bias** estimates
- **Single Imputation** (SI):
  - Impute $y^{\text{mis}}$ with a **single value** $y^{\text{mis}} = y^{\star}$ (eg mean, zero, etc.)
  - **Ignores** uncertainty with *restrictive* assumptions
- **Multiple Imputation** (MI):
  - **Impute** $y^{\text{mis}}$ $M$ times with an *imputation model*
  - **Analyse** each dataset and derive estimates $\hat{\beta}_m$ for $m = 1, \dots, M$
  - **Combine** $M$ estimates into a single quantity $\hat{\beta}$
  - Validity relies on **correct** model specification

Maastricht University

# Missing data - alternative methods



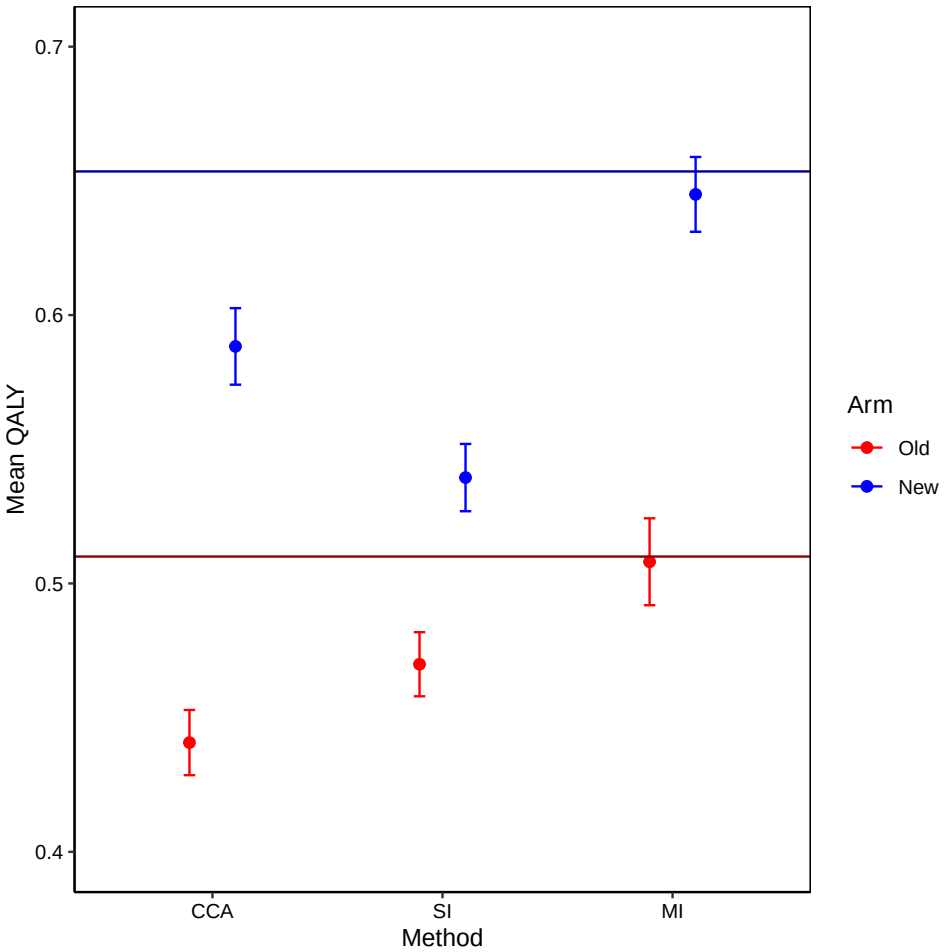

- **Mean QALY difference $\Delta_e$**

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.148*** | *0.129* | *0.167* | *CCA* |
| ***0.070*** | *0.052* | *0.087* | *SI* |
| ***0.137*** | *0.115* | *0.159* | *MI* |

- **Mean TC difference $\Delta_c$**

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***15.197*** | *7.082* | *23.313* | *CCA* |
| ***6.728*** | *-4.236* | *17.691* | *SI* |
| ***11.268*** | *4.434* | *18.102* | *MI* |

# Missing data - alternative methods




- **Mean QALY** for arm = "Old"

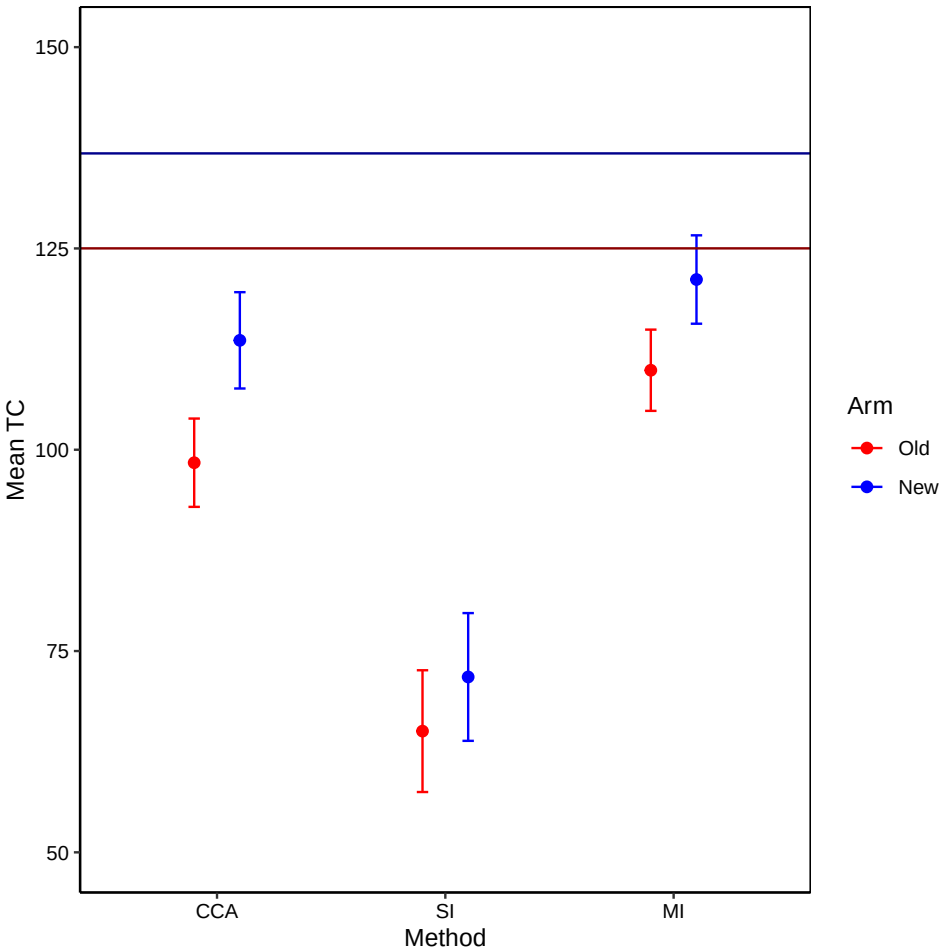
| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.441*** | *0.429* | *0.453* | *CCA* |
| ***0.470*** | *0.458* | *0.482* | *SI* |
| ***0.508*** | *0.492* | *0.524* | *MI* |

