Marginalization of regression-adjusted treatment effects

Antonio Remiro-Azócar, MSc

R-HTA Showcase | 1 July 2021

About me

> My name is Antonio and I am a final-year PhD candidate based at the Statistics for Health Economic Evaluation Group at UCL, where I am supervised by Gianluca Baio and Anna Heath. I am also an independent consultant specializing in the statistical aspects of evidence synthesis, HTA and outcomes research.

About me

> My name is Antonio and I am a final-year PhD candidate based at the Statistics for Health Economic Evaluation Group at UCL, where I am supervised by Gianluca Baio and Anna Heath. I am also an independent consultant specializing in the statistical aspects of evidence synthesis, HTA and outcomes research.

Impact statement

> I recently submitted my PhD thesis. There are two sides to the story it tells. One addresses a substantive problem in HTA, which is the application of population-adjusted indirect comparisons (e.g. MAIC, STC). The other side of the story highlights the importance of carefully considering whether a marginal or conditional treatment effect is of interest in HTA.

Multivariable outcome regression in HTA

The treatment coefficient of a multivariable regression of outcome on treatment and baseline covariates often informs average effectiveness in health economic evaluations. Some examples:

Multivariable outcome regression in HTA

The treatment coefficient of a multivariable regression of outcome on treatment and baseline covariates often informs average effectiveness in health economic evaluations. Some examples:

1. Controlling for the effect of prognostic factors with individual patient data (IPD) from an observational or non-randomized study

2. Correcting for empirical confounding caused by chance imbalances in baseline covariates with IPD from a randomized controlled trial (RCT)

3. Accounting for differences in effect measure modifiers across a connected network of RCTs in a network meta-regression, either with IPD or aggregate-level data (ALD)

4. *Transporting* or *generalizing* inferences from a study lacking external validity to the target population for the decision

5. **Performing a pairwise population-adjusted indirect comparison to compare treatments with a common comparator arm across trials**

Multivariable outcome regression in HTA

The treatment coefficient of a multivariable regression of outcome on treatment and baseline covariates often informs average effectiveness in health economic evaluations. Some examples:

1. Controlling for the effect of prognostic factors with individual patient data (IPD) from an observational or non-randomized study

2. Correcting for empirical confounding caused by chance imbalances in baseline covariates with IPD from a randomized controlled trial (RCT)

3. Accounting for differences in effect measure modifiers across a connected network of RCTs in a network meta-regression, either with IPD or aggregate-level data (ALD)

4. *Transporting* or *generalizing* inferences from a study lacking external validity to the target population for the decision

5. **Performing a pairwise population-adjusted indirect comparison to compare treatments with a common comparator arm across trials**

This presentation deals with Scenario 5, a special case of Scenario 4

Issues

- The estimated effect may have a conditional interpretation, as opposed to the population-level interpretation that is required for reimbursement decisions made by bodies such as NICE
- When effect measures are non-collapsible, there may be sizable differences between marginal and conditional estimands, even in an ideal RCT
- Non-collapsibility occurs in logistic regression analysis for the odds ratio, in the Cox proportional hazards model for the hazard ratio, and for most measures of effect involving non-linear regressions
- Estimators targeting different estimands will have different variances for both collapsible and non-collapsible measures of effect. Hence, these quantify parametric uncertainty differently.
- This leads to the incorrect propagation of uncertainty to the wider health economic decision model. Dangerous for probabilistic sensitivity analyses.

Issues

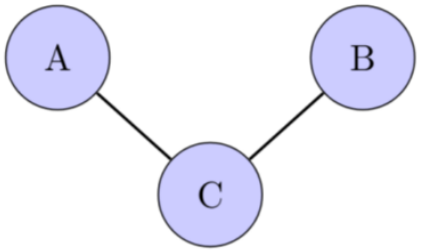
- The estimated effect may have a conditional interpretation, as opposed to the population-level interpretation that is required for reimbursement decisions made by bodies such as NICE
- When effect measures are non-collapsible, there may be sizable differences between marginal and conditional estimands, even in an ideal RCT
- Non-collapsibility occurs in logistic regression analysis for the odds ratio, in the Cox proportional hazards model for the hazard ratio, and for most measures of effect involving non-linear regressions
- Estimators targeting different estimands will have different variances for both collapsible and non-collapsible measures of effect. Hence, these quantify parametric uncertainty differently.
- This leads to the incorrect propagation of uncertainty to the wider health economic decision model. Dangerous for probabilistic sensitivity analyses.

The solution is the *marginalization* of the conditional effect estimates

Population-adjusted indirect comparisons

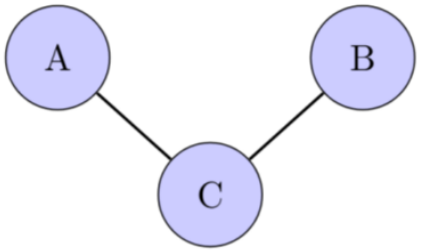
Population-adjusted indirect comparisons

Our case study is a very common scenario in oncology HTAs:



Population-adjusted indirect comparisons

Our case study is a very common scenario in oncology HTAs:



- Active treatment *A* needs to be compared to active treatment *B* for reimbursement purposes
- Anchored scenario: both treatments have been evaluated in RCTs against a common comparator *C*, but not against each other
- The manufacturer submitting evidence to HTA bodies has access to IPD from its own *AC* RCT. No IPD, only published ALD, are available for the competitor's *BC* RCT.
- Standard methods are biased where there is treatment effect heterogeneity over variables that vary in distribution across trials

