A gentle introduction to marginal and conditional estimands in causal inference

Rhian Daniel, Cardiff University
Workshop on targeted learning in RCTs
Ghent University, 29th June 2021

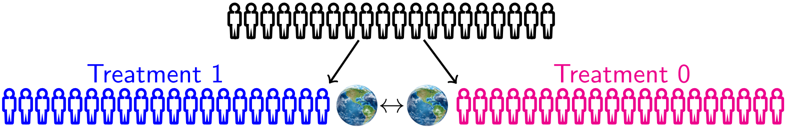
Outline

Outline

Potential outcomes

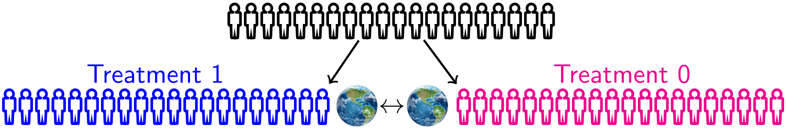
- Take an eligible patient population

Potential outcomes

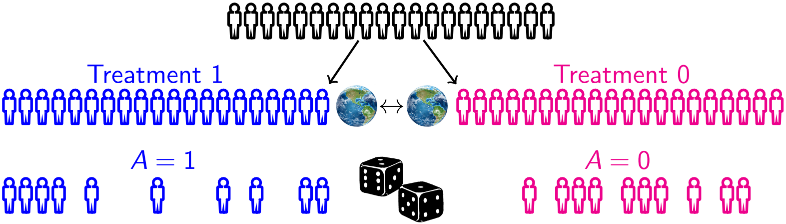


- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.

Potential outcomes

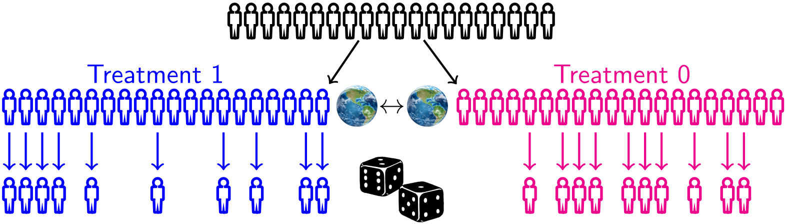


- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.

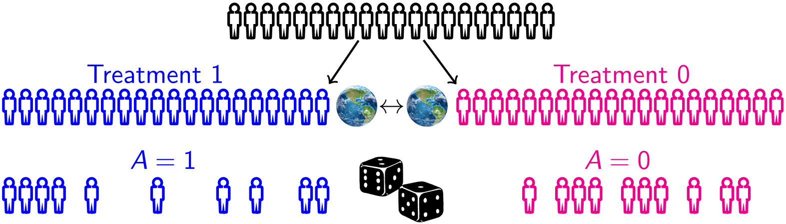


- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.
- In fact patients are randomised to one group or the other. The randomised group is denoted A and the factual outcome is denoted Y.

Potential outcomes

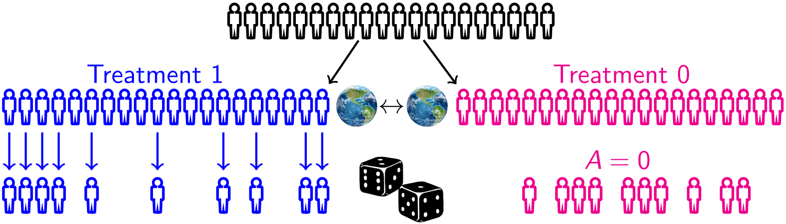


- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.
- In fact patients are randomised to one group or the other. The randomised group is denoted A and the factual outcome is denoted Y.
- Consistency: $Y = Y^1$ if $A = 1$ and $Y = Y^0$ if $A = 0$.



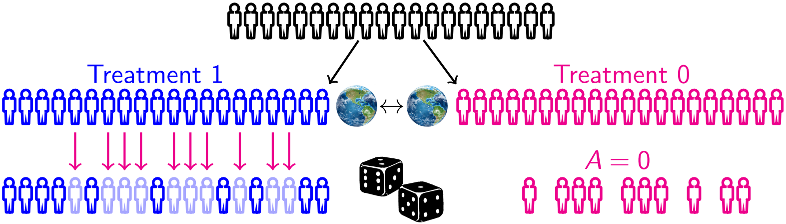
- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.
- In fact patients are randomised to one group or the other. The randomised group is denoted A and the factual outcome is denoted Y.
- Consistency: $Y = Y^1$ if $A = 1$ and $Y = Y^0$ if $A = 0$.
- Exchangeability: Y^1 and Y^0 are both independent of A.

