Marginal/conditional estimands and the issue of (non)collapsibility

Rhian Daniel, Cardiff University
PSI Webinar on Covariate Adjustment
1st December 2022

Outline

1. Re-cap elements of previous talks
2. (Non)collapsibility
3. The choice between marginal and conditional estimands
4. For more...

Outline

1. Re-cap elements of previous talks
2. (Non)collapsibility
3. The choice between marginal and conditional estimands
4. For more. . .

Conditioning vs. adjusting

- Traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.

Conditioning vs. adjusting

- Traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- Useful suggestion from causal inference (heeded by Jonathan): use marginal/conditional for the estimand of interest and unadjusted/adjusted for the analysis performed.

Conditioning vs. adjusting

- Traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- Useful suggestion from causal inference (heeded by Jonathan): use marginal/conditional for the estimand of interest and unadjusted/adjusted for the analysis performed.
- Can obtain an adjusted estimator of a marginal estimand.

Conditioning vs. adjusting

- Traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- Useful suggestion from causal inference (heeded by Jonathan): use marginal/conditional for the estimand of interest and unadjusted/adjusted for the analysis performed.
- Can obtain an adjusted estimator of a marginal estimand.
  - As discussed by Jonathan.

Conditioning vs. adjusting

- Traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- Useful suggestion from causal inference (heeded by Jonathan): use marginal/conditional for the estimand of interest and unadjusted/adjusted for the analysis performed.
- Can obtain an adjusted estimator of a marginal estimand.
  - As discussed by Jonathan.
- When comparing estimators of a marginal estimand, an adjusted estimator is usually more precise than an unadjusted estimator.

Conditioning vs. adjusting

- Traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- Useful suggestion from causal inference (heeded by Jonathan): use marginal/conditional for the estimand of interest and unadjusted/adjusted for the analysis performed.
- Can obtain an adjusted estimator of a marginal estimand.
  — As discussed by Jonathan.
- When comparing estimators of a marginal estimand, an adjusted estimator is usually more precise than an unadjusted estimator.
  — This is the case for binary and survival outcomes, not just for continuous outcomes, contrary to some discussions.

Conditioning vs. adjusting

- Traditionally conditional and adjusted are used interchangeably, likewise marginal and unadjusted.
- Useful suggestion from causal inference (heeded by Jonathan): use marginal/conditional for the estimand of interest and unadjusted/adjusted for the analysis performed.
- Can obtain an adjusted estimator of a marginal estimand.
  - As discussed by Jonathan.
- When comparing estimators of a marginal estimand, an adjusted estimator is usually more precise than an unadjusted estimator.
  - This is the case for binary and survival outcomes, not just for continuous outcomes, contrary to some discussions.
  - 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.

Heterogeneity and (non)collapsibility

- A focus on conditional estimands can mean e.g. separately estimating

$$\mathbb{E}(Y^1 - Y^0 | X = 0) \text{ and } \mathbb{E}(Y^1 - Y^0 | X = 1)$$

Heterogeneity and (non)collapsibility

- A focus on conditional estimands can mean e.g. separately estimating

$$\mathbb{E}(Y^1 - Y^0 | X = 0) \text{ and } \mathbb{E}(Y^1 - Y^0 | X = 1)$$

  — e.g. what is the treatment effect in smokers? What is it in non-smokers?

Heterogeneity and (non)collapsibility

- A focus on conditional estimands can mean e.g. separately estimating

$$\mathbb{E}(Y^1 - Y^0 | X = 0) \text{ and } \mathbb{E}(Y^1 - Y^0 | X = 1)$$

  - e.g. what is the treatment effect in smokers? What is it in non-smokers?
  - But note caution discussed by Florian.

Heterogeneity and (non)collapsibility

- A focus on conditional estimands can mean e.g. separately estimating

$$\mathbb{E}(Y^1 - Y^0|X = 0) \text{ and } \mathbb{E}(Y^1 - Y^0|X = 1)$$

  - e.g. what is the treatment effect in smokers? What is it in non-smokers?
  - But note caution discussed by Florian.

- Or focus is on a single conditional treatment effect

$$\mathbb{E}(Y^1 - Y^0|X)$$

assumed not to vary with X.

Heterogeneity and (non)collapsibility

- A focus on conditional estimands can mean e.g. separately estimating

$$\mathbb{E}(Y^1 - Y^0|X = 0) \text{ and } \mathbb{E}(Y^1 - Y^0|X = 1)$$

  — e.g. what is the treatment effect in smokers? What is it in non-smokers?
  — But note caution discussed by Florian.

- Or focus is on a single conditional treatment effect

$$\mathbb{E}(Y^1 - Y^0|X)$$

assumed not to vary with X.

  — This is the default implied by a regression model.

Heterogeneity and (non)collapsibility

- A focus on conditional estimands can mean e.g. separately estimating

$$\mathbb{E}(Y^1 - Y^0|X = 0) \text{ and } \mathbb{E}(Y^1 - Y^0|X = 1)$$

  — e.g. what is the treatment effect in smokers? What is it in non-smokers?
  — But note caution discussed by Florian.

- Or focus is on a single conditional treatment effect

$$\mathbb{E}(Y^1 - Y^0|X)$$

assumed not to vary with X.

  — This is the default implied by a regression model.

- A conditional estimand (correctly) assumed homogeneous across levels of X may or may not be equal to the corresponding marginal estimand.

Heterogeneity and (non)collapsibility

- A focus on conditional estimands can mean e.g. separately estimating

$$\mathbb{E}(Y^1 - Y^0|X = 0) \text{ and } \mathbb{E}(Y^1 - Y^0|X = 1)$$

  — e.g. what is the treatment effect in smokers? What is it in non-smokers?
  — But note caution discussed by Florian.

- Or focus is on a single conditional treatment effect

$$\mathbb{E}(Y^1 - Y^0|X)$$

assumed not to vary with X.

  — This is the default implied by a regression model.

- A conditional estimand (correctly) assumed homogeneous across levels of X may or may not be equal to the corresponding marginal estimand.

  — (Non)collapsibility.

Outline

1. Re-cap elements of previous talks
2. (Non)collapsibility
3. The choice between marginal and conditional estimands
4. For more. . .

(Non)collapsibility

- To explain (non)collapsibility, consider a binary outcome Y, binary treatment trt and (for simplicity) a single continuous baseline covariate X (e.g. age).