- **Mean QALY** for arm = "New"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***0.588*** | *0.574* | *0.603* | *CCA* |
| ***0.539*** | *0.527* | *0.552* | *SI* |
| ***0.645*** | *0.631* | *0.659* | *MI* |

Maastricht University

# Missing data - alternative methods




- **Mean TC** for arm = "Old"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***98.389*** | *92.904* | *103.874* | *CCA* |
| ***65.052*** | *57.490* | *72.614* | *SI* |
| ***109.871*** | *104.826* | *114.917* | *MI* |

- **Mean TC** for arm = "New"

| Estimate | CI.low | CI.high | method |
|---|---|---|---|
| ***113.586*** | *107.605* | *119.567* | *CCA* |
| ***71.780*** | *63.845* | *79.714* | *SI* |
| ***121.139*** | *115.650* | *126.628* | *MI* |

Maastricht University

# Missing data - things to consider

- "Simple" methods **(CCA/SI)**:
  - **Easy** to implement but likely **inefficient** and/or **biased**
  - Often valid under *quite restrictive* assumptions (eg MCAR)

# Missing data - things to consider

- "Simple" methods **(CCA/SI)**:
  - **Easy** to implement but likely **inefficient** and/or **biased**
  - Often valid under *quite restrictive* assumptions (eg MCAR)
- "Advanced" methods **(MI)**:
  - Account for missingness **uncertainty**
  - Require specification of **imputation model** for each variable (MICE)
  - Valid under *less restrictive* assumptions (eg MAR)

Maastricht University

# Missing data - things to consider

- "Simple" methods **(CCA/SI)**:
  - **Easy** to implement but likely **inefficient** and/or **biased**
  - Often valid under *quite restrictive* assumptions (eg MCAR)
- "Advanced" methods **(MI)**:
  - Account for missingness **uncertainty**
  - Require specification of **imputation model** for each variable (MICE)
  - Valid under *less restrictive* assumptions (eg MAR)
- **Further issues** to address:
  - Check **sensitivity** of results to departures (eg MNAR)
    Use specific *MNAR approaches* (Leurent et al. 2018)
  - Imputation of **longitudinal** variables $(u_{ij}, c_{ij})$ challenging
    Use mixed models under *MAR* (Gabrio et al. 2022)
  - "Best" way to **combine** MI/bootstrap *unclear* (Brand et al. 2019)

Maastricht University

# Are we still alive?

Section 4

# Reporting of CEAs

# Data & Methods

- **Patient characteristics** (eg baseline var) & **effectiveness** by arm

# Data & Methods

- **Patient characteristics** (eg baseline var) & **effectiveness** by arm
- **Costs** & **Quality of Life** :
  - **Prices** (with year) and **volumes** of all cost components
  - QoL **instrument** and summary of **measurement moments**
  - Valuation of health states and methods used (if applicable)

Maastricht University

# Data & Methods

- **Patient characteristics** (eg baseline var) & **effectiveness** by arm
- **Costs** & **Quality of Life** :
  - **Prices** (with year) and **volumes** of all cost components
  - QoL **instrument** and summary of **measurement moments**
  - Valuation of health states and methods used (if applicable)
- **Missing Data** :
  - Description of **amounts** and **patterns**
  - How it was handled (including *type of imputation*)

Maastricht University

# Data & Methods

- **Patient characteristics** (eg baseline var) & **effectiveness** by arm
- **Costs** & **Quality of Life** :
  - **Prices** (with year) and **volumes** of all cost components
  - QoL **instrument** and summary of **measurement moments**
  - Valuation of health states and methods used (if applicable)
- **Missing Data** :
  - Description of **amounts** and **patterns**
  - How it was handled (including *type of imputation*)
- **Extrapolation** & **Validation** :
  - Measures for checking the validity of extrapolation (time-to-event data)
  - *AdVISHE*[2] checklist (validation of input data and outcomes)

[2]Assessment of the Validation Status of Health-Economic

Maastricht University

# Results: base-case analysis

- **Costs** & **Effects** :
  - *Absolute* & *incremental* estimates with CIs
  - *Absolute* & *incremental* costs by cost category and total sums
  - *Absolute* & *incremental* QoL by health state (if applicable)

Maastricht University

# Results: base-case analysis

- **Costs & Effects** :
  - *Absolute* & *incremental* estimates with CIs
  - *Absolute* & *incremental* costs by cost category and total sums
  - *Absolute* & *incremental* QoL by health state (if applicable)
- **CE quantities** :
  - **Incremental Cost-Effectiveness Ratio** : ICER $= \frac{\Delta_e}{\Delta_c}$
  - **Net Monetary Benefit** : NMB $= k \times \Delta_e - \Delta_c$