Requirements

Covariate-adjusted effect for *A* vs. *B* estimated in the *BC* population. Indirect comparison carried out in the “linear predictor” scale; using additive effects for a given linear predictor:

$$\hat{\Delta}_{AB}^{(BC)} = \hat{\Delta}_{AC}^{(BC)} - \hat{\Delta}_{BC}^{(BC)}$$

Requirements

Covariate-adjusted effect for *A* vs. *B* estimated in the *BC* population. Indirect comparison carried out in the "linear predictor" scale; using additive effects for a given linear predictor:

$$\hat{\Delta}_{AB}^{(BC)} = \hat{\Delta}_{AC}^{(BC)} - \hat{\Delta}_{BC}^{(BC)}$$

- $\hat{\Delta}_{BC}^{(BC)}$ is the estimated *marginal* effect of *B* vs. *C*, available from the RCT publication. Any conditional estimate is likely incompatible.
- $\hat{\Delta}_{AB}^{(BC)}$ should target a *marginal* effect to inform reimbursement decisions at the population level
- $\hat{\Delta}_{AC}^{(BC)}$ must target a *marginal* effect that is compatible with $\hat{\Delta}_{BC}^{(BC)}$. Estimand incompatibility may produce bias (Remiro-Azócar et al. 2021a).

Requirements

Covariate-adjusted effect for *A* vs. *B* estimated in the *BC* population. Indirect comparison carried out in the "linear predictor" scale; using additive effects for a given linear predictor:

$$\hat{\Delta}_{AB}^{(BC)} = \hat{\Delta}_{AC}^{(BC)} - \hat{\Delta}_{BC}^{(BC)}$$

- $\hat{\Delta}_{BC}^{(BC)}$ is the estimated *marginal* effect of *B* vs. *C*, available from the RCT publication. Any conditional estimate is likely incompatible.
- $\hat{\Delta}_{AB}^{(BC)}$ should target a *marginal* effect to inform reimbursement decisions at the population level
- $\hat{\Delta}_{AC}^{(BC)}$ must target a *marginal* effect that is compatible with $\hat{\Delta}_{BC}^{(BC)}$. Estimand incompatibility may produce bias (Remiro-Azócar et al. 2021a).

Weighting (MAIC) or outcome regression can be used to generate $\hat{\Delta}_{AC}^{(BC)}$. Outcome regression is more statistically precise and efficient than weighting (Remiro-Azócar et al. 2021b).

Setup

```
## load packages
# for non-parametric bootstrap in maximum-likelihood G-computation
if (!require("boot")) install.packages("boot")
# for simulating BC (ALD study) covariates from Gaussian copula
if (!require("copula")) install.packages("copula")
# for outcome regression and prediction in Bayesian G-computation
if (!require("rstanarm")) install.packages("rstanarm")

set.seed(555) # set seed for reproducibility
rm(list = ls(all = TRUE)) # clear directory
```

Setup

```
## load packages
# for non-parametric bootstrap in maximum-likelihood G-computation
if (!require("boot")) install.packages("boot")
# for simulating BC (ALD study) covariates from Gaussian copula
if (!require("copula")) install.packages("copula")
# for outcome regression and prediction in Bayesian G-computation
if (!require("rstanarm")) install.packages("rstanarm")

set.seed(555) # set seed for reproducibility
rm(list = ls(all = TRUE)) # clear directory


getRversion()
```

```
## [1] '4.1.0'
```

Setup

```
## load packages
# for non-parametric bootstrap in maximum-likelihood G-computation
if (!require("boot")) install.packages("boot")
# for simulating BC (ALD study) covariates from Gaussian copula
if (!require("copula")) install.packages("copula")
# for outcome regression and prediction in Bayesian G-computation
if (!require("rstanarm")) install.packages("rstanarm")

set.seed(555) # set seed for reproducibility
rm(list = ls(all = TRUE)) # clear directory


getRversion()
```

```
## [1] '4.1.0'
```

Data

```
# Load fake (simulated) data
AC.IPD <- read.csv("AC_IPD.csv") # load AC patient-level data
BC.ALD <- read.csv("BC_ALD.csv") # load BC aggregate-level data
```

The *AC* IPD (200 subjects) consists of individual-level baseline covariates, treatment and binary outcomes, e.g. the occurrence of an adverse event

```
knitr::kable(round(head(AC.IPD),digits=2))
```

X1	X2	X3	X4	trt	y
0.44	0.67	0.93	0.09	1	0
0.06	0.60	0.04	0.60	1	1
-0.08	0.68	0.93	-0.11	1	0
-0.39	0.57	-0.32	0.03	1	0
1.01	0.82	0.93	0.84	1	1
0.19	0.20	0.35	0.16	1	0

The *AC* IPD (200 subjects) consists of individual-level baseline covariates, treatment and binary outcomes, e.g. the occurrence of an adverse event

```
knitr::kable(round(head(AC.IPD),digits=2))
```

X1	X2	X3	X4	trt	y
0.44	0.67	0.93	0.09	1	0
0.06	0.60	0.04	0.60	1	1
-0.08	0.68	0.93	-0.11	1	0
-0.39	0.57	-0.32	0.03	1	0
1.01	0.82	0.93	0.84	1	1
0.19	0.20	0.35	0.16	1	0