(Non)collapsibility

- To explain (non)collapsibility, consider a binary outcome Y, binary treatment *trt* and (for simplicity) a single continuous baseline covariate X (e.g. age).
- A GLM for Y given *trt* and X with link function f, assuming no effect modification on that scale:

$$\Pr(Y = 1|trt, X) = \mathbb{E}(Y|trt, X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

(Non)collapsibility

- To explain (non)collapsibility, consider a binary outcome Y, binary treatment *trt* and (for simplicity) a single continuous baseline covariate X (e.g. age).
- A GLM for Y given *trt* and X with link function f, assuming no effect modification on that scale:

$$\Pr(Y = 1|trt, X) = \mathbb{E}(Y|trt, X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

- By randomisation, β_1 can be given a causal interpretation:

$$\Pr(Y^{trt} = 1|X) = \mathbb{E}(Y^{trt}|X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

(Non)collapsibility

- To explain (non)collapsibility, consider a binary outcome Y, binary treatment trt and (for simplicity) a single continuous baseline covariate X (e.g. age).
- A GLM for Y given trt and X with link function f, assuming no effect modification on that scale:

$$\Pr(Y = 1|trt, X) = \mathbb{E}(Y|trt, X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

- By randomisation, β_1 can be given a causal interpretation:

$$\Pr(Y^{trt} = 1|X) = \mathbb{E}(Y^{trt}|X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

- Consider instead the marginal 'model':

$$\mathbb{E}(Y|trt) = f^{-1}(\alpha_0 + \alpha_1 trt)$$

(Non)collapsibility

- To explain (non)collapsibility, consider a binary outcome Y, binary treatment *trt* and (for simplicity) a single continuous baseline covariate X (e.g. age).
- A GLM for Y given *trt* and X with link function f, assuming no effect modification on that scale:

$$\Pr(Y = 1|trt, X) = \mathbb{E}(Y|trt, X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

- By randomisation, β_1 can be given a causal interpretation:

$$\Pr(Y^{trt} = 1|X) = \mathbb{E}(Y^{trt}|X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

- Consider instead the marginal 'model':

$$\mathbb{E}(Y|trt) = f^{-1}(\alpha_0 + \alpha_1 trt)$$

- By randomisation, α_1 too can be given a causal interpretation:

$$\mathbb{E}(Y^{trt}) = f^{-1}(\alpha_0 + \alpha_1 trt)$$

(Non)collapsibility

- To explain (non)collapsibility, consider a binary outcome Y, binary treatment trt and (for simplicity) a single continuous baseline covariate X (e.g. age).
- A GLM for Y given trt and X with link function f, assuming no effect modification on that scale:

$$\Pr(Y = 1|trt, X) = \mathbb{E}(Y|trt, X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

- By randomisation, β_1 can be given a causal interpretation:

$$\Pr(Y^{trt} = 1|X) = \mathbb{E}(Y^{trt}|X) = f^{-1}(\beta_0 + \beta_1 trt + \beta_2 X)$$

- Consider instead the marginal 'model':

$$\mathbb{E}(Y|trt) = f^{-1}(\alpha_0 + \alpha_1 trt)$$

- By randomisation, α_1 too can be given a causal interpretation:

$$\mathbb{E}(Y^{trt}) = f^{-1}(\alpha_0 + \alpha_1 trt)$$

- Depending on f, α_1 and β_1 may or 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_{β_1} (g with the conditional parameter) to the marginal $\mathbb{E}(Y^0)$:

$$g_{\beta_1}\left\{\mathbb{E}(Y^0)\right\} = g_{\beta_1}\left[\mathbb{E}\left\{\mathbb{E}(Y^0|X)\right\}\right]$$

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
$$g_\theta(p) = f^{-1}\{f(p) + \theta\}$$
- Consider what happens if we apply g_{β_1} (g with the conditional parameter) to the marginal $\mathbb{E}(Y^0)$:
$$g_{\beta_1}\{\mathbb{E}(Y^0)\} = g_{\beta_1}\left[\mathbb{E}\{\mathbb{E}(Y^0|X)\}\right]$$
- **IF** $g_{\beta_1}(p)$ is a linear function of p then we can interchange g_{β_1} and $\mathbb{E}$:
$$g_{\beta_1}\{\mathbb{E}(Y^0)\} = \mathbb{E}\left[g_{\beta_1}\{\mathbb{E}(Y^0|X)\}\right] = \mathbb{E}\{\mathbb{E}(Y^1|X)\} = \mathbb{E}(Y^1)$$

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_{β_1} (g with the conditional parameter) to the marginal $\mathbb{E}(Y^0)$:

$$g_{\beta_1}\{\mathbb{E}(Y^0)\} = g_{\beta_1}\left[\mathbb{E}\{\mathbb{E}(Y^0|X)\}\right]$$

- **IF** $g_{\beta_1}(p)$ is a linear function of p then we can interchange g_{β_1} and $\mathbb{E}$:

$$g_{\beta_1}\{\mathbb{E}(Y^0)\} = \mathbb{E}\left[g_{\beta_1}\{\mathbb{E}(Y^0|X)\}\right] = \mathbb{E}\{\mathbb{E}(Y^1|X)\} = \mathbb{E}(Y^1)$$

- Thus **IF** $g_{.}(p)$ is a linear function of p then $\alpha_1 = \beta_1$.

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_{β_1} (g with the conditional parameter) to the marginal $\mathbb{E}(Y^0)$:

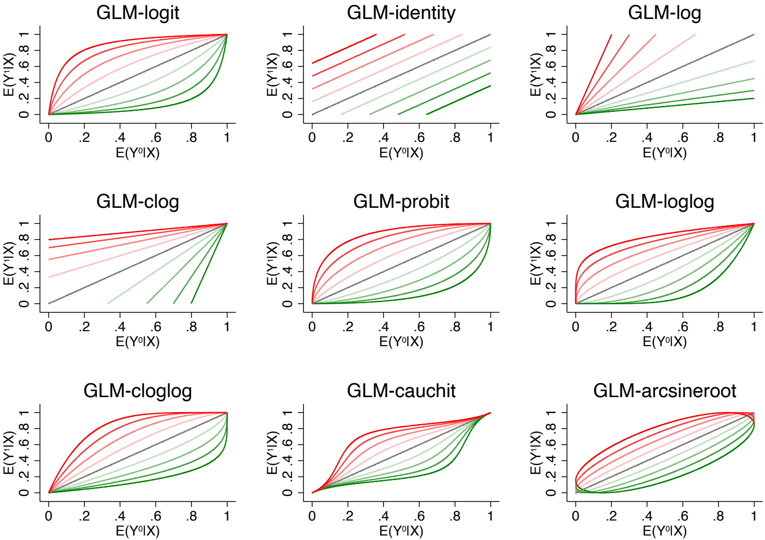
$$g_{\beta_1}\{\mathbb{E}(Y^0)\} = g_{\beta_1}\left[\mathbb{E}\{\mathbb{E}(Y^0|X)\}\right]$$