    Use **reference CE threshold** $k$ –> **CE in practice Manual**

Maastricht University

# Results: base-case analysis

- **Costs & Effects** :
  - *Absolute & incremental* estimates with CIs
  - *Absolute & incremental* costs by cost category and total sums
  - *Absolute & incremental* QoL by health state (if applicable)
- **CE quantities** :
  - **Incremental Cost-Effectiveness Ratio** : ICER $= \frac{\Delta_e}{\Delta_c}$
  - **Net Monetary Benefit** : NMB $= k \times \Delta_e - \Delta_c$

    Use **reference CE threshold** $k$ –> **CE in practice Manual**
- **Subgroup analyses** (if applicable):
  - All CE results *presented separately* for each subgroup

Maastricht University

# Results: base-case analysis - an example

- Summary of CE results in **table** format:

| | Mean | SE | CI(low) | CI(high) |
|---|---|---|---|---|
| ***Old(QALY)*** | ***0.511*** | *0.012* | *0.487* | *0.535* |
| ***New(QALY)*** | ***0.557*** | *0.013* | *0.532* | *0.582* |
| ***Old(TC)*** | ***125.010*** | *2.169* | *120.739* | *129.281* |
| ***New(TC)*** | ***147.056*** | *2.275* | *142.575* | *151.538* |
| **Incr(QALY)** | **0.046** | 0.018 | 0.012 | 0.081 |
| **Incr(TC)** | **22.046** | 3.144 | 15.854 | 28.239 |
| **NMB(k=5000)** | **24.385** | | | |
| **ICER** | **474.820** | | | |

Maastricht University

# Results: base-case analysis - an example

- Summary of CE results in **table** format:

| | Mean | SE | CI(low) | CI(high) |
|---|---|---|---|---|
| ***Old(QALY)*** | ***0.511*** | *0.012* | *0.487* | *0.535* |
| ***New(QALY)*** | ***0.557*** | *0.013* | *0.532* | *0.582* |
| ***Old(TC)*** | ***125.010*** | *2.169* | *120.739* | *129.281* |
| ***New(TC)*** | ***147.056*** | *2.275* | *142.575* | *151.538* |
| **Incr(QALY)** | **0.046** | 0.018 | 0.012 | 0.081 |
| **Incr(TC)** | **22.046** | 3.144 | 15.854 | 28.239 |
| **NMB(k=5000)** | **24.385** | | | |
| **ICER** | **474.820** | | | |

- In **Base-Case** analysis, focus on:
  - **Point** & **CI** estimates for QALY/TC by arm (over the study period)
  - **Point** & **CI** estimates for incremental QALY/TC (New-Old)
  - **Point** estimate for *ICER* and *NMB* (assuming **reference value** for $k$)

Maastricht University

# Results: sensitivity analyses

- **Deterministic** & **scenario** analyses (*model-based*):
  - Chosen parameter distribution
  - ICER and incremental quantities lower/upper values
  - Always compared with base-case results

Maastricht University

# Results: sensitivity analyses

- **Deterministic** & **scenario** analyses (*model-based*):
  - Chosen parameter distribution
  - ICER and incremental quantities lower/upper values
  - Always compared with base-case results
- **Probabilistic** analyses:
  - Graphically presented using:
    **Cost-Effectiveness Plane** (*CE plane*)
    **Cost-Effectiveness Acceptability Curve** (*CEAC*)