Potential outcomes



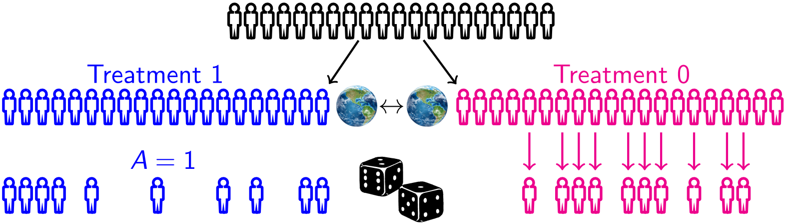
- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.
- In fact patients are randomised to one group or the other. The randomised group is denoted A and the factual outcome is denoted Y.
- Consistency: $Y = Y^1$ if $A = 1$ and $Y = Y^0$ if $A = 0$.
- Exchangeability: Y^1 and Y^0 are both independent of A.

Potential outcomes



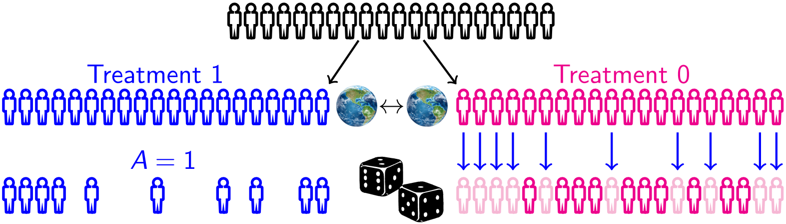
- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.
- In fact patients are randomised to one group or the other. The randomised group is denoted A and the factual outcome is denoted Y.
- Consistency: $Y = Y^1$ if $A = 1$ and $Y = Y^0$ if $A = 0$.
- Exchangeability: Y^1 and Y^0 are both independent of A.

Potential outcomes



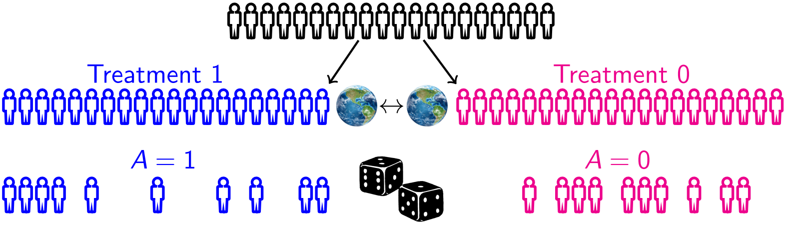
- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.
- In fact patients are randomised to one group or the other. The randomised group is denoted A and the factual outcome is denoted Y.
- Consistency: $Y = Y^1$ if $A = 1$ and $Y = Y^0$ if $A = 0$.
- Exchangeability: Y^1 and Y^0 are both independent of A.

Potential outcomes



- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.
- In fact patients are randomised to one group or the other. The randomised group is denoted A and the factual outcome is denoted Y.
- Consistency: $Y = Y^1$ if $A = 1$ and $Y = Y^0$ if $A = 0$.
- Exchangeability: Y^1 and Y^0 are both independent of A.

Potential outcomes



- Take an eligible patient population, and imagine two parallel worlds: in one they are all assigned treatment 1 and in the other treatment 0.
- Y^1 and Y^0 denote the *potential outcomes* in the two parallel worlds.
- In fact patients are randomised to one group or the other. The randomised group is denoted A and the factual outcome is denoted Y.
- Consistency: $Y = Y^1$ if $A = 1$ and $Y = Y^0$ if $A = 0$.
- Exchangeability: Y^1 and Y^0 are both independent of A.
- This ensures e.g. $\mathbb{E}(Y^1) = \mathbb{E}(Y|A = 1)$ and $\mathbb{E}(Y^0) = \mathbb{E}(Y|A = 0)$.

Outline

'Treatment policy' estimands

- The causal contrast of interest in an RCT is often a contrast between some summary of the distributions of Y^1 and Y^0.

‘Treatment policy’ estimands

- The causal contrast of interest in an RCT is often a contrast between some summary of the distributions of Y^1 and Y^0.
  - $a = 1, 0$ here denotes allocation to treatment, and so this is a ‘treatment policy’ (or intention-to-treat) estimand.

'Treatment policy' estimands

- The causal contrast of interest in an RCT is often a contrast between some summary of the distributions of Y^1 and Y^0.
  - $a = 1, 0$ here denotes allocation to treatment, and so this is a 'treatment policy' (or intention-to-treat) estimand.
  - Other types of estimands exist ('dealing' with intercurrent events somehow) but are not the focus of today's workshop.

‘Treatment policy’ estimands

- The causal contrast of interest in an RCT is often a contrast between some summary of the distributions of Y^1 and Y^0.
  - $a = 1, 0$ here denotes allocation to treatment, and so this is a ‘treatment policy’ (or intention-to-treat) estimand.
  - Other types of estimands exist (‘dealing’ with intercurrent events somehow) but are not the focus of today’s workshop.
- Often the summary is simply the mean: $\mathbb{E}(Y^1)$ vs. $\mathbb{E}(Y^0)$.