- **IF** $g_{\beta_1}(p)$ is a linear function of p then we can interchange g_{β_1} and $\mathbb{E}$:

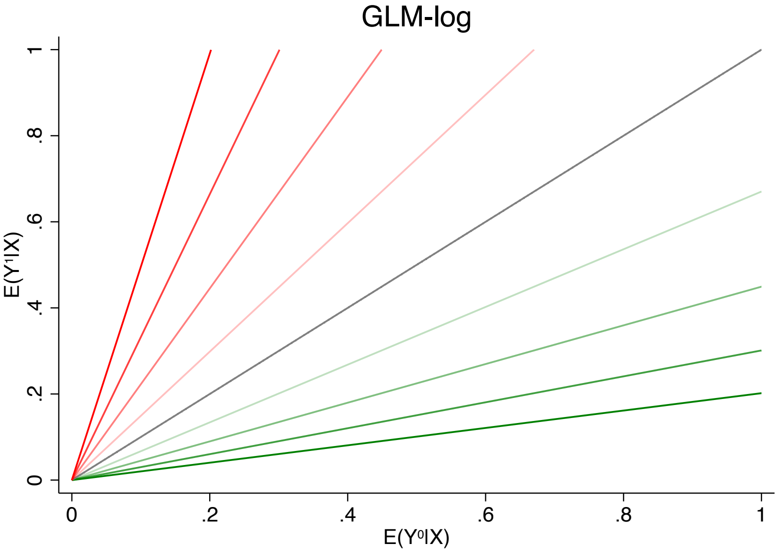
$$g_{\beta_1}\{\mathbb{E}(Y^0)\} = \mathbb{E}\left[g_{\beta_1}\{\mathbb{E}(Y^0|X)\}\right] = \mathbb{E}\{\mathbb{E}(Y^1|X)\} = \mathbb{E}(Y^1)$$

- Thus **IF** $g_{\cdot}(p)$ is a linear function of p then $\alpha_1 = \beta_1$.
- But for many link functions f, e.g. logit, $g_{\cdot}(p)$ is non-linear.

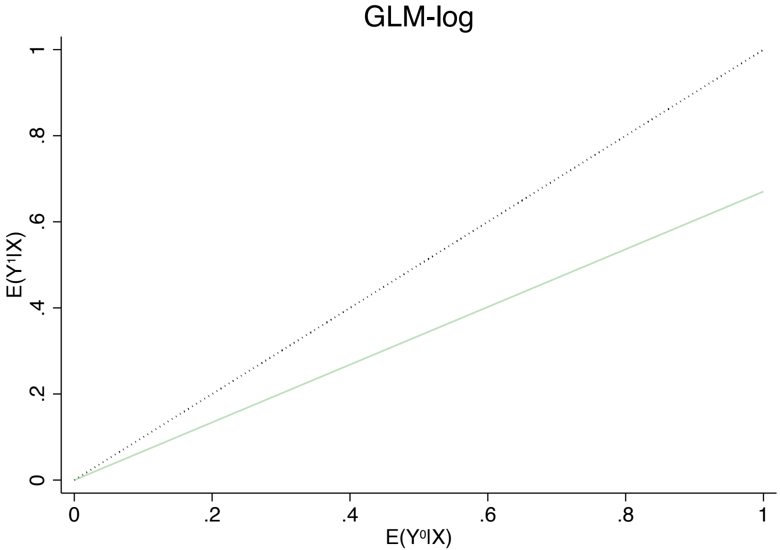
$g_{\cdot}(p)$ for common link functions



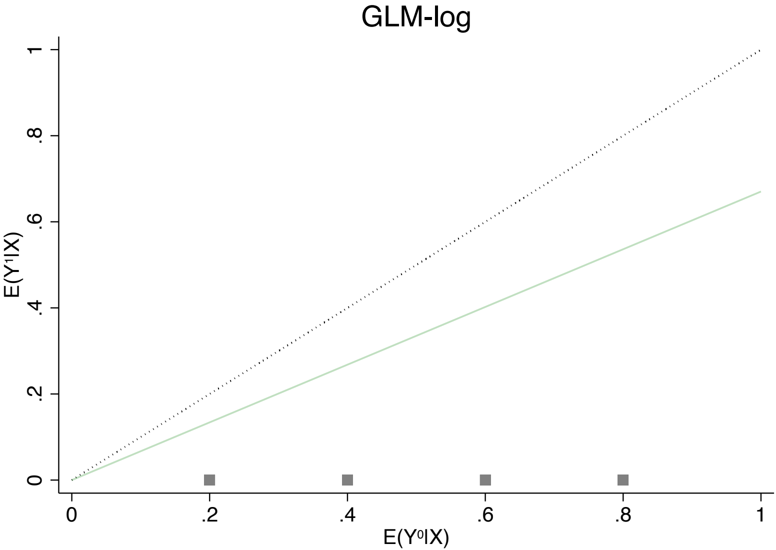
A closer look at the log link



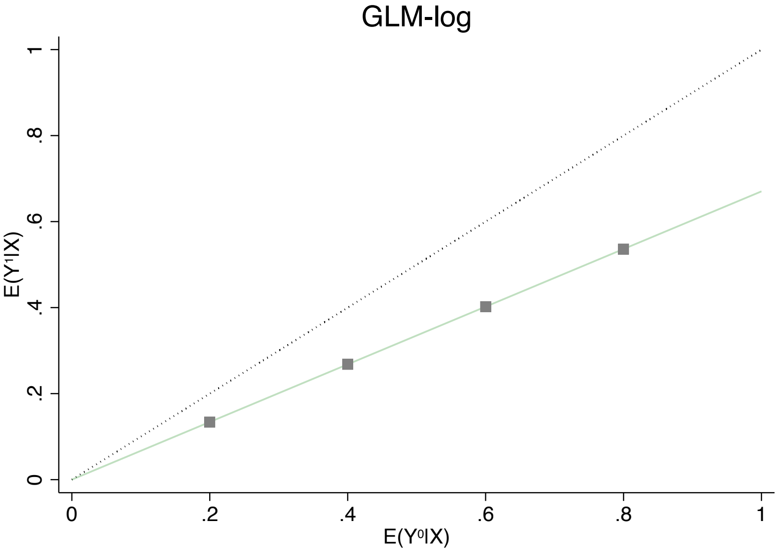
Suppose no heterogeneity and $\beta_1 = -0.4$