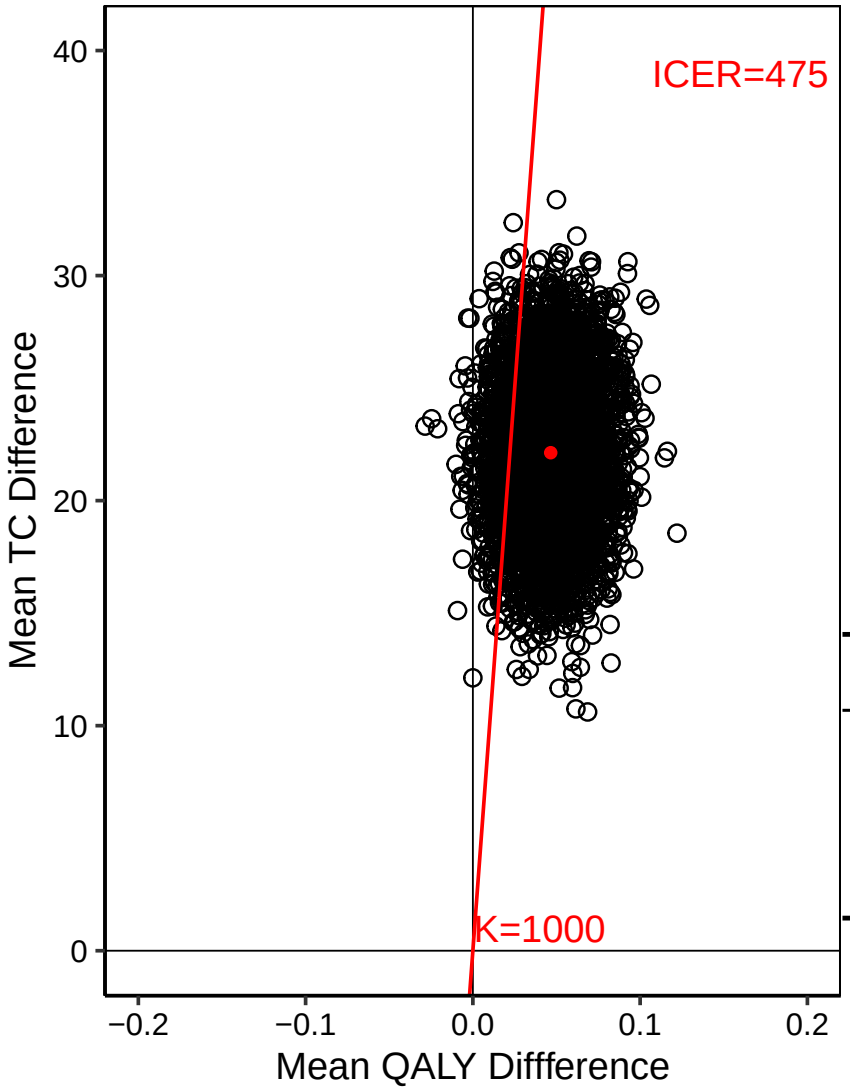
Maastricht University

# Results: sensitivity analyses

- **Deterministic** & **scenario** analyses (*model-based*):
  - Chosen parameter distribution
  - ICER and incremental quantities lower/upper values
  - Always compared with base-case results
- **Probabilistic** analyses:
  - Graphically presented using:
    **Cost-Effectiveness Plane** (*CE plane*)
    **Cost-Effectiveness Acceptability Curve** (*CEAC*)
- **Value of Information** analyses (*model-based*):
  - **Expected Value of (Partial) Perfect Information** (*EV(P)PI*) plot
  - EVSI and ENBS results may be presented –> **VoI analyses Manual**

Maastricht University

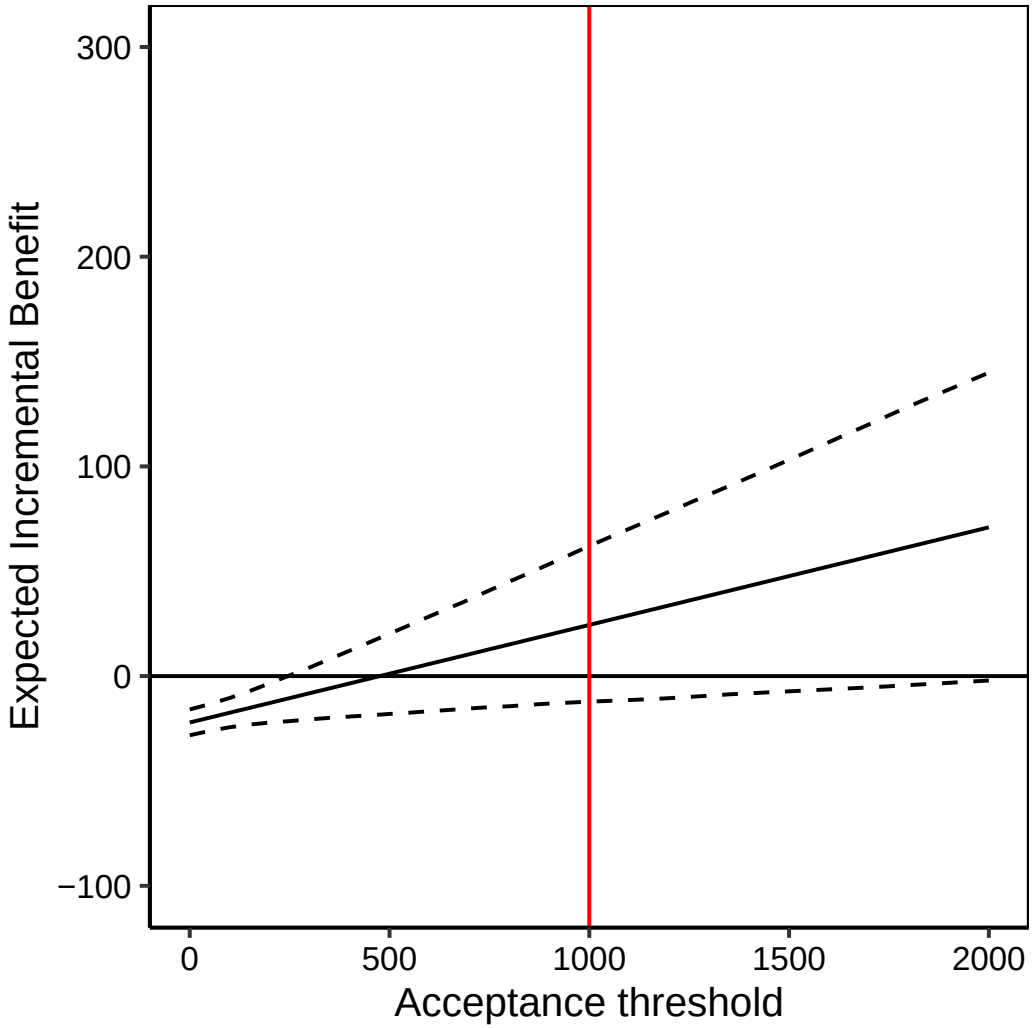
# Results: probabilistic analyses - an example




- **CE plane**
  - Scatter plot of **boostrapped** estimates $(\Delta_e^b, \Delta_c^b)$
  - Red line slope equal to **reference value** for $K$
  - ICER $= \frac{\mathsf{E}[\Delta_c^b]}{\mathsf{E}[\Delta_e^b]}$

| | Mean | NE | NW | SE | SW | K=1000 |
|---|---|---|---|---|---|---|
| *%* | | *0.994* | *0.006* | *0* | *0* | ***0.909*** |
| **QALY** | 0.047 | | | | | |
| **TC** | 22.129 | | | | | |
| **ICER** | 475.487 | | | | | |