The *BC* ALD (600 subjects) consists of aggregate-level baseline covariates and summary outcomes, i.e., the marginal covariate moments ("Table 1" of the RCT publication) and a contingency table for the event counts

```
round(BC.ALD,digits=2)
```

```
##   mean.X1 mean.X2 mean.X3 mean.X4 sd.X1 sd.X2 sd.X3 sd.X4 y.B.sum y.B.bar N.B
## 1    0.59    0.64    0.59     0.6  0.39   0.4  0.41   0.4     182    0.46 400
##   y.C.sum y.C.bar N.C
## 1     149    0.74 200
```

Covariate simulation

We will marginalize with respect to a hypothetical *BC* pseudo-population. Individual-level covariates $\boldsymbol{x}^*$ are generated using a Gaussian copula.

We use normally-distributed marginals with the *BC* means and standard deviations, and the pairwise linear correlations of the *AC* IPD. $N^* = 1000$ subjects are simulated, large enough to minimize sampling variability.

Covariate simulation

We will marginalize with respect to a hypothetical *BC* pseudo-population. Individual-level covariates $\boldsymbol{x}^*$ are generated using a Gaussian copula.

We use normally-distributed marginals with the *BC* means and standard deviations, and the pairwise linear correlations of the *AC* IPD. $N^* = 1000$ subjects are simulated, large enough to minimize sampling variability.

```
# matrix of pairwise correlations between IPD covariates
rho <- cor(AC.IPD[,c("X1","X2","X3","X4")])
#  covariate simulation for BC trial using copula package
cop <- normalCopula(param=c(rho[1,2],rho[1,3],rho[1,4],rho[2,3],
                            rho[2,4],rho[3,4]),
                    dim=4, dispstr="un") # AC IPD pairwise correlations
# sample covariates from approximate joint distribution using copula
mvd <- mvdc(copula=cop, margins=c("norm", "norm", # Gaussian marginals
                                  "norm", "norm"),
            # BC covariate means and standard deviations
            paramMargins=list(list(mean=BC.ALD$mean.X1, sd=BC.ALD$sd.X1),
                              list(mean=BC.ALD$mean.X2, sd=BC.ALD$sd.X2),
                              list(mean=BC.ALD$mean.X3, sd=BC.ALD$sd.X3),
                              list(mean=BC.ALD$mean.X4, sd=BC.ALD$sd.X4)))
# simulated BC pseudo-population of size 1000 to stabilize sampling distribution
x_star <- as.data.frame(rMvdc(n=1000, mvd))
colnames(x_star) <- c("X1", "X2", "X3", "X4")
```

Outcome regression

Our working regression is a generalized linear model of the observed outcome $\boldsymbol{y}$ on the covariates $\boldsymbol{x}$ and treatment $\boldsymbol{t}$, fitted to the *AC* IPD:

$$g(\mu_n) = \beta_0 + \boldsymbol{x}_n \boldsymbol{\beta_1} + \left(\beta_t + \boldsymbol{x}_n^{\boldsymbol{(EM)}} \boldsymbol{\beta_2}\right) \mathbf{1}\left(t_n = A\right)$$

Outcome regression

Our working regression is a generalized linear model of the observed outcome $\boldsymbol{y}$ on the covariates $\boldsymbol{x}$ and treatment $\boldsymbol{t}$, fitted to the *AC* IPD:

$$g(\mu_n) = \beta_0 + \boldsymbol{x}_n \boldsymbol{\beta_1} + \left(\beta_t + \boldsymbol{x}_n^{(\boldsymbol{EM})} \boldsymbol{\beta_2}\right) \mathbf{1}\left(t_n = A\right)$$

- μ_n: expected outcome of subject n on the natural outcome scale, e.g. the probability scale for binary outcomes
- $g(\cdot)$: appropriate invertible canonical link function, e.g. the $\text{logit}(\mu_n) = \ln\left(\mu_n / (1 - \mu_n)\right)$ for binary outcomes in logistic regression
- $\boldsymbol{\beta_1}$: vector of regression coefficients for the prognostic variables
- $\boldsymbol{\beta_2}$: vector of interaction coefficients for the effect modifiers (modifying the effect of treatment *A* vs. *C*)
- β_t: conditional treatment effect for *A* vs. *C*

Outcome regression

Our working regression is a generalized linear model of the observed outcome $\boldsymbol{y}$ on the covariates $\boldsymbol{x}$ and treatment $\boldsymbol{t}$, fitted to the *AC* IPD:

$$g(\mu_n) = \beta_0 + \boldsymbol{x}_n\boldsymbol{\beta_1} + \left(\beta_t + \boldsymbol{x}_n^{(\boldsymbol{EM})}\boldsymbol{\beta_2}\right)\mathbf{1}\left(t_n = A\right)$$

- μ_n: expected outcome of subject n on the natural outcome scale, e.g. the probability scale for binary outcomes
- $g(\cdot)$: appropriate invertible canonical link function, e.g. the $\text{logit}(\mu_n) = \ln\left(\mu_n/(1-\mu_n)\right)$ for binary outcomes in logistic regression
- $\boldsymbol{\beta_1}$: vector of regression coefficients for the prognostic variables
- $\boldsymbol{\beta_2}$: vector of interaction coefficients for the effect modifiers (modifying the effect of treatment *A* vs. *C*)
- β_t: conditional treatment effect for *A* vs. *C*