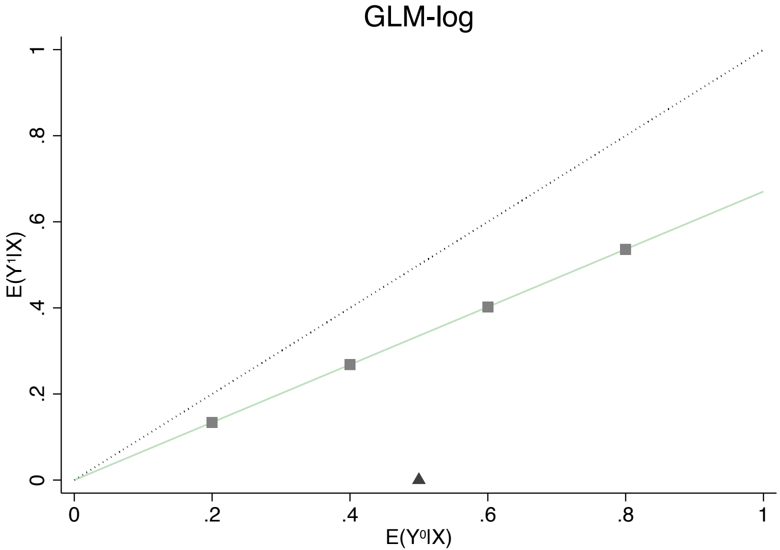
Suppose 4 (equiprobable) values of $\mathbb{E}(Y^0|X)$

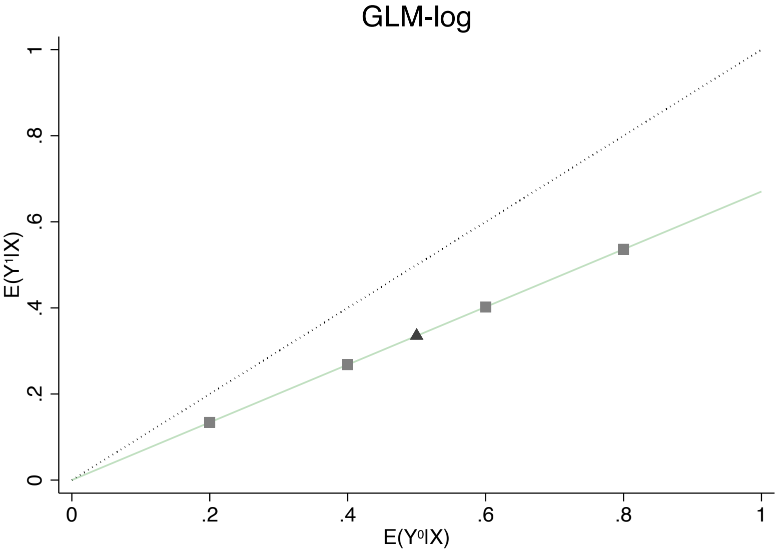


These map like this

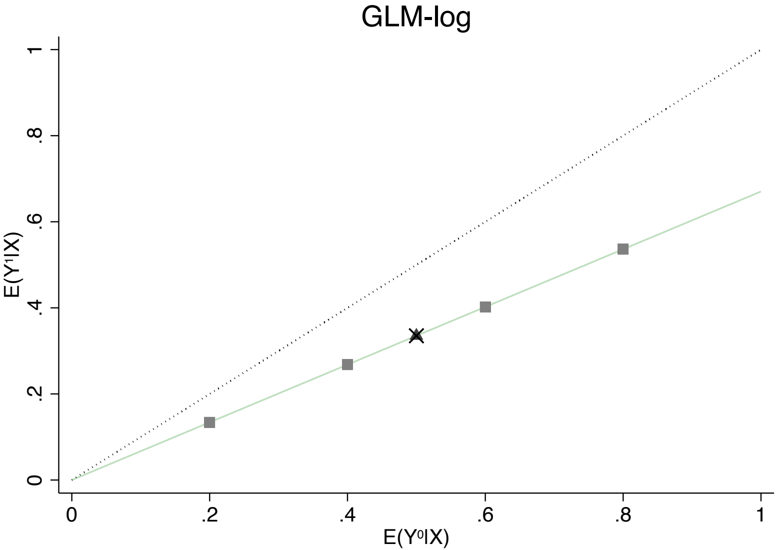


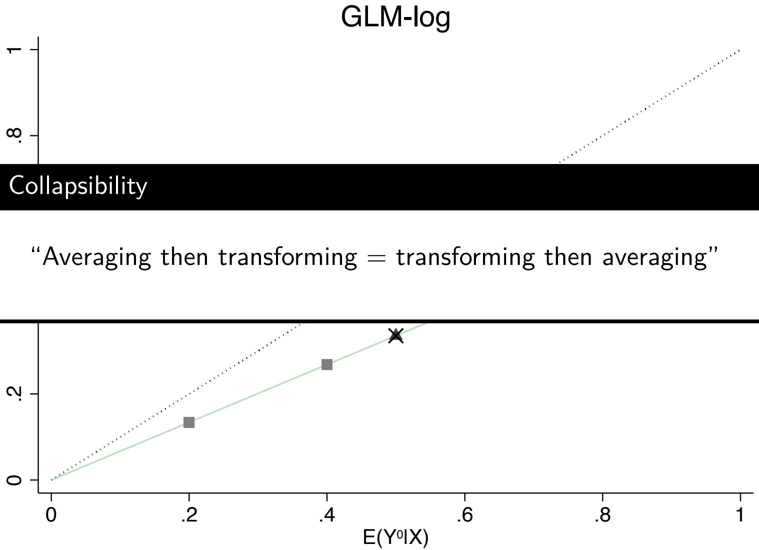
And the ('horizontal') average $\mathbb{E}(Y^0)$...