'Treatment policy' estimands

- The causal contrast of interest in an RCT is often a contrast between some summary of the distributions of Y^1 and Y^0.
  - $a = 1, 0$ here denotes allocation to treatment, and so this is a 'treatment policy' (or intention-to-treat) estimand.
  - Other types of estimands exist ('dealing' with intercurrent events somehow) but are not the focus of today's workshop.
- Often the summary is simply the mean: $\mathbb{E}(Y^1)$ vs. $\mathbb{E}(Y^0)$.
  - But of course not always, e.g. with time-to-event outcomes.

'Treatment policy' estimands

- The causal contrast of interest in an RCT is often a contrast between some summary of the distributions of Y^1 and Y^0.
  - $a = 1, 0$ here denotes allocation to treatment, and so this is a 'treatment policy' (or intention-to-treat) estimand.
  - Other types of estimands exist ('dealing' with intercurrent events somehow) but are not the focus of today's workshop.
- Often the summary is simply the mean: $\mathbb{E}(Y^1)$ vs. $\mathbb{E}(Y^0)$.
  - But of course not always, e.g. with time-to-event outcomes.
- Several contrasts are possible, e.g. mean difference $\mathbb{E}(Y^1) - \mathbb{E}(Y^0)$, mean ratio $\mathbb{E}(Y^1)/\mathbb{E}(Y^0)$, (for binary outcomes) odds ratio

$$\frac{\mathbb{E}(Y^1)/\{1 - \mathbb{E}(Y^1)\}}{\mathbb{E}(Y^0)/\{1 - \mathbb{E}(Y^0)\}}, \ldots$$

'Treatment policy' estimands

- The causal contrast of interest in an RCT is often a contrast between some summary of the distributions of Y^1 and Y^0.
  - $a = 1, 0$ here denotes allocation to treatment, and so this is a 'treatment policy' (or intention-to-treat) estimand.
  - Other types of estimands exist ('dealing' with intercurrent events somehow) but are not the focus of today's workshop.
- Often the summary is simply the mean: $\mathbb{E}(Y^1)$ vs. $\mathbb{E}(Y^0)$.
  - But of course not always, e.g. with time-to-event outcomes.
- Several contrasts are possible, e.g. mean difference $\mathbb{E}(Y^1) - \mathbb{E}(Y^0)$, mean ratio $\mathbb{E}(Y^1)/\mathbb{E}(Y^0)$, (for binary outcomes) odds ratio

$$\frac{\mathbb{E}(Y^1)/\{1 - \mathbb{E}(Y^1)\}}{\mathbb{E}(Y^0)/\{1 - \mathbb{E}(Y^0)\}}, \ldots$$

- These are all marginal causal contrasts.

Outline

Baseline covariates

- Randomisation ensures that these causal contrasts correspond to statistical contrasts, e.g.

$$\mathbb{E}(Y^1) - \mathbb{E}(Y^0) = \mathbb{E}(Y|A=1) - \mathbb{E}(Y|A=0).$$

Baseline covariates

- Randomisation ensures that these causal contrasts correspond to statistical contrasts, e.g.

$$\mathbb{E}(Y^1) - \mathbb{E}(Y^0) = \mathbb{E}(Y|A = 1) - \mathbb{E}(Y|A = 0).$$

- Exchangeability ($A \perp\!\!\!\perp Y^a,\ a = 1, 0$) is rarely plausible in observational studies. Instead, we typically rely on conditional exchangeability

$$A \perp\!\!\!\perp Y^a | W,\ a = 1, 0$$

given a set of baseline covariates (confounders) W.

Baseline covariates

- Randomisation ensures that these causal contrasts correspond to statistical contrasts, e.g.

$$\mathbb{E}(Y^1) - \mathbb{E}(Y^0) = \mathbb{E}(Y|A = 1) - \mathbb{E}(Y|A = 0).$$

- Exchangeability ($A \perp\!\!\!\perp Y^a,\ a = 1, 0$) is rarely plausible in observational studies. Instead, we typically rely on conditional exchangeability

$$A \perp\!\!\!\perp Y^a | W,\ a = 1, 0$$

given a set of baseline covariates (confounders) W.

- W play a vital role in observational studies: controlling for confounding.

Baseline covariates

- Randomisation ensures that these causal contrasts correspond to statistical contrasts, e.g.

$$\mathbb{E}(Y^1) - \mathbb{E}(Y^0) = \mathbb{E}(Y|A=1) - \mathbb{E}(Y|A=0).$$

- Exchangeability ($A \perp\!\!\!\perp Y^a,\ a = 1, 0$) is rarely plausible in observational studies. Instead, we typically rely on conditional exchangeability

$$A \perp\!\!\!\perp Y^a | W,\ a = 1, 0$$

given a set of baseline covariates (confounders) W.

- W play a vital role in observational studies: controlling for confounding.
- In RCTs, where confounding is not an issue, it is perhaps not surprising that W have traditionally received less attention.

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

1. First a description of the distribution of baseline covariates by arm is almost always given.

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

1. First a description of the distribution of baseline covariates by arm is almost always given.
   - Used to illustrate 'successful' randomisation etc.

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

1. First a description of the distribution of baseline covariates by arm is almost always given.
   - Used to illustrate ‘successful’ randomisation etc.
   - This is not our focus today.