In the context of G-computation, the working model is called the *Q-model*

Parametric G-computation

The goal is to *integrate*, *average* or *marginalize* out the model for the conditional expectation over the relevant joint covariate distribution

Parametric G-computation

The goal is to *integrate*, *average* or *marginalize* out the model for the conditional expectation over the relevant joint covariate distribution

Maximum-likelihood estimation (MLE)

Fit the Q-model to the AC IPD, $\mathcal{D}_{AC} = (\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{y})$, using MLE:

```
# outcome logistic regression fitted to IPD using maximum likelihood
outcome.model <- glm(y~X3+X4+trt*X1+trt*X2, data=AC.IPD, family=binomial)
```

Parametric G-computation

The goal is to *integrate*, *average* or *marginalize* out the model for the conditional expectation over the relevant joint covariate distribution

Maximum-likelihood estimation (MLE)

Fit the Q-model to the *AC* IPD, $\mathcal{D}_{AC} = (\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{y})$, using MLE:

```
# outcome logistic regression fitted to IPD using maximum likelihood
outcome.model <- glm(y~X3+X4+trt*X1+trt*X2, data=AC.IPD, family=binomial)
```

Leaving the simulated covariates $\boldsymbol{x}^*$ at their set values, we apply the maximum-likelihood coefficients $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\boldsymbol{\beta}}_{\mathbf{1}}, \hat{\boldsymbol{\beta}}_{\mathbf{2}}, \hat{\beta}_t)$ to predict a pair of hypothetical outcomes for each subject (under treatments *A* and *C*):

```
# hypothetical datasets
data.trtA <- data.trtC <- x_star
# intervene on treatment while keeping set covariates fixed
data.trtA$trt <- 1 # dataset where everyone receives treatment A
data.trtC$trt <- 0 # dataset where all observations receive C
```

Plug treatment A into the regression fit for each simulated individual to compute the conditional expectation when all subjects are under A:

$\hat{\mu}_A(\boldsymbol{x^*}) = \frac{1}{N^*} \sum_{i=1}^{N^*} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_i^* \hat{\boldsymbol{\beta}}_{\mathbf{1}} + \hat{\beta}_t + \boldsymbol{x}_i^{*(\boldsymbol{EM})} \hat{\boldsymbol{\beta}}_{\mathbf{2}})$

Plug treatment A into the regression fit for each simulated individual to compute the conditional expectation when all subjects are under A:

$$\hat{\mu}_A(\boldsymbol{x^*}) = \frac{1}{N^*} \sum_{i=1}^{N^*} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_i^* \hat{\boldsymbol{\beta}}_{\mathbf{1}} + \hat{\beta}_t + \boldsymbol{x}_i^{*(\boldsymbol{EM})} \hat{\boldsymbol{\beta}}_{\mathbf{2}})$$

```
# predict hypothetical event probs, conditional on treatment/covariates
hat.mu.A.i <- predict(outcome.model, type="response", newdata=data.trtA)
data.trtA$hat.mu <- hat.mu.A.i
hat.mu.A <- mean(hat.mu.A.i) # mean probability prediction under A
```

Plug treatment A into the regression fit for each simulated individual to compute the conditional expectation when all subjects are under A:

$$\hat{\mu}_A(\boldsymbol{x^*}) = \frac{1}{N^*}\sum_{i=1}^{N^*} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_i^*\hat{\boldsymbol{\beta}}_{\mathbf{1}} + \hat{\beta}_t + \boldsymbol{x}_i^{*(\boldsymbol{EM})}\hat{\boldsymbol{\beta}}_{\mathbf{2}})$$

```
# predict hypothetical event probs, conditional on treatment/covariates
hat.mu.A.i <- predict(outcome.model, type="response", newdata=data.trtA)
data.trtA$hat.mu <- hat.mu.A.i
hat.mu.A <- mean(hat.mu.A.i) # mean probability prediction under A
```

By plugging treatment C into the regression fit for every simulated observation, we obtain the mean predicted outcome when all units are under C:

$$\hat{\mu}_C(\boldsymbol{x^*}) = \frac{1}{N^*}\sum_{i=1}^{N^*} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_i^*\hat{\boldsymbol{\beta}}_{\mathbf{1}})$$

Plug treatment A into the regression fit for each simulated individual to compute the conditional expectation when all subjects are under A:

$$\hat{\mu}_A(\boldsymbol{x^*}) = \frac{1}{N^*}\sum_{i=1}^{N^*} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_i^*\hat{\boldsymbol{\beta}}_{\mathbf{1}} + \hat{\beta}_t + \boldsymbol{x}_i^{*(\boldsymbol{EM})}\hat{\boldsymbol{\beta}}_{\mathbf{2}})$$

```
# predict hypothetical event probs, conditional on treatment/covariates
hat.mu.A.i <- predict(outcome.model, type="response", newdata=data.trtA)
data.trtA$hat.mu <- hat.mu.A.i
hat.mu.A <- mean(hat.mu.A.i) # mean probability prediction under A
```