GLM-log
E(Y¹|X)
0
.2
.4
.6
.8
1
E(Y⁰|X)

. . . which is the 'vertical average', $\mathbb{E}(Y^1)$

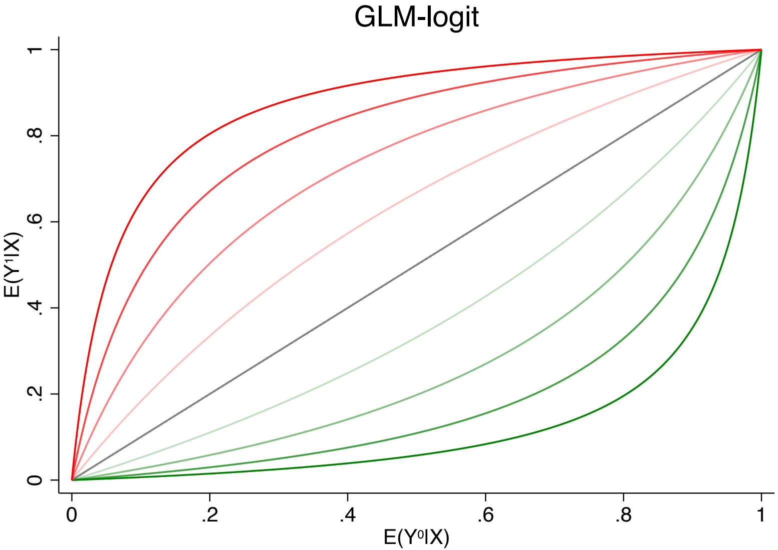




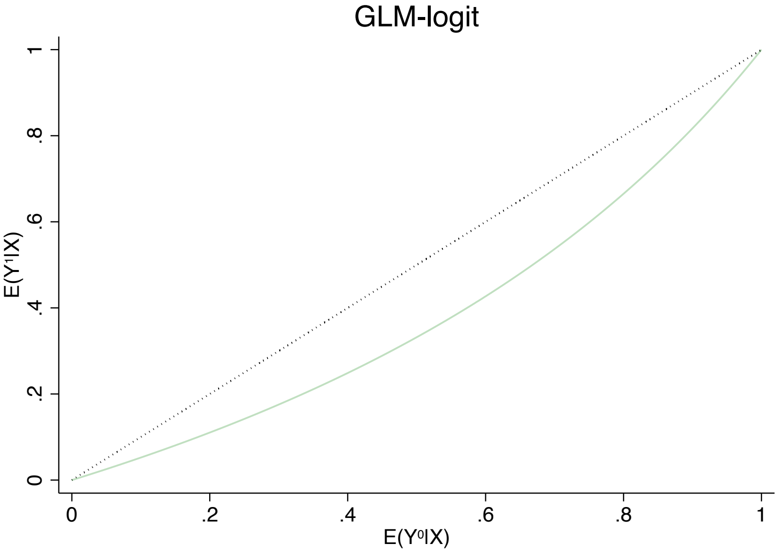
Collapsibility

"Averaging then transforming = transforming then averaging"

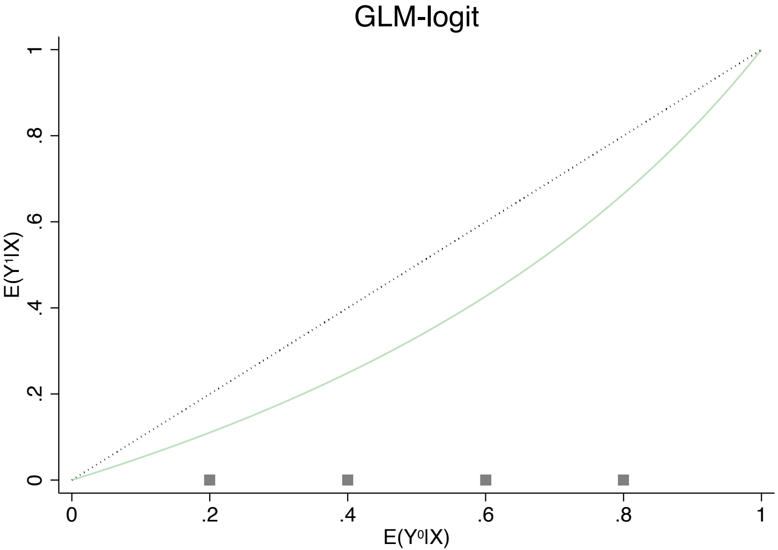
Now take the logit link



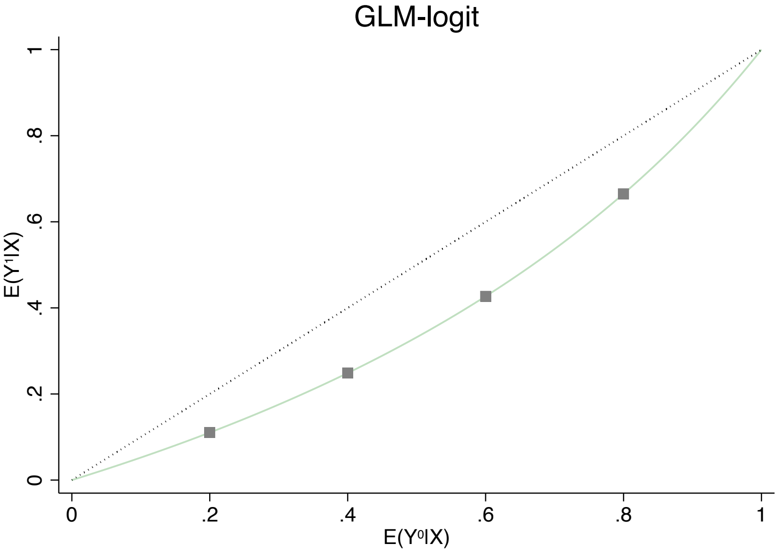
Suppose no heterogeneity and $\beta_1 = -0.7$



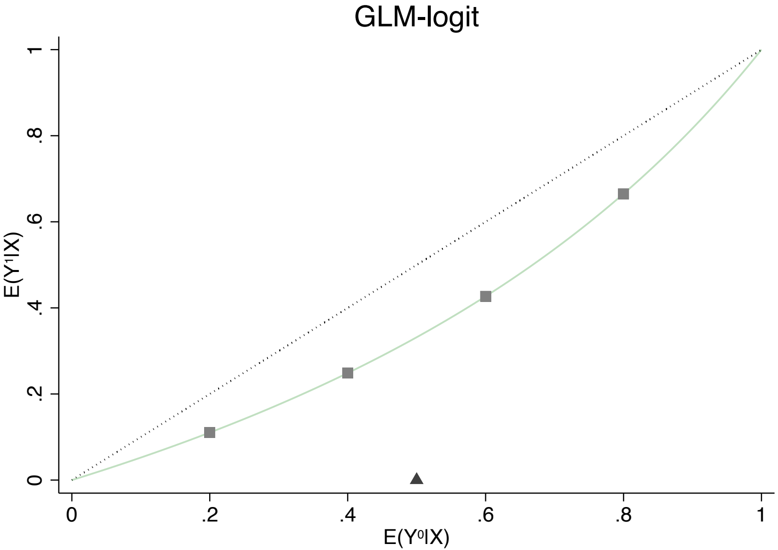
Suppose 4 (equiprobable) values of $\mathbb{E}(Y^0|X)$