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

1. First a description of the distribution of baseline covariates by arm is almost always given.
   - Used to illustrate ‘successful’ randomisation etc.
   - This is not our focus today.
2. Baseline covariates W can be used in the definition of the causal estimand of interest.

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

1. First a description of the distribution of baseline covariates by arm is almost always given.
   - Used to illustrate ‘successful’ randomisation etc.
   - This is not our focus today.
2. Baseline covariates W can be used in the definition of the causal estimand of interest.
   - e.g. subgroup analyses.

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

1. First a description of the distribution of baseline covariates by arm is almost always given.
   - Used to illustrate 'successful' randomisation etc.
   - This is not our focus today.
2. Baseline covariates W can be used in the definition of the causal estimand of interest.
   - e.g. subgroup analyses.
3. They can also be used in the analysis even when they don't appear in the estimand.

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

1. First a description of the distribution of baseline covariates by arm is almost always given.
   - Used to illustrate 'successful' randomisation etc.
   - This is not our focus today.
2. Baseline covariates W can be used in the definition of the causal estimand of interest.
   - e.g. subgroup analyses.
3. They can also be used in the analysis even when they don't appear in the estimand.
   - e.g. to increase efficiency / power.

Three ways of using baseline covariates

Baseline covariates are still used in RCTs too, however, in (at least) three ways:

1. First a description of the distribution of baseline covariates by arm is almost always given.
   - Used to illustrate ‘successful’ randomisation etc.
   - This is not our focus today.
2. Baseline covariates W can be used in the definition of the causal estimand of interest.
   - e.g. subgroup analyses.
3. They can also be used in the analysis even when they don’t appear in the estimand.
   - e.g. to increase efficiency / power.
   - Some confusion persists over this issue.

Outline

Conditioning vs. adjusting

- One partial explanation for the confusion is that traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.

Conditioning vs. adjusting

- One partial explanation for the confusion is that traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- A useful suggestion is that marginal/conditional is used to describe the estimand of interest and unadjusted/adjusted for the analysis performed.

Conditioning vs. adjusting

- One partial explanation for the confusion is that traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- A useful suggestion is that marginal/conditional is used to describe the estimand of interest and unadjusted/adjusted for the analysis performed.
- It is perfectly possible to obtain an adjusted estimator of a marginal estimand, for example.

Conditioning vs. adjusting

- One partial explanation for the confusion is that traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- A useful suggestion is that marginal/conditional is used to describe the estimand of interest and unadjusted/adjusted for the analysis performed.
- It is perfectly possible to obtain an adjusted estimator of a marginal estimand, for example.
  - See Kelly's and Alex's talks.

Conditioning vs. adjusting

- One partial explanation for the confusion is that traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- A useful suggestion is that marginal/conditional is used to describe the estimand of interest and unadjusted/adjusted for the analysis performed.
- It is perfectly possible to obtain an adjusted estimator of a marginal estimand, for example.
  - See Kelly's and Alex's talks.
- Adjusted estimators of marginal estimands are almost always more precise than unadjusted estimators.

Conditioning vs. adjusting

- One partial explanation for the confusion is that traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- A useful suggestion is that marginal/conditional is used to describe the estimand of interest and unadjusted/adjusted for the analysis performed.
- It is perfectly possible to obtain an adjusted estimator of a marginal estimand, for example.
  - See Kelly's and Alex's talks.
- Adjusted estimators of marginal estimands are almost always more precise than unadjusted estimators.
  - This is the case for binary and time-to-event outcomes, not just for continuous outcomes as is sometimes said.

Conditioning vs. adjusting

- One partial explanation for the confusion is that traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- A useful suggestion is that marginal/conditional is used to describe the estimand of interest and unadjusted/adjusted for the analysis performed.
- It is perfectly possible to obtain an adjusted estimator of a marginal estimand, for example.
  - See Kelly's and Alex's talks.
- Adjusted estimators of marginal estimands are almost always more precise than unadjusted estimators.
  - This is the case for binary and time-to-event outcomes, not just for continuous outcomes as is sometimes said.
  - Confusion enters when people compare the SE of an (adjusted) estimator of a conditional estimand with the SE of an unadjusted estimator of a marginal estimand: apple vs. orange.

Outline

Heterogeneity and non-collapsibility

- Another important distinction concerns heterogeneity.

- Another important distinction concerns heterogeneity.
- A focus on conditional estimands can mean e.g. separately estimating

$$\phi_{\mathsf{F}} = \mathbb{E}(Y^1 - Y^0 | \mathsf{Sex} = \mathsf{F}) \text{ and } \phi_{\mathsf{M}} = \mathbb{E}(Y^1 - Y^0 | \mathsf{Sex} = \mathsf{M})$$

Heterogeneity and non-collapsibility

- Another important distinction concerns heterogeneity.
- A focus on conditional estimands can mean e.g. separately estimating

$$\phi_{\mathsf{F}} = \mathbb{E}(Y^1 - Y^0|\mathsf{Sex} = \mathsf{F}) \text{ and } \phi_{\mathsf{M}} = \mathbb{E}(Y^1 - Y^0|\mathsf{Sex} = \mathsf{M})$$

  — What is the treatment effect in women? What is it in men? (Are they similar?)