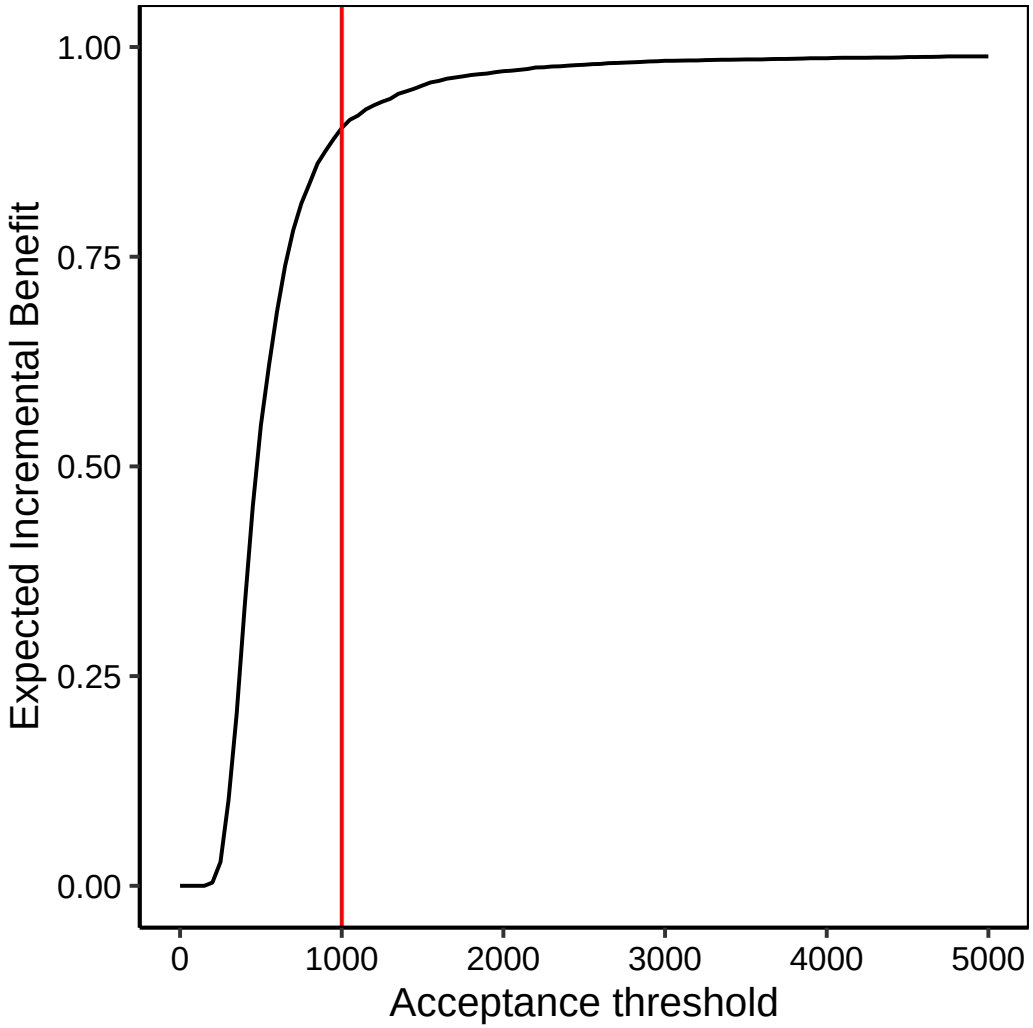
# Results: probabilistic analyses - an example




- **NMB**
  - Plot of **bootstrapped** $\text{NMB}^b$ estimates for $K \in [0, K_{\text{max}}]$
  - Red line equal to **reference value** for $K$
  - Dashed lines are $95\%$ interval based on **bootstrapped** estimates
  - $\text{NMB}^b = K \times \Delta_e^b - \Delta_c^b$

| | Mean | CI(low) | CI(high) |
|---|---|---|---|
| ***K=1000*** | *24.41* | *-12.159* | *62.059* |
| ***K=475.487*** | *0.00* | *-18.302* | *18.195* |

# Results: probabilistic analyses - an example




- **CEAC**
  - Plot of % of **boostrapped** estimates $\frac{\Delta_c^b}{\Delta_e^b}$ below $K$ for $K \in [0, K_{\text{max}}]$
  - Often interpreted as **"probability"** that New is CE compared to Old
  - Red line equal to **reference value** for $K$

| | K=500 | K=1000 | K=2000 |
|---|---|---|---|
| *% CE* | *0.549* | ***0.903*** | *0.971* |

Section 5

# Conclusions

# Summary

- Trial-based analysis considered the **gold standard** for evaluating effectiveness and cost-effectiveness of new interventions
  - Randomisation & ITT analysis minimises chance of **bias**
  - Provide *prospective* patient-level cost and effect data

Maastricht University

# Summary

- Trial-based analysis considered the **gold standard** for evaluating effectiveness and cost-effectiveness of new interventions
  - Randomisation & ITT analysis minimises chance of **bias**
  - Provide *prospective* patient-level cost and effect data
- Use of **invalid** stat methods may still lead to bias
  - **Distort** CE conclusions with "waste" of resources
  - *Statistical quality* of analyses is generally poor (Gabrio et al., 2017; Ling et al. 2022)

Maastricht University

# Summary

- Trial-based analysis considered the **gold standard** for evaluating effectiveness and cost-effectiveness of new interventions
  - Randomisation & ITT analysis minimises chance of **bias**
  - Provide *prospective* patient-level cost and effect data
- Use of **invalid** stat methods may still lead to bias
  - **Distort** CE conclusions with "waste" of resources
  - *Statistical quality* of analyses is generally poor (Gabrio et al., 2017; Ling et al. 2022)
- Methods for health effects often **not appropriate** for CEA:
  - Many *complexities* present simultaneously in CEA:
    Baseline imbalance, Correlation, Skewness, Clustering, Missing data
  - Require a **combination** of different statistical methods use

Maastricht University

# Summary

- Trial-based analysis considered the **gold standard** for evaluating effectiveness and cost-effectiveness of new interventions
  - Randomisation & ITT analysis minimises chance of **bias**
  - Provide *prospective* patient-level cost and effect data
- Use of **invalid** stat methods may still lead to bias
  - **Distort** CE conclusions with "waste" of resources
  - *Statistical quality* of analyses is generally poor (Gabrio et al., 2017; Ling et al. 2022)
- Methods for health effects often **not appropriate** for CEA:
  - Many *complexities* present simultaneously in CEA:
    Baseline imbalance, Correlation, Skewness, Clustering, Missing data
  - Require a **combination** of different statistical methods use
- Important to identify **barriers** and **facilitators** of methods

Maastricht University

# R you still using Excel/SPSS?

# R you still using Excel? (Incerti et. al 2019)

- Historically, HTAs conducted with **commercial software** (eg SPSS) or **spreadsheet software** (eg Excel) which:
  - Are sufficient for simple analyses **BUT**
  - Put **constraints** that limit credibility and relevance