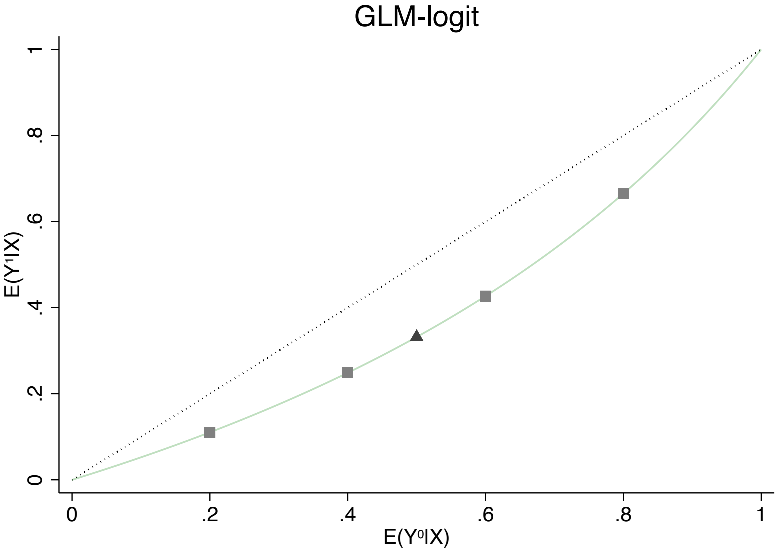


These map like this

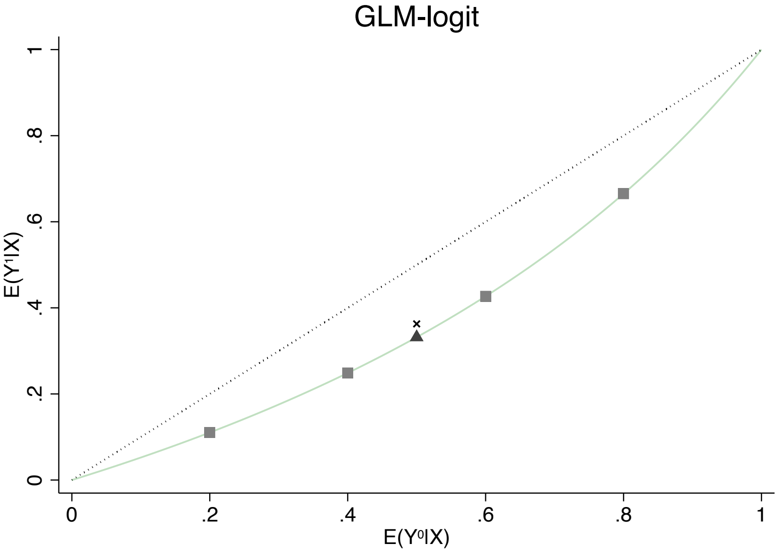


And the average $\mathbb{E}(Y^0)$...

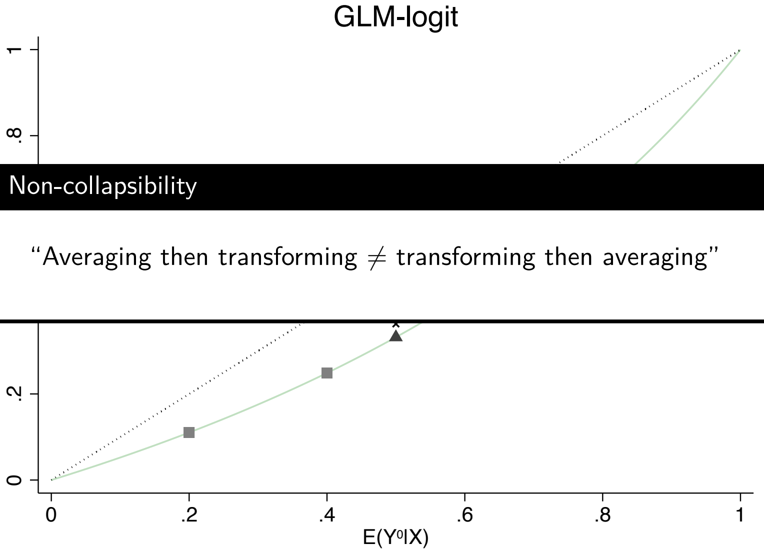



GLM-logit
E(Y¹|X)
E(Y⁰|X)
0
.2
.4
.6
.8
1

...which is NOT $\mathbb{E}(Y^1)$



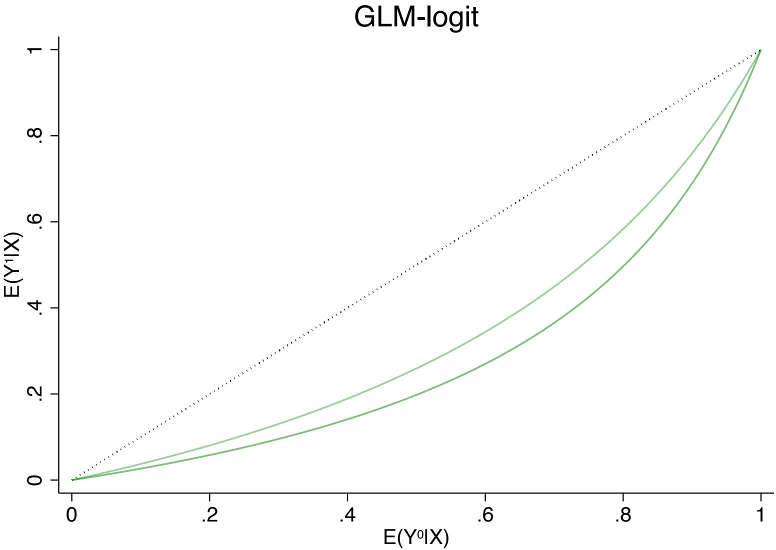
. . . which is NOT $\mathbb{E}(Y^1)$



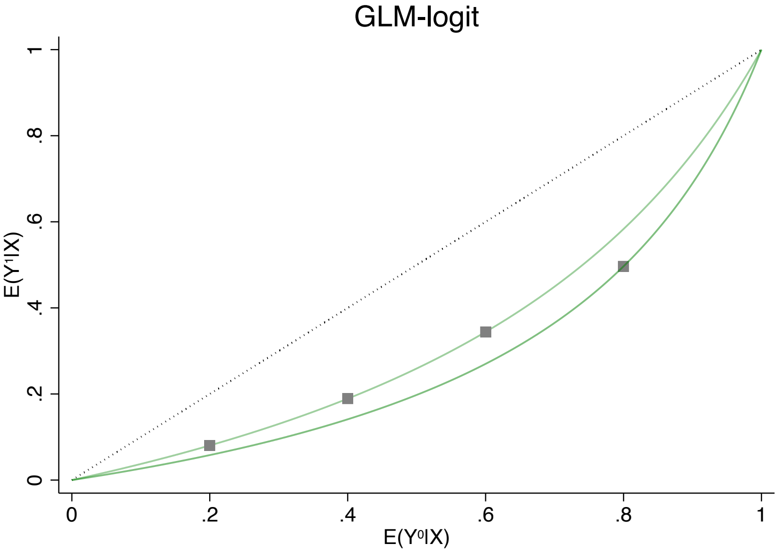
Non-collapsibility

"Averaging then transforming $\neq$ transforming then averaging"