Heterogeneity and non-collapsibility

- Another important distinction concerns heterogeneity.
- A focus on conditional estimands can mean e.g. separately estimating

$$\phi_{\mathsf{F}} = \mathbb{E}(Y^1 - Y^0 | \mathsf{Sex} = \mathsf{F}) \text{ and } \phi_{\mathsf{M}} = \mathbb{E}(Y^1 - Y^0 | \mathsf{Sex} = \mathsf{M})$$

  — What is the treatment effect in women? What is it in men? (Are they similar?)
- It's also possible, however, to have a single conditional treatment effect

$$\phi = \mathbb{E}(Y^1 - Y^0 | W)$$

assumed not to vary with W.

Heterogeneity and non-collapsibility

- Another important distinction concerns heterogeneity.
- A focus on conditional estimands can mean e.g. separately estimating

$$\phi_{\mathsf{F}} = \mathbb{E}(Y^1 - Y^0 | \mathsf{Sex} = \mathsf{F}) \text{ and } \phi_{\mathsf{M}} = \mathbb{E}(Y^1 - Y^0 | \mathsf{Sex} = \mathsf{M})$$

  — What is the treatment effect in women? What is it in men? (Are they similar?)

- It's also possible, however, to have a single conditional treatment effect

$$\phi = \mathbb{E}(Y^1 - Y^0 | W)$$

assumed not to vary with W.

  — This is the default implied by a regression model.

Heterogeneity and non-collapsibility

- Another important distinction concerns heterogeneity.
- A focus on conditional estimands can mean e.g. separately estimating

$$\phi_{\mathsf{F}} = \mathbb{E}(Y^1 - Y^0 | \mathsf{Sex} = \mathsf{F}) \text{ and } \phi_{\mathsf{M}} = \mathbb{E}(Y^1 - Y^0 | \mathsf{Sex} = \mathsf{M})$$

  — What is the treatment effect in women? What is it in men? (Are they similar?)

- It's also possible, however, to have a single conditional treatment effect

$$\phi = \mathbb{E}(Y^1 - Y^0 | W)$$

assumed not to vary with W.

  — This is the default implied by a regression model.

- Depending on the effect measure, a conditional estimand assumed homogeneous across levels of W may or may not be equal to the corresponding marginal estimand.

Heterogeneity and non-collapsibility

- Another important distinction concerns heterogeneity.
- A focus on conditional estimands can mean e.g. separately estimating

$$\phi_\mathsf{F} = \mathbb{E}(Y^1 - Y^0|\mathsf{Sex} = \mathsf{F}) \text{ and } \phi_\mathsf{M} = \mathbb{E}(Y^1 - Y^0|\mathsf{Sex} = \mathsf{M})$$

  - What is the treatment effect in women? What is it in men? (Are they similar?)

- It's also possible, however, to have a single conditional treatment effect

$$\phi = \mathbb{E}(Y^1 - Y^0|W)$$

assumed not to vary with W.

  - This is the default implied by a regression model.

- Depending on the effect measure, a conditional estimand assumed homogeneous across levels of W may or may not be equal to the corresponding marginal estimand.
  - This is known as non-collapsibility and is certainly a source of much confusion.

Outline

Non-collapsibility

- To explain non-collapsibility, consider a binary outcome Y, a binary treatment A and (for simplicity) a single continuous baseline covariate W.

Non-collapsibility

- To explain non-collapsibility, consider a binary outcome Y, a binary treatment A and (for simplicity) a single continuous baseline covariate W.
- A GLM for Y given A and W with link function f, assuming no effect modification, is given by:

$$\Pr(Y = 1|A = a, W) = \mathbb{E}(Y|A = a, W) = f^{-1}(\eta + \gamma W + \nu a)$$

Non-collapsibility

- To explain non-collapsibility, consider a binary outcome Y, a binary treatment A and (for simplicity) a single continuous baseline covariate W.
- A GLM for Y given A and W with link function f, assuming no effect modification, is given by:

$$\Pr(Y = 1|A = a, W) = \mathbb{E}(Y|A = a, W) = f^{-1}(\eta + \gamma W + \nu a)$$

- By randomisation, ν can be given a causal interpretation:

$$\Pr(Y^a = 1|W) = \mathbb{E}(Y^a|W) = f^{-1}(\eta + \gamma W + \nu a)$$

Non-collapsibility

- To explain non-collapsibility, consider a binary outcome Y, a binary treatment A and (for simplicity) a single continuous baseline covariate W.
- A GLM for Y given A and W with link function f, assuming no effect modification, is given by:

$$\Pr(Y = 1|A = a, W) = \mathbb{E}(Y|A = a, W) = f^{-1}(\eta + \gamma W + \nu a)$$