Maastricht University

# R you still using Excel? (Incerti et. al 2019)

- Historically, HTAs conducted with **commercial software** (eg SPSS) or **spreadsheet software** (eg Excel) which:
  - Are sufficient for simple analyses **BUT**
  - Put **constraints** that limit credibility and relevance
- **Modern programming languages** (eg R) facilitate the development of models that are:
  - Increasingly sophisticated and realistic
  - Capable of quantifying decision **uncertainty**
  - Tansparent and reproducible
  - Reusable and adaptable

Maastricht University

# R you still using Excel? (Incerti et. al 2019)

- Historically, HTAs conducted with **commercial software** (eg SPSS) or **spreadsheet software** (eg Excel) which:
  - Are sufficient for simple analyses **BUT**
  - Put **constraints** that limit credibility and relevance
- **Modern programming languages** (eg R) facilitate the development of models that are:
  - Increasingly sophisticated and realistic
  - Capable of quantifying decision **uncertainty**
  - Tansparent and reproducible
  - Reusable and adaptable
- **R** user/developer communities well suited to:
  - Develop/implement HTA models in a **single software environment**
  - Catch up with **methodological advances**
  - Spot and correct code errors via **open-source** nature of packages

Maastricht University

# A path forward for HTA

- Still a general lack of **software experience** in the HTA community:
  - *Insufficient* training in script-based prog software
  - *Limited* guidance on how to implement standard models

Maastricht University

# A path forward for HTA

- Still a general lack of **software experience** in the HTA community:
  - *Insufficient* training in script-based prog software
  - *Limited* guidance on how to implement standard models
- Critical to **train** the next generation of health economists in *state of the art* methods and *software* to implement them

Maastricht University

# A path forward for HTA

- Still a general lack of **software experience** in the HTA community:
  - *Insufficient* training in script-based prog software
  - *Limited* guidance on how to implement standard models
- Critical to **train** the next generation of health economists in *state of the art* methods and *software* to implement them
- **How to do this**?
  - Developing university *courses* & *workshops*
  - Writing *tutorial papers*
  - Making code *freely available* on repositories (eg *GitHub*)
  - *Encouraging* the use of programming languages among researchers

Maastricht University

# Key references


El Alili, M., Dongen, J. M. van, Esser, J. L., Heymans, M. W., Tulder, M. W. van, & Bosmans, J. E. (2022). A scoping review of statistical methods for trial-based economic evaluations: The current state of play. *Health Economics*, *31*(12), 2680–2699.

Gabrio, A., Mason, A. J., & Baio, G. (2019). A full bayesian model to handle structural ones and missingness in economic evaluations from individual-level data. *Statistics in Medicine*, *38*(8), 1399–1420.

Gomes, M., Ng, E. S.-W., Grieve, R., Nixon, R., Carpenter, J., & Thompson, S. G. (2012). Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Medical Decision Making*, *32*(2), 350–361.

Incerti, D., Thom, H., Baio, G., & Jansen, J. P. (2019). R you still using excel? The advantages of modern software tools for health technology assessment. *Value in Health*, *22*(5), 575–579.

Leurent, B., Gomes, M., Faria, R., Morris, S., Grieve, R., & Carpenter, J. R. (2018). Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: A tutorial. *Pharmacoeconomics*, *36*(8), 889–901.

Manca, A., Hawkins, N., & Sculpher, M. J. (2005). Estimating mean QALYs in trial-based cost-effectiveness analysis: The importance of controlling for baseline utility. *Health Economics*, *14*(5), 487–496.

Nederland, Z. (2024). Guideline for economic evaluations in healthcare (2024 version). *Zorginstituut Nederland: Diemen*.


Maastricht University

Section 6

# Appendix - R code

# Appendix - R code - Baseline adjustment

- Full code available **@AnGabrio GitHub**

```
> #fit QALY OLS model adjusting for baseline utilities
> lm2 <- lm(QALY ~ trt + u, data = dataset)
> #get mean QALY in each trt group
> lm2em <- emmeans(lm2, ~ trt)
> #get mean difference between groups
> contrast2_1vs0 <- list("New vs Old" = c(-1, 1))
> #summarise estimates
> confint(contrast(lm2em, contrast2_1vs0))
```

```
 contrast    estimate     SE  df lower.CL upper.CL
 New vs Old    0.0765 0.0108 297   0.0552   0.0979

Confidence level used: 0.95
```

# Appendix - R code - Dealing with correlation

- Full code available **@AnGabrio GitHub**

```
> #fit SUR model to QALY and TC
> sur <- systemfit(list(QALYreg = QALY ~ trt + u,
+     TCreg = TC ~ trt + c), method="SUR", data=dataset)
> #get mean QALY difference CI
> sur_e.trt.ci <- confint(sur, level=0.95)[2,]
> #summarise mean QALY difference between groups
> sur_e.trt.sum <- c(coef(summary(sur))[2,
+     c("Estimate", "Std. Error")],sur_e.trt.ci)
> sur_e.trt.sum
```

```
  Estimate Std. Error      2.5 %     97.5 %
0.07641998 0.01083463 0.05509761 0.09774235
```

# Appendix - R code - Dealing with correlation

- Full code available **@AnGabrio GitHub**

```
> #fit OLS/SUR model and get bootrap estimates
> boot_res <- boot_ec(dataset, QALYreg = QALY ~ trt + u,
+     TCreg = TC ~ trt + c, method = "OLS", B=200)
> summary(boot_res$QALY_boot$Delta_e)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.04894 0.06982 0.07790 0.07711 0.08325 0.10448
```

```
> #compute percentile or BCa CIs
> boot_ci_bca <- boot_ci(x = boot_res, method = "BCa")
> boot_ci_bca$Delta_e
```

```
 1.638952%  96.28698%
0.05464895 0.09617978
```