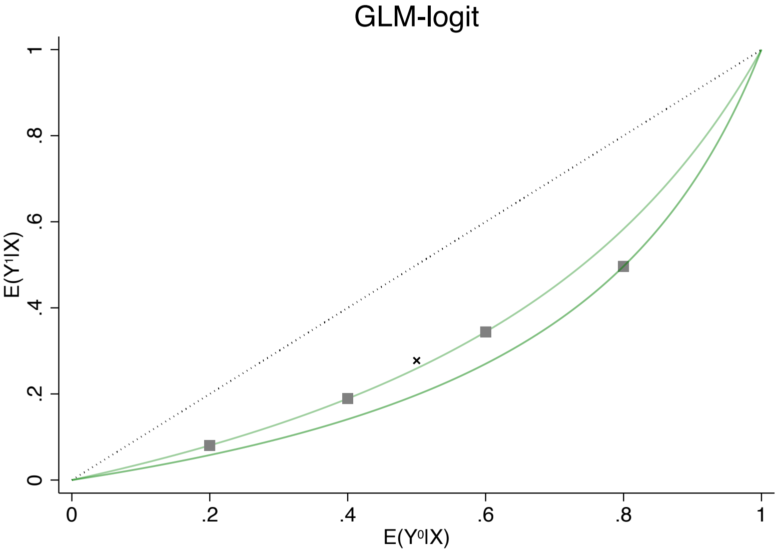
More generally, with heterogeneity. . .



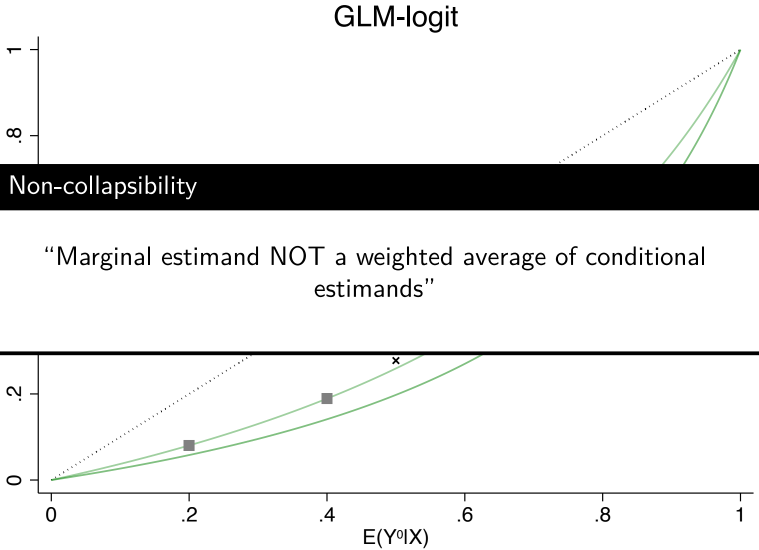
More generally, with heterogeneity...



More generally, with heterogeneity...



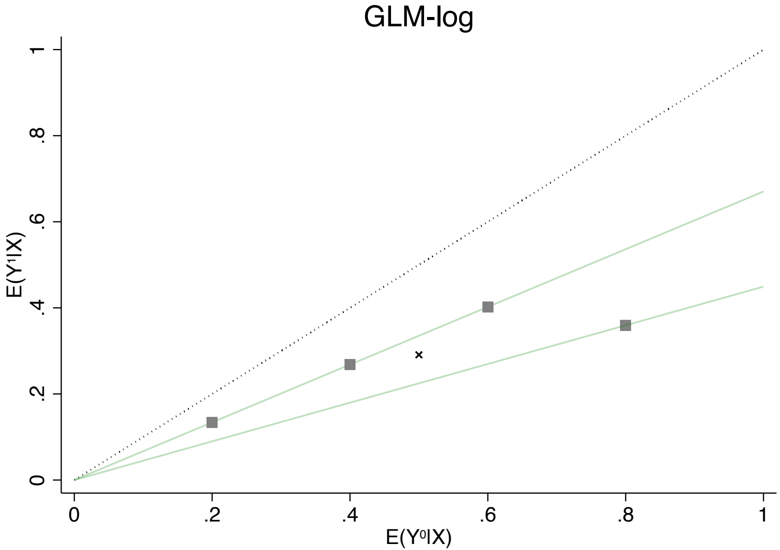
More generally, with heterogeneity...



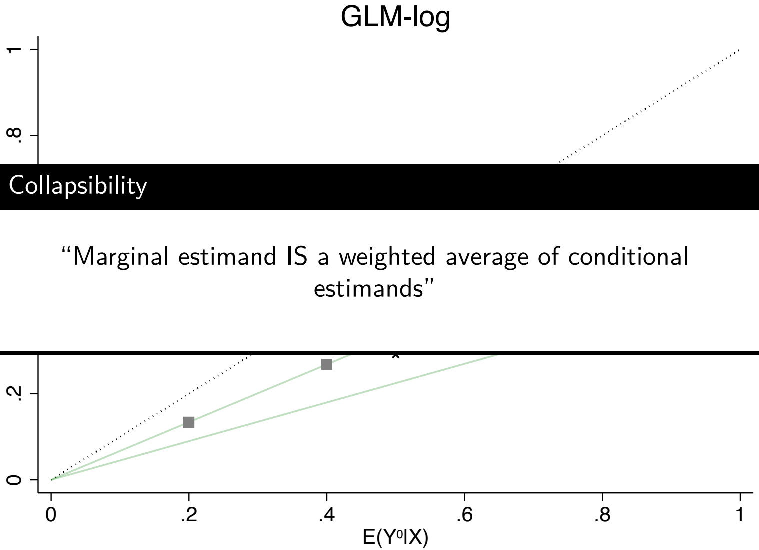
Non-collapsibility

"Marginal estimand NOT a weighted average of conditional estimands"

Whereas for the log link



Whereas for the log link



Collapsibility

"Marginal estimand IS a weighted average of conditional estimands"

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**
  - If non-collapsible, every estimand conditional on different sets of X's can each be different, and each different from the marginal.