- By randomisation, ν can be given a causal interpretation:

$$\Pr(Y^a = 1|W) = \mathbb{E}(Y^a|W) = f^{-1}(\eta + \gamma W + \nu a)$$

- Consider instead the marginal 'model':

$$\mathbb{E}(Y|A = a) = f^{-1}(\alpha + \beta a)$$

Non-collapsibility

- To explain non-collapsibility, consider a binary outcome Y, a binary treatment A and (for simplicity) a single continuous baseline covariate W.
- A GLM for Y given A and W with link function f, assuming no effect modification, is given by:

$$\Pr(Y = 1 | A = a, W) = \mathbb{E}(Y | A = a, W) = f^{-1}(\eta + \gamma W + \nu a)$$

- By randomisation, ν can be given a causal interpretation:

$$\Pr(Y^a = 1 | W) = \mathbb{E}(Y^a | W) = f^{-1}(\eta + \gamma W + \nu a)$$

- Consider instead the marginal 'model':

$$\mathbb{E}(Y | A = a) = f^{-1}(\alpha + \beta a)$$

- By randomisation, β too can be given a causal interpretation:

$$\mathbb{E}(Y^a) = f^{-1}(\alpha + \beta a)$$

- To explain non-collapsibility, consider a binary outcome Y, a binary treatment A and (for simplicity) a single continuous baseline covariate W.
- A GLM for Y given A and W with link function f, assuming no effect modification, is given by:

$$\Pr(Y = 1|A = a, W) = \mathbb{E}(Y|A = a, W) = f^{-1}(\eta + \gamma W + \nu a)$$

- By randomisation, ν can be given a causal interpretation:

$$\Pr(Y^a = 1|W) = \mathbb{E}(Y^a|W) = f^{-1}(\eta + \gamma W + \nu a)$$

- Consider instead the marginal 'model':

$$\mathbb{E}(Y|A = a) = f^{-1}(\alpha + \beta a)$$

- By randomisation, β too can be given a causal interpretation:

$$\mathbb{E}(Y^a) = f^{-1}(\alpha + \beta a)$$

- Depending on f, β and ν may not be equal: non-collapsibility.

Why?

- The essence of the GLM (for a given link function f) when used in this context is that it specifies how $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$.

Why?

- The essence of the GLM (for a given link function f) when used in this context is that it specifies how $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$.
- Specifically, $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$ via the function g with parameter θ where

$$g_\theta(p) = f^{-1}\left\{f(p) + \theta\right\}$$

Why?

- The essence of the GLM (for a given link function f) when used in this context is that it specifies how $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$.
- Specifically, $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$ via the function g with parameter θ where

$$g_\theta(p) = f^{-1}\left\{f(p) + \theta\right\}$$

- Consider what happens if we apply g_ν (g with the conditional parameter) to the marginal $\mathbb{E}(Y^0)$:

$$g_\nu\left\{\mathbb{E}(Y^0)\right\} = g_\nu\left[\mathbb{E}\left\{\mathbb{E}(Y^0|W)\right\}\right]$$

Why?

- The essence of the GLM (for a given link function f) when used in this context is that it specifies how $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$.
- Specifically, $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$ via the function g with parameter θ where

$$g_\theta(p) = f^{-1}\left\{f(p) + \theta\right\}$$

- Consider what happens if we apply g_ν (g with the conditional parameter) to the marginal $\mathbb{E}(Y^0)$:

$$g_\nu\left\{\mathbb{E}(Y^0)\right\} = g_\nu\left[\mathbb{E}\left\{\mathbb{E}(Y^0|W)\right\}\right]$$

- **IF** $g_\nu(p)$ is a linear function of p then we can interchange g_ν and $\mathbb{E}$:

$$g_\nu\left\{\mathbb{E}(Y^0)\right\} = \mathbb{E}\left[g_\nu\left\{\mathbb{E}(Y^0|W)\right\}\right] = \mathbb{E}\left\{\mathbb{E}(Y^1|W)\right\} = \mathbb{E}(Y^1)$$

Why?

- The essence of the GLM (for a given link function f) when used in this context is that it specifies how $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$.
- Specifically, $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$ via the function g with parameter θ where

$$g_\theta(p) = f^{-1}\{f(p) + \theta\}$$

- Consider what happens if we apply g_ν (g with the conditional parameter) to the marginal $\mathbb{E}(Y^0)$:

$$g_\nu\{\mathbb{E}(Y^0)\} = g_\nu\left[\mathbb{E}\{\mathbb{E}(Y^0|W)\}\right]$$

- **IF** $g_\nu(p)$ is a linear function of p then we can interchange g_ν and $\mathbb{E}$:

$$g_\nu\{\mathbb{E}(Y^0)\} = \mathbb{E}\left[g_\nu\{\mathbb{E}(Y^0|W)\}\right] = \mathbb{E}\{\mathbb{E}(Y^1|W)\} = \mathbb{E}(Y^1)$$

- Thus **IF** $g_{\cdot}(p)$ is a linear function of p then $\beta = \nu$.