By plugging treatment C into the regression fit for every simulated observation, we obtain the mean predicted outcome when all units are under C:

$$\hat{\mu}_C(\boldsymbol{x^*}) = \frac{1}{N^*}\sum_{i=1}^{N^*} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_i^*\hat{\boldsymbol{\beta}}_{\mathbf{1}})$$

```
# predict hypothetical event probs, conditional on treatment/covariates
hat.mu.C.i <- predict(outcome.model, type="response", newdata=data.trtC)
data.trtC$hat.mu <- hat.mu.C.i
hat.mu.C <- mean(hat.mu.C.i) # mean probability prediction under C
```

Plug treatment A into the regression fit for each simulated individual to compute the conditional expectation when all subjects are under A:

$$\hat{\mu}_A(\boldsymbol{x^*}) = \frac{1}{N^*} \sum_{i=1}^{N^*} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_i^* \hat{\boldsymbol{\beta}}_{\mathbf{1}} + \hat{\beta}_t + \boldsymbol{x}_i^{*(\boldsymbol{EM})} \hat{\boldsymbol{\beta}}_{\mathbf{2}})$$

```
# predict hypothetical event probs, conditional on treatment/covariates
hat.mu.A.i <- predict(outcome.model, type="response", newdata=data.trtA)
data.trtA$hat.mu <- hat.mu.A.i
hat.mu.A <- mean(hat.mu.A.i) # mean probability prediction under A
```

By plugging treatment C into the regression fit for every simulated observation, we obtain the mean predicted outcome when all units are under C:

$$\hat{\mu}_C(\boldsymbol{x^*}) = \frac{1}{N^*} \sum_{i=1}^{N^*} g^{-1}(\hat{\beta}_0 + \boldsymbol{x}_i^* \hat{\boldsymbol{\beta}}_{\mathbf{1}})$$

```
# predict hypothetical event probs, conditional on treatment/covariates
hat.mu.C.i <- predict(outcome.model, type="response", newdata=data.trtC)
data.trtC$hat.mu <- hat.mu.C.i
hat.mu.C <- mean(hat.mu.C.i) # mean probability prediction under C
```

We now have two *counterfactual* datasets: what outcomes might have been observed had subjects in a different population, in which the *A* vs. *C* trial was not conducted, received treatment?

The *BC* pseudo-population under treatment *A*:

```
knitr::kable(round(head(data.trtA),digits=2))
```

X1	X2	X3	X4	trt	hat.mu
-0.77	0.25	0.22	0.08	1	0.10
0.49	0.67	0.13	0.81	1	0.40
0.39	1.23	0.74	0.99	1	0.81
0.54	0.63	0.66	0.24	1	0.38
1.32	0.41	0.59	0.54	1	0.42
0.17	0.64	0.16	-0.22	1	0.13

The *BC* pseudo-population under treatment *A*:

```
knitr::kable(round(head(data.trtA),digits=2))
```

X1	X2	X3	X4	trt	hat.mu
-0.77	0.25	0.22	0.08	1	0.10
0.49	0.67	0.13	0.81	1	0.40
0.39	1.23	0.74	0.99	1	0.81
0.54	0.63	0.66	0.24	1	0.38
1.32	0.41	0.59	0.54	1	0.42
0.17	0.64	0.16	-0.22	1	0.13

The *BC* pseudo-population under treatment *C*:

```
knitr::kable(round(head(data.trtC),digits=2))
```

X1	X2	X3	X4	trt	hat.mu
-0.77	0.25	0.22	0.08	0	0.47
0.49	0.67	0.13	0.81	0	0.75
0.39	1.23	0.74	0.99	0	0.97
0.54	0.63	0.66	0.24	0	0.72
1.32	0.41	0.59	0.54	0	0.62
0.17	0.64	0.16	-0.22	0	0.45

Estimate the marginal treatment effect for *A* vs. *C* by transforming from the natural outcome scale to the linear predictor scale and calculating the difference between the average linear predictions:

$$\hat{\Delta}_{AC}^{(BC)} = g\left(\hat{\mu}_A(\boldsymbol{x}^*)\right) - g\left(\hat{\mu}_C(\boldsymbol{x}^*)\right)$$

Estimate the marginal treatment effect for A vs. C by transforming from the natural outcome scale to the linear predictor scale and calculating the difference between the average linear predictions:

$$\hat{\Delta}_{AC}^{(BC)} = g\left(\hat{\mu}_A(\boldsymbol{x}^*)\right) - g\left(\hat{\mu}_C(\boldsymbol{x}^*)\right)$$

```
# marginal A vs. C log-odds ratio (mean difference in expected log-odds)
# estimated by transforming from probability to linear predictor scale
hat.Delta.AC <- log(hat.mu.A/(1-hat.mu.A)) - log(hat.mu.C/(1-hat.mu.C))
# hat.Delta.AC <- qlogis(hat.mu.A) - qlogis(hat.mu.C)
hat.Delta.AC
```

```
## [1] -1.106043
```

Estimate the marginal treatment effect for A vs. C by transforming from the natural outcome scale to the linear predictor scale and calculating the difference between the average linear predictions:

$$\hat{\Delta}_{AC}^{(BC)} = g\left(\hat{\mu}_A(\boldsymbol{x}^*)\right) - g\left(\hat{\mu}_C(\boldsymbol{x}^*)\right)$$

```
# marginal A vs. C log-odds ratio (mean difference in expected log-odds)
# estimated by transforming from probability to linear predictor scale
hat.Delta.AC <- log(hat.mu.A/(1-hat.mu.A)) - log(hat.mu.C/(1-hat.mu.C))
# hat.Delta.AC <- qlogis(hat.mu.A) - qlogis(hat.mu.C)
hat.Delta.AC
```

```
## [1] -1.106043
```