# Appendix - R code - Dealing with skewness

- Full code available **@AnGabrio GitHub**

```
> #fit GLM  model for TC
> glm_c <- glm(TC ~ trt + c, data = dataset_sam_skew.df,
+   family = Gamma(link = "identity"))
> #get mean TC in each trt group
> glm_c.em <- emmeans(glm_c, ~ trt)
> #get mean difference between groups
> contrast1_1vs0 <- list("New vs Old" = c(-1, 1))
> #summarise estimates
> confint(contrast(glm_c.em, contrast1_1vs0))
```

```
 contrast    estimate   SE df lower.CL upper.CL
 New vs Old      35.6 6.65 17     21.6     49.6

Confidence level used: 0.95
```

Maastricht University

# Appendix - R code - Dealing with clustering

- Full code available **@AnGabrio GitHub**

```
> #fit MLM  model for QALY
> mlm_e <- lme(QALY ~ trt, random = ~1|cluster,
+     data = data.clus.df)
> #get mean QALY in each trt group
> mlm1em_e <- emmeans(mlm_e, ~ trt)
> #get mean difference between groups
> contrast1_1vs0 <- list("New vs Old" = c(-1, 1))
> confint(contrast(mlm1em_e, contrast1_1vs0))
```

```
 contrast   estimate     SE df lower.CL upper.CL
 New vs Old    0.116 0.0744 24  -0.0374     0.27

Degrees-of-freedom method: containment
Confidence level used: 0.95
```

# Appendix - R code - Dealing with clustering

- Full code available **@AnGabrio GitHub**

```
> #fit OLS/SUR model and get TSB estimates
> tsboot_res <- tsboot_ec(data = data.clus.df,
+   QALYreg = QALY ~ trt, TCreg = TC ~ trt,
+   cluster = "cluster", B=200)
> summary(tsboot_res$QALY_boot$Delta_e)

    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.08989  0.06294  0.11181  0.11045  0.15426  0.32714

> #compute percentile or BCa CIs
> tsboot_ci_bca <- boot_ci(x = tsboot_res, method = "BCa")
> tsboot_ci_bca$Delta_e

  2.971331%   97.90686%
-0.05569824  0.27476579
```

# Appendix - R code - Dealing with missing data

- Full code available **@AnGabrio GitHub**

```
> #impute data using MICE with pmm
> mice_data <- mice(full.mar.MI, predictorMatrix = pM,
+   method='pmm', m = 20, print = FALSE)
> #analyse imputed data with OLS for QALY
> lme_mice_data <- with(mice_data, lm(QALY_obs ~ trt + u))
> #get pooled mean QALY per group
> em_lme_mu_data.mi <- emmeans(lme_mice_data, ~ trt)
> #get pooled mean difference between groups
> contrast1_1vs0 <- list("New vs Old" = c(-1, 1))
> confint(contrast(em_lme_mu_data.mi, contrast1_1vs0))
```

```
 contrast    estimate       SE   df lower.CL upper.CL
 New vs Old     0.138 0.00948 41.8    0.119    0.157

Confidence level used: 0.95
```

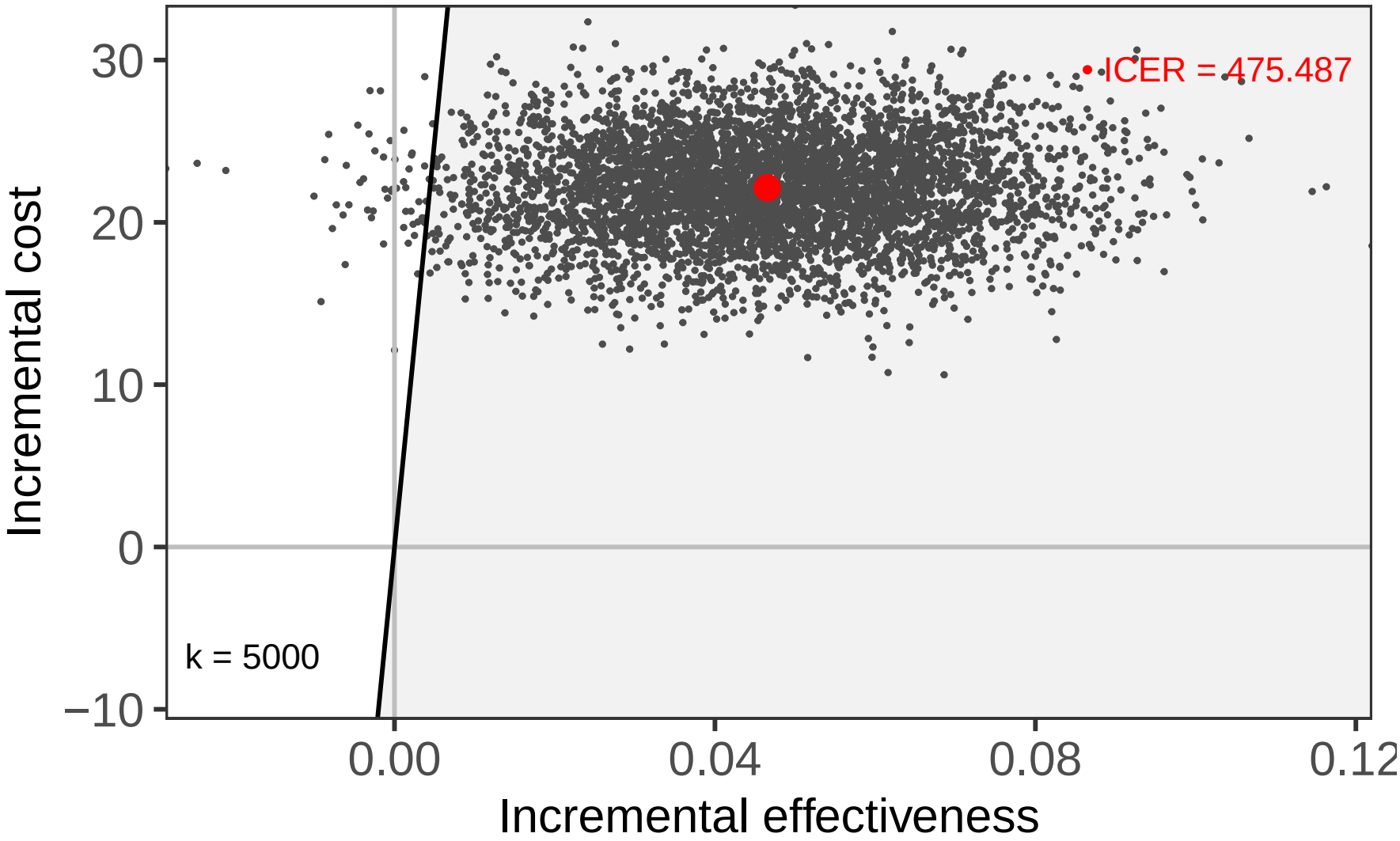
# Appendix - R code - probabilistic analysis

```
> #get CEA output from bootstrapped estimates
> library(BCEA)
> cea_res <- bcea(eff = boo.mu.e, cost = boo.mu.c,
+     Kmax = 25000, ref = 2)
> summary(cea_res)
```

```
Cost-effectiveness analysis summary

Reference intervention:  intervention 2
Comparator intervention: intervention 1

Optimal decision: choose intervention 1 for k < 500 and intervention 2 for k >= 500


Analysis for willingness to pay parameter k = 25000

              Expected net benefit
intervention 1                12642
intervention 2                13783

                                  EIB   CEAC   ICER
intervention 2 vs intervention 1 1141.3 0.9934 475.49

Optimal intervention (max expected net benefit) for k = 25000: intervention 2

EVPI 1.0828
```