Why?

- The essence of the GLM (for a given link function f) when used in this context is that it specifies how $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$.
- Specifically, $\mathbb{E}(Y^0|\cdot)$ maps to $\mathbb{E}(Y^1|\cdot)$ via the function g with parameter θ where

$$g_\theta(p) = f^{-1}\left\{f(p) + \theta\right\}$$

- Consider what happens if we apply g_ν (g with the conditional parameter) to the marginal $\mathbb{E}(Y^0)$:

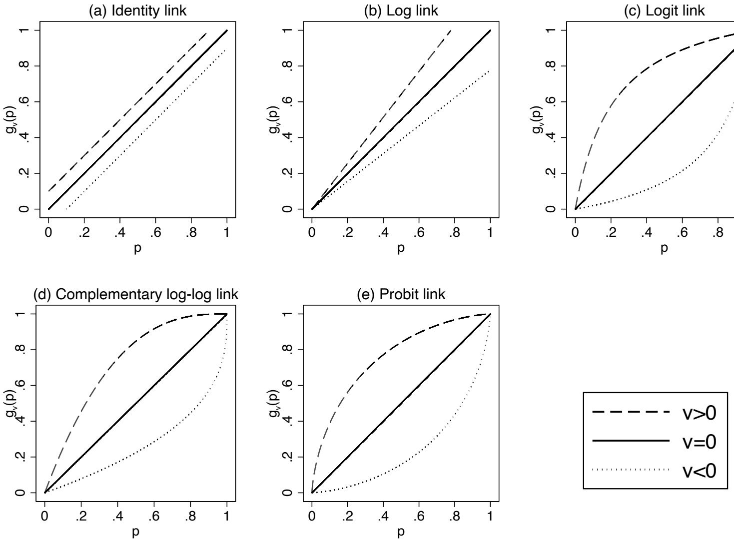
$$g_\nu\left\{\mathbb{E}(Y^0)\right\} = g_\nu\left[\mathbb{E}\left\{\mathbb{E}(Y^0|W)\right\}\right]$$

- **IF** $g_\nu(p)$ is a linear function of p then we can interchange g_ν and $\mathbb{E}$:

$$g_\nu\left\{\mathbb{E}(Y^0)\right\} = \mathbb{E}\left[g_\nu\left\{\mathbb{E}(Y^0|W)\right\}\right] = \mathbb{E}\left\{\mathbb{E}(Y^1|W)\right\} = \mathbb{E}(Y^1)$$

- Thus **IF** $g_.(p)$ is a linear function of p then $\beta = \nu$.
- But for some link functions f, e.g. logit, $g_.(p)$ is non-linear.

$g_{\cdot}(p)$ for common link functions



Outline

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  — Yes, but not to be used as an argument for an unadjusted analysis.

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  — Yes, but not to be used as an argument for an unadjusted analysis.
- **Useful for making blanket policy decisions e.g. should this drug be approved?**

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - Yes, but not to be used as an argument for an unadjusted analysis.
- **Useful for making blanket policy decisions e.g. should this drug be approved?**
  - Yes, but only if the population to which the decision is relevant is similar to the RCT population.

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - Yes, but not to be used as an argument for an unadjusted analysis.
- **Useful for making blanket policy decisions e.g. should this drug be approved?**
  - Yes, but only if the population to which the decision is relevant is similar to the RCT population.
- **Less risk that model misspecification invalidates the analysis.**

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  — Yes, but not to be used as an argument for an unadjusted analysis.
- **Useful for making blanket policy decisions e.g. should this drug be approved?**
  — Yes, but only if the population to which the decision is relevant is similar to the RCT population.
- **Less risk that model misspecification invalidates the analysis.**
  — Sometimes used in defense of an unadjusted analysis; also used in defense of an adjusted analysis of a marginal estimand.

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  — Yes, but not to be used as an argument for an unadjusted analysis.
- **Useful for making blanket policy decisions e.g. should this drug be approved?**
  — Yes, but only if the population to which the decision is relevant is similar to the RCT population.
- **Less risk that model misspecification invalidates the analysis.**
  — Sometimes used in defense of an unadjusted analysis; also used in defense of an adjusted analysis of a marginal estimand.
  — Either way, these arguments are doubtful. Randomisation often gives protection *even when it may appear that* we are relying on additional assumptions in order to bring in W: see all remaining talks.