Different summary measures of the marginal contrast, e.g. odds ratios, relative risks or risk differences, can be produced by manipulating the conditional expectations differently, mapping these to other scales:

```
MOR <- (hat.mu.A/(1-hat.mu.A))/(hat.mu.C/(1-hat.mu.C)) # marginal odds ratio for
MRR <- hat.mu.A/hat.mu.C # marginal relative risk for A vs. C
MRD <- hat.mu.A-hat.mu.C # marginal risk difference for A vs. C
```

Estimate the marginal treatment effect for A vs. C by transforming from the natural outcome scale to the linear predictor scale and calculating the difference between the average linear predictions:

$$\hat{\Delta}_{AC}^{(BC)} = g\left(\hat{\mu}_A(\boldsymbol{x}^*)\right) - g\left(\hat{\mu}_C(\boldsymbol{x}^*)\right)$$

```
# marginal A vs. C log-odds ratio (mean difference in expected log-odds)
# estimated by transforming from probability to linear predictor scale
hat.Delta.AC <- log(hat.mu.A/(1-hat.mu.A)) - log(hat.mu.C/(1-hat.mu.C))
# hat.Delta.AC <- qlogis(hat.mu.A) - qlogis(hat.mu.C)
hat.Delta.AC
```

```
## [1] -1.106043
```

Different summary measures of the marginal contrast, e.g. odds ratios, relative risks or risk differences, can be produced by manipulating the conditional expectations differently, mapping these to other scales:

```
MOR <- (hat.mu.A/(1-hat.mu.A))/(hat.mu.C/(1-hat.mu.C)) # marginal odds ratio for
MRR <- hat.mu.A/hat.mu.C # marginal relative risk for A vs. C
MRD <- hat.mu.A-hat.mu.C # marginal risk difference for A vs. C
```

The estimated absolute outcomes $\hat{\mu}_A(\boldsymbol{x}^*)$ and $\hat{\mu}_C(\boldsymbol{x}^*)$ are sometimes desirable in health economic models and in unanchored comparisons

Variance estimation

It is not easy to derive the standard error analytically when the marginal estimate is a non-linear function of the components of $\hat{\beta}$

We shall resample using the ordinary non-parametric bootstrap

Variance estimation

It is not easy to derive the standard error analytically when the marginal estimate is a non-linear function of the components of $\hat{\beta}$

We shall resample using the ordinary non-parametric bootstrap

```
# this function will be bootstrapped
gcomp.ml <- function(data, indices) {
  dat = data[indices,]
  # outcome logistic regression fitted to IPD using maximum likelihood
  outcome.model <- glm(y~X3+X4+trt*X1+trt*X2, data=dat, family=binomial)
  # hypothetical datasets
  data.trtA <- data.trtC <- x_star
  # intervene on treatment while keeping set covariates fixed
  data.trtA$trt <- 1 # dataset where everyone receives treatment A
  data.trtC$trt <- 0 # dataset where all observations receive C
  # predict hypothetical event probs, conditional on treatment/covariates
  hat.mu.A.i <- predict(outcome.model, type="response", newdata=data.trtA)
  hat.mu.C.i <- predict(outcome.model, type="response", newdata=data.trtC)
  hat.mu.A <- mean(hat.mu.A.i) # mean probability prediction under A
  hat.mu.C <- mean(hat.mu.C.i) # mean probability prediction under C
  # marginal A vs. C log-odds ratio (mean difference in expected log-odds)
  # estimated by transforming from probability to linear predictor scale
  hat.Delta.AC <- log(hat.mu.A/(1-hat.mu.A)) - log(hat.mu.C/(1-hat.mu.C))
  # hat.Delta.AC <- qlogis(hat.mu.A) - qlogis(hat.mu.C)
  return(hat.Delta.AC)
}
```

We use 1,000 resamples, as increasing further the number of resamples produces minimal gains in estimation precision and accuracy

```
# non-parametric bootstrap with 1000 resamples
resamples <- 1000
boot.object <- boot::boot(data=AC.IPD, statistic=gcomp.ml, R=resamples)
```

We use 1,000 resamples, as increasing further the number of resamples produces minimal gains in estimation precision and accuracy

```
# non-parametric bootstrap with 1000 resamples
resamples <- 1000
boot.object <- boot::boot(data=AC.IPD, statistic=gcomp.ml, R=resamples)
```

We can recover a point estimate $\hat{\Delta}_{AC}^{(BC)}$ of the marginal effect for A vs. C in the BC population. This is the average across the resamples:

```
# bootstrap mean of marginal A vs. C treatment effect estimate
hat.Delta.AC <- mean(boot.object$t)
hat.Delta.AC
```

```
## [1] -1.119884
```

We use 1,000 resamples, as increasing further the number of resamples produces minimal gains in estimation precision and accuracy

```
# non-parametric bootstrap with 1000 resamples
resamples <- 1000
boot.object <- boot::boot(data=AC.IPD, statistic=gcomp.ml, R=resamples)
```