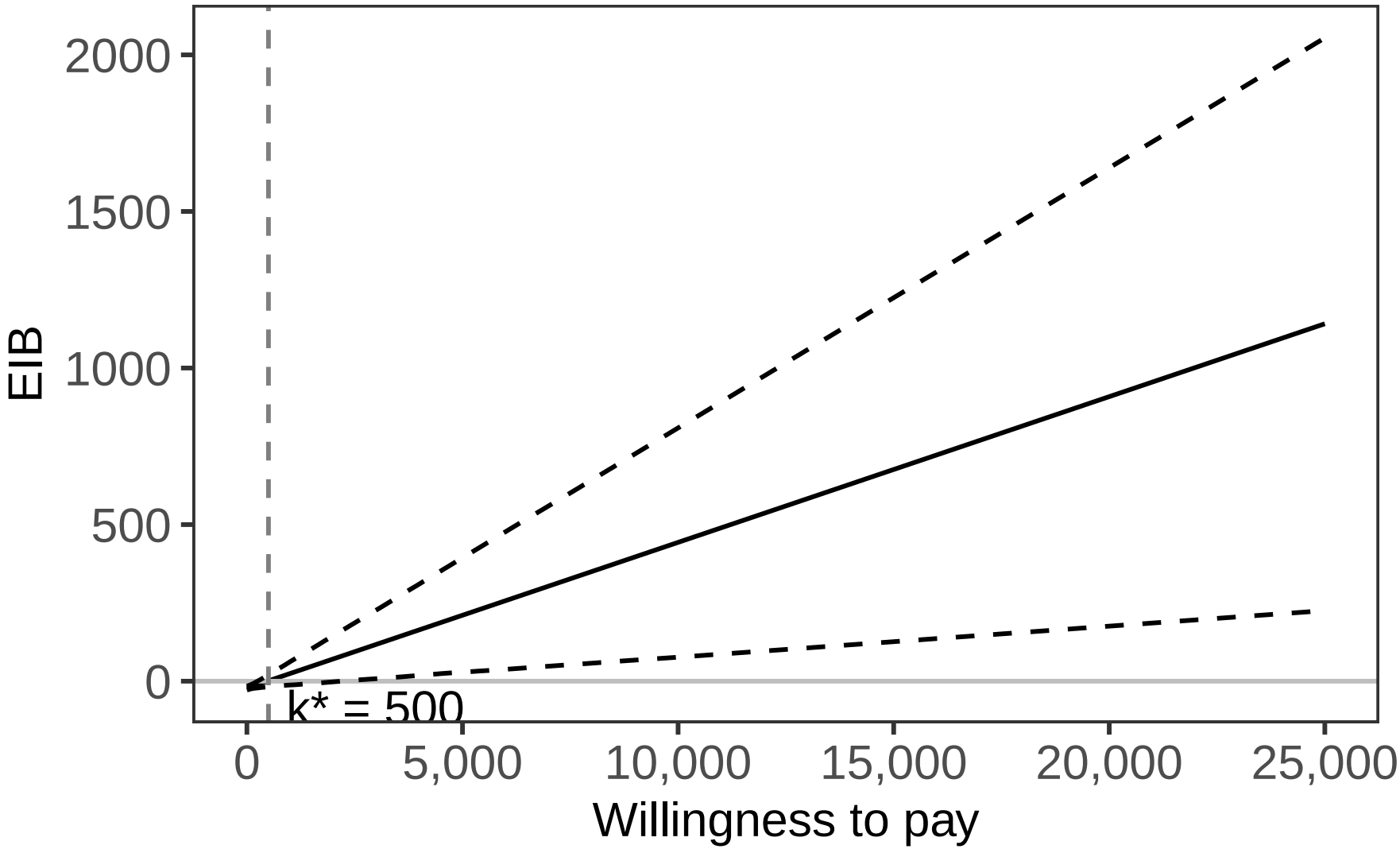
# Appendix - R code - probabilistic analysis

```
> ceplane.plot(cea_res, graph = "ggplot2", wtp = 5000) + labs(title="")
```



# Appendix - R code - probabilistic analysis

```
> eib.plot(cea_res, graph = "ggplot2", plot.cri = TRUE) + labs(title="")
```



# Appendix - R code - probabilistic analysis

```
> ceac.plot(cea_res, graph = "ggplot2") + labs(title="")
```