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - Yes, but not to be used as an argument for an unadjusted analysis.
- **Useful for making blanket policy decisions e.g. should this drug be approved?**
  - Yes, but only if the population to which the decision is relevant is similar to the RCT population.
- **Less risk that model misspecification invalidates the analysis.**
  - Sometimes used in defense of an unadjusted analysis; also used in defense of an adjusted analysis of a marginal estimand.
  - Either way, these arguments are doubtful. Randomisation often gives protection *even when it may appear that* we are relying on additional assumptions in order to bring in W: see all remaining talks.
- **Eliminates concerns over multiple testing / 'gaming' the analysis.**

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - Yes, but not to be used as an argument for an unadjusted analysis.
- **Useful for making blanket policy decisions e.g. should this drug be approved?**
  - Yes, but only if the population to which the decision is relevant is similar to the RCT population.
- **Less risk that model misspecification invalidates the analysis.**
  - Sometimes used in defense of an unadjusted analysis; also used in defense of an adjusted analysis of a marginal estimand.
  - Either way, these arguments are doubtful. Randomisation often gives protection *even when it may appear that* we are relying on additional assumptions in order to bring in W: see all remaining talks.
- **Eliminates concerns over multiple testing / 'gaming' the analysis.**
  - Possibly, but with modern approaches, pre-specification is feasible even when using W: see all remaining talks.

Arguments made for conditional estimands

- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**

Arguments made for conditional estimands

- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**
  — Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity).

Arguments made for conditional estimands

- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**
  - Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity).
- **Estimators of conditional estimands are more precise.**

Arguments made for conditional estimands

- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**
  - Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity).
- **Estimators of conditional estimands are more precise.**
  - This is an argument for adjusted analyses; not (necessarily) for conditional estimands.

Arguments made for conditional estimands

- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**
  - Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity).
- **Estimators of conditional estimands are more precise.**
  - This is an argument for adjusted analyses; not (necessarily) for conditional estimands.
- **Conditional estimands are more relevant to an individual and more transportable to different populations.**

Arguments made for conditional estimands

- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**
  - Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity).
- **Estimators of conditional estimands are more precise.**
  - This is an argument for adjusted analyses; not (necessarily) for conditional estimands.
- **Conditional estimands are more relevant to an individual and more transportable to different populations.**
  - Yes. This is obviously the case if there is heterogeneity and the heterogeneity is modelled.

Arguments made for conditional estimands

- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**
  - Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity).
- **Estimators of conditional estimands are more precise.**
  - This is an argument for adjusted analyses; not (necessarily) for conditional estimands.
- **Conditional estimands are more relevant to an individual and more transportable to different populations.**
  - Yes. This is obviously the case if there is heterogeneity and the heterogeneity is modelled.
  - It even tends to be the case if the conditional effects are wrongly assumed homogeneous.

Arguments made for conditional estimands

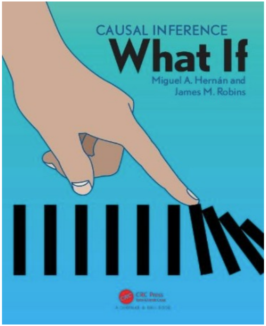
- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**
  — Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity).
- **Estimators of conditional estimands are more precise.**
  — This is an argument for adjusted analyses; not (necessarily) for conditional estimands.
- **Conditional estimands are more relevant to an individual and more transportable to different populations.**
  — Yes. This is obviously the case if there is heterogeneity and the heterogeneity is modelled.
  — It even tends to be the case if the conditional effects are wrongly assumed homogeneous.
  — The choice of effect measure can be important when considering transportability, however.

Arguments made for conditional estimands

- **A more comprehensive understanding of treatment effect, e.g. groups for whom treatment may be especially beneficial.**
  - Yes, but for this, the conditional estimands must be allowed to differ (heterogeneity).
- **Estimators of conditional estimands are more precise.**
  - This is an argument for adjusted analyses; not (necessarily) for conditional estimands.
- **Conditional estimands are more relevant to an individual and more transportable to different populations.**
  - Yes. This is obviously the case if there is heterogeneity and the heterogeneity is modelled.
  - It even tends to be the case if the conditional effects are wrongly assumed homogeneous.
  - The choice of effect measure can be important when considering transportability, however.
  - See work by Anders Huitfeldt.

Outline

More on causal inference, collapsibility, etc



JCE
JOURNAL OF CLINICAL EPIDEMIOLOGY

COMMENTARY | ARTICLES IN PRESS

Noncollapsibility, confounding, and sparse-data bias. Part 1: The oddities of odds

Sander Greenland

Published: June 10, 2021 • DOI: https://doi.org/10.1016/j.jclinepi.2021.06.007

Noncollapsibility, confounding, and sparse-data bias. Part 2: What should researchers make of persistent controversies about the odds ratio?

Sander Greenland

Published: June 10, 2021 • DOI: https://doi.org/10.1016/j.jclinepi.2021.06.004

Huitfeldt *et al. Emerg Themes Epidemiol* (2019) 16:1
https://doi.org/10.1186/s12982-018-0083-9

Emerging Themes in Epidemiology

RESEARCH ARTICLE Open Access

On the collapsibility of measures of effect in the counterfactual causal framework

Anders Huitfeldt[1*], Mats J. Stensrud[2,4] and Etsuji Suzuki[3,4]

Received: 1 October 2019 | Revised: 27 May 2020 | Accepted: 23 July 2020
DOI: 10.1002/bimj.201900297

Biometrical Journal

RESEARCH PAPER

Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets

Rhian Daniel | Jingjing Zhang | Daniel Farewell