We can recover a point estimate $\hat{\Delta}_{AC}^{(BC)}$ of the marginal effect for A vs. C in the BC population. This is the average across the resamples:

```
# bootstrap mean of marginal A vs. C treatment effect estimate
hat.Delta.AC <- mean(boot.object$t)
hat.Delta.AC
```

```
## [1] -1.119884
```

An estimate of the variance is the sample variance across the resamples:

```
# bootstrap variance of A vs. C treatment effect estimate
hat.var.Delta.AC <- var(boot.object$t)
hat.var.Delta.AC
```

```
##             [,1]
## [1,] 0.09622426
```

Bayesian parametric G-computation

Fit the Q-model using Markov chain Monte Carlo (MCMC). We use default "weakly informative" priors, 2 Markov chains with 4,000 iterations each (2,000 warmup), which gives $L = 4000$ iterations in total for the analysis.

```
# outcome logistic regression fitted to IPD using MCMC (Stan)
outcome.model <- stan_glm(y~X3+X4+trt*X1+trt*X2, data=AC.IPD, family=binomial,
                          algorithm="sampling", iter=4000, warmup=2000, chains=2
```

Bayesian parametric G-computation

Fit the Q-model using Markov chain Monte Carlo (MCMC). We use default "weakly informative" priors, 2 Markov chains with 4,000 iterations each (2,000 warmup), which gives $L = 4000$ iterations in total for the analysis.

```
# outcome logistic regression fitted to IPD using MCMC (Stan)
outcome.model <- stan_glm(y~X3+X4+trt*X1+trt*X2, data=AC.IPD, family=binomial,
                          algorithm="sampling", iter=4000, warmup=2000, chains=2
```

We now marginalize over the joint posterior distribution of the conditional regression parameters $\boldsymbol{\beta}$, as well as the joint BC covariate distribution

We draw a vector $\boldsymbol{y}^*_{t^*}$ of size N^* of predicted outcomes under each intervention $t^* \in \{A, C\}$ from its posterior predictive distribution:

$$
\begin{aligned}
p(\boldsymbol{y}^*_{t^*} \mid \mathcal{D}_{AC}) &= \int_{\boldsymbol{\beta}} p(\boldsymbol{y}^*_{t^*} \mid \boldsymbol{\beta}) p(\boldsymbol{\beta} \mid \mathcal{D}_{AC}) d\boldsymbol{\beta} \\
&= \int_{\boldsymbol{x}^*} p(\boldsymbol{y}^* \mid \boldsymbol{t}^*, \boldsymbol{x}^*, \mathcal{D}_{AC}) p(\boldsymbol{x}^* \mid \mathcal{D}_{AC}) d\boldsymbol{x}^* \\
&= \int_{\boldsymbol{x}^*} \int_{\boldsymbol{\beta}} p(\boldsymbol{y}^* \mid \boldsymbol{t}^*, \boldsymbol{x}^*, \boldsymbol{\beta}) p(\boldsymbol{x}^* \mid \boldsymbol{\beta}) p(\boldsymbol{\beta} \mid \mathcal{D}_{AC}) d\boldsymbol{\beta} d\boldsymbol{x}^*
\end{aligned}
$$

The integrals can be approximated numerically using MCMC sampling

Leaving the simulated covariates at their set values, we fix the value of treatment to create counterfactual datasets under A and C

The integrals can be approximated numerically using MCMC sampling

Leaving the simulated covariates at their set values, we fix the value of treatment to create counterfactual datasets under *A* and *C*

Where all simulated subjects are under treatment A, the l-th draw of the conditional expectation for subject i is:

$$\hat{\mu}_{A,i}^{(l)} = g^{-1}(\hat{\beta}_0^{(l)} + \boldsymbol{x}_i^* \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(l)} + \hat{\beta}_t^{(l)} + \boldsymbol{x}_i^{*(\boldsymbol{EM})} \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{(l)}).$$

Above, $\hat{\boldsymbol{\beta}}^{(l)} = (\hat{\beta}_0^{(l)}, \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(l)}, \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{(l)}, \hat{\beta}_t^{(l)})$ is the l-th posterior draw of the regression coefficients

Where all simulated subjects are set to treatment C, the l-th draw of the conditional expectation for subject i is:

$$\hat{\mu}_{C,i}^{(l)} = g^{-1}(\hat{\beta}_0^{(l)} + \boldsymbol{x}_i^* \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(l)})$$

The integrals can be approximated numerically using MCMC sampling

Leaving the simulated covariates at their set values, we fix the value of treatment to create counterfactual datasets under *A* and *C*

Where all simulated subjects are under treatment A, the l-th draw of the conditional expectation for subject i is:

$$\hat{\mu}_{A,i}^{(l)} = g^{-1}(\hat{\beta}_0^{(l)} + \boldsymbol{x}_i^* \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(l)} + \hat{\beta}_t^{(l)} + \boldsymbol{x}_i^{*(\boldsymbol{EM})} \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{(l)}).$$

Above, $\hat{\boldsymbol{\beta}}^{(l)} = (\hat{\beta}_0^{(l)}, \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(l)}, \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{(l)}, \hat{\beta}_t^{(l)})$ is the l-th posterior draw of the regression coefficients

Where all simulated subjects are set to treatment C, the l-th draw of the conditional expectation for subject i is:

$$\hat{\mu}_{C,i}^{(l)} = g^{-1}(\hat{\beta}_0^{(l)} + \boldsymbol{x}_i^* \hat{\boldsymbol{\beta}}_{\mathbf{1}}^{(l)})$$