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**
  - If non-collapsible, every estimand conditional on different sets of X's can each be different, and each different from the marginal.
- **Consequences for counterfactual risk prediction**

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**
  - If non-collapsible, every estimand conditional on different sets of X's can each be different, and each different from the marginal.
- **Consequences for counterfactual risk prediction**
  - Suppose a published study estimand is conditional on age, sex, comorbidities, genotype.

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**
  — If non-collapsible, every estimand conditional on different sets of X's can each be different, and each different from the marginal.
- **Consequences for counterfactual risk prediction**
  — Suppose a published study estimand is conditional on age, sex, comorbidities, genotype.
  — Suppose the published conditional odds ratio estimates for treatment vs. placebo are 0.6 for one genotype and 0.8 for the other.

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**
  - If non-collapsible, every estimand conditional on different sets of X's can each be different, and each different from the marginal.
- **Consequences for counterfactual risk prediction**
  - Suppose a published study estimand is conditional on age, sex, comorbidities, genotype.
  - Suppose the published conditional odds ratio estimates for treatment vs. placebo are 0.6 for one genotype and 0.8 for the other.
  - A decision is needed on whether or not to treat a new patient, whose $\mathbb{E}(Y^0|\text{age, sex, comorbidities})$ is thought to be ~ 0.2, but with genotype unknown.

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**
  - If non-collapsible, every estimand conditional on different sets of X's can each be different, and each different from the marginal.
- **Consequences for counterfactual risk prediction**
  - Suppose a published study estimand is conditional on age, sex, comorbidities, genotype.
  - Suppose the published conditional odds ratio estimates for treatment vs. placebo are 0.6 for one genotype and 0.8 for the other.
  - A decision is needed on whether or not to treat a new patient, whose $\mathbb{E}(Y^0|\text{age, sex, comorbidities})$ is thought to be ~ 0.2, but with genotype unknown.
  - Tempting to think that we can apply the estimated odds ratios of 0.6 and 0.8 to this 0.2 to bound $\mathbb{E}(Y^1|\text{age, sex, comorbidities})$, but we can't because of non-collapsibility.

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**
  - If non-collapsible, every estimand conditional on different sets of X's can each be different, and each different from the marginal.
- **Consequences for counterfactual risk prediction**
  - Suppose a published study estimand is conditional on age, sex, comorbidities, genotype.
  - Suppose the published conditional odds ratio estimates for treatment vs. placebo are 0.6 for one genotype and 0.8 for the other.
  - A decision is needed on whether or not to treat a new patient, whose $\mathbb{E}(Y^0|\text{age, sex, comorbidities})$ is thought to be ~ 0.2, but with genotype unknown.
  - Tempting to think that we can apply the estimated odds ratios of 0.6 and 0.8 to this 0.2 to bound $\mathbb{E}(Y^1|\text{age, sex, comorbidities})$, but we can't because of non-collapsibility.
- **The issue is not specific to binary outcomes**

Three more remarks about (non)collapsibility

- **It's not just about conditional vs. marginal**
  - If non-collapsible, every estimand conditional on different sets of X's can each be different, and each different from the marginal.
- **Consequences for counterfactual risk prediction**
  - Suppose a published study estimand is conditional on age, sex, comorbidities, genotype.
  - Suppose the published conditional odds ratio estimates for treatment vs. placebo are 0.6 for one genotype and 0.8 for the other.
  - A decision is needed on whether or not to treat a new patient, whose $\mathbb{E}(Y^0|\text{age, sex, comorbidities})$ is thought to be ~ 0.2, but with genotype unknown.
  - Tempting to think that we can apply the estimated odds ratios of 0.6 and 0.8 to this 0.2 to bound $\mathbb{E}(Y^1|\text{age, sex, comorbidities})$, but we can't because of non-collapsibility.
- **The issue is not specific to binary outcomes**
  - e.g. hazard ratios in Cox PH models are non-collapsible

Outline

1. Re-cap elements of previous talks
2. (Non)collapsibility
3. The choice between marginal and conditional estimands
4. For more. . .

Arguments made for marginal estimands

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

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  — What ‘population’?

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - What 'population'? Counter: always marginalising over *something*.

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.
  - Not an argument for an unadjusted analysis.

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.
  - Not 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.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.
  - Not 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.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.
  - Not 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.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.
  - Not 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.**
  - Intended as a defense of an unadjusted analysis?

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.
  - Not 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.**
  - Intended as a defense of an unadjusted analysis?
  - But recall Jonathan: randomisation often protects against misspecification *even when it may appear that* we are relying on additional assumptions in order to bring in X.

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.
  - Not 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.**
  - Intended as a defense of an unadjusted analysis?
  - But recall Jonathan: randomisation often protects against misspecification *even when it may appear that* we are relying on additional assumptions in order to bring in X.
- **Eliminates concerns over multiple testing / 'gaming'.**

Arguments made for marginal estimands

- **A single number with a simple(?) interpretation.**
  - What 'population'? Counter: always marginalising over *something*.
  - Conditional estimand can also be a single number.
  - Not 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.**
  - Intended as a defense of an unadjusted analysis?
  - But recall Jonathan: randomisation often protects against misspecification *even when it may appear that* we are relying on additional assumptions in order to bring in X.
- **Eliminates concerns over multiple testing / 'gaming'.**
  - Possibly, but pre-specification is possible even when using X: as discussed by Florian.

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.**
  — 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.**
  - 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.**
  - 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.
  - Choice of effect measure (odds ratio, risk ratio, . . .) also important when considering transportability, however.

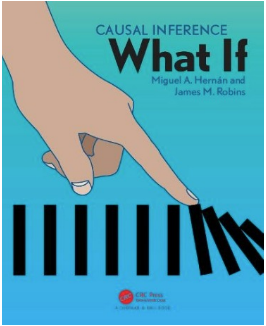
Arguments made for conditional estimands

- **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.**
  - 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.
  - Choice of effect measure (odds ratio, risk ratio, . . .) also important when considering transportability, however.
  - See work by Anders Huitfeldt.

Outline

1. Re-cap elements of previous talks
2. (Non)collapsibility
3. The choice between marginal and conditional estimands
4. For more. . .

More on (non)collapsibility



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