These are used to impute the individual-level outcomes as independent draws from their posterior predictive distribution at each iteration, e.g. for logistic regression:

$$y_{t^*,i}^{*(l)} \sim \text{Bernoulli}(\hat{\mu}_{t^*,i}^{(l)})$$

Producing draws from the posterior predictive distribution of outcomes is fairly simple using the package `rstanarm`:

```
# draw binary responses from posterior predictive distribution
# LxN* matrix of posterior predictive draws under A
y.star.A <- posterior_predict(outcome.model, newdata=data.trtA)
# LxN* matrix of posterior predictive draws under C
y.star.C <- posterior_predict(outcome.model, newdata=data.trtC)
```

Producing draws from the posterior predictive distribution of outcomes is fairly simple using the package `rstanarm`:

```
# draw binary responses from posterior predictive distribution
# LxN* matrix of posterior predictive draws under A
y.star.A <- posterior_predict(outcome.model, newdata=data.trtA)
# LxN* matrix of posterior predictive draws under C
y.star.C <- posterior_predict(outcome.model, newdata=data.trtC)
```

For the l-th draw (the l-th row of each matrix), the A vs. C marginal treatment effect estimate is:

$$\hat{\Delta}_{AC}^{(BC,l)} = g\left(\frac{1}{N^*}\sum_{i=1}^{N^*} y_{A,i}^{*(l)}\right) - g\left(\frac{1}{N^*}\sum_{i=1}^{N^*} y_{C,i}^{*(l)}\right)$$

Producing draws from the posterior predictive distribution of outcomes is fairly simple using the package `rstanarm`:

```
# draw binary responses from posterior predictive distribution
# LxN* matrix of posterior predictive draws under A
y.star.A <- posterior_predict(outcome.model, newdata=data.trtA)
# LxN* matrix of posterior predictive draws under C
y.star.C <- posterior_predict(outcome.model, newdata=data.trtC)
```

For the l-th draw (the l-th row of each matrix), the A vs. C marginal treatment effect estimate is:

$$\hat{\Delta}_{AC}^{(BC,l)} = g\left(\frac{1}{N^*}\sum_{i=1}^{N^*} y_{A,i}^{*(l)}\right) - g\left(\frac{1}{N^*}\sum_{i=1}^{N^*} y_{C,i}^{*(l)}\right)$$

We average out the imputed outcome predictions in each draw over the rows and take the difference in the means on a suitably transformed scale:

```
# compute marginal log-odds ratio for A vs. C for each MCMC sample
# by transforming from probability to linear predictor scale
hat.delta.AC <- qlogis(rowMeans(y.star.A)) - qlogis(rowMeans(y.star.C))
```

Producing draws from the posterior predictive distribution of outcomes is fairly simple using the package `rstanarm`:

```
# draw binary responses from posterior predictive distribution
# LxN* matrix of posterior predictive draws under A
y.star.A <- posterior_predict(outcome.model, newdata=data.trtA)
# LxN* matrix of posterior predictive draws under C
y.star.C <- posterior_predict(outcome.model, newdata=data.trtC)
```

For the l-th draw (the l-th row of each matrix), the A vs. C marginal treatment effect estimate is:

$$\hat{\Delta}_{AC}^{(BC,l)} = g\left(\frac{1}{N^*}\sum_{i=1}^{N^*} y_{A,i}^{*(l)}\right) - g\left(\frac{1}{N^*}\sum_{i=1}^{N^*} y_{C,i}^{*(l)}\right)$$

We average out the imputed outcome predictions in each draw over the rows and take the difference in the means on a suitably transformed scale:

```
# compute marginal log-odds ratio for A vs. C for each MCMC sample
# by transforming from probability to linear predictor scale
hat.delta.AC <- qlogis(rowMeans(y.star.A)) - qlogis(rowMeans(y.star.C))
```

The average and variance of the marginal effect can be derived empirically from the draws, which approximate the posterior distribution:

```
hat.Delta.AC <- mean(hat.delta.AC) # average over samples
hat.var.Delta.AC <- var(hat.delta.AC) # sample variance
```

References

For the dangers of combining incompatible effect estimates in indirect treatment comparisons:

- Remiro-Azócar, A., Heath, A., and Baio, G. "Methods for Population Adjustment with Limited Access to Individual Patient Data: A Review and Simulation Study". *In Press, Research Synthesis Methods*, 2021a. Extended version available at: https://arxiv.org/abs/2004.14800

For a novel marginalization method (multiple imputation marginalization) and an outline of parametric G-computation in the context of the Cox proportional hazards regression:

- Remiro-Azócar, A., Heath, A., and Baio, G. "Marginalization of Regression-Adjusted Treatment Effects in Indirect Comparisons with Limited Patient-Level Data". Working Paper, 2021b. Available at: https://arxiv.org/abs/2008.05951

More content and code soon TBA at: https://soranalytics.substack.com/

Acknowledgements

- **Gianluca Baio (co-author)**
- **Anna Heath (co-author)**
- Anonymous Peer Reviewer #2
- Tim Morris
- David Phillippo

Acknowledgements

- **Gianluca Baio (co-author)**
- **Anna Heath (co-author)**
- Anonymous Peer Reviewer #2
- Tim Morris
- David Phillippo

Thank you!