# Adjusted results in randomized clinical trials

## G-computation

Séminaire de Biostatistique

Joe DE KEIZER

14/05/2024

CHU Poitiers
CENTRE HOSPITALIER UNIVERSITAIRE

# Summary

# I. Introduction

- Randomized clinical trials (RCTs) aims to obtain the balance of confounders between the compared arms
- The food and drug administration (FDA) recommends reporting adjusted analyses (residual confounders, and increase in power)
- Multivariate regressions provide estimates of conditional (subject-specific) treatment effects, while RCTs aim to estimate marginal (population-average) treatment effects
- Causal inference methods

Ref : *Research CfDEa. Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products. 2023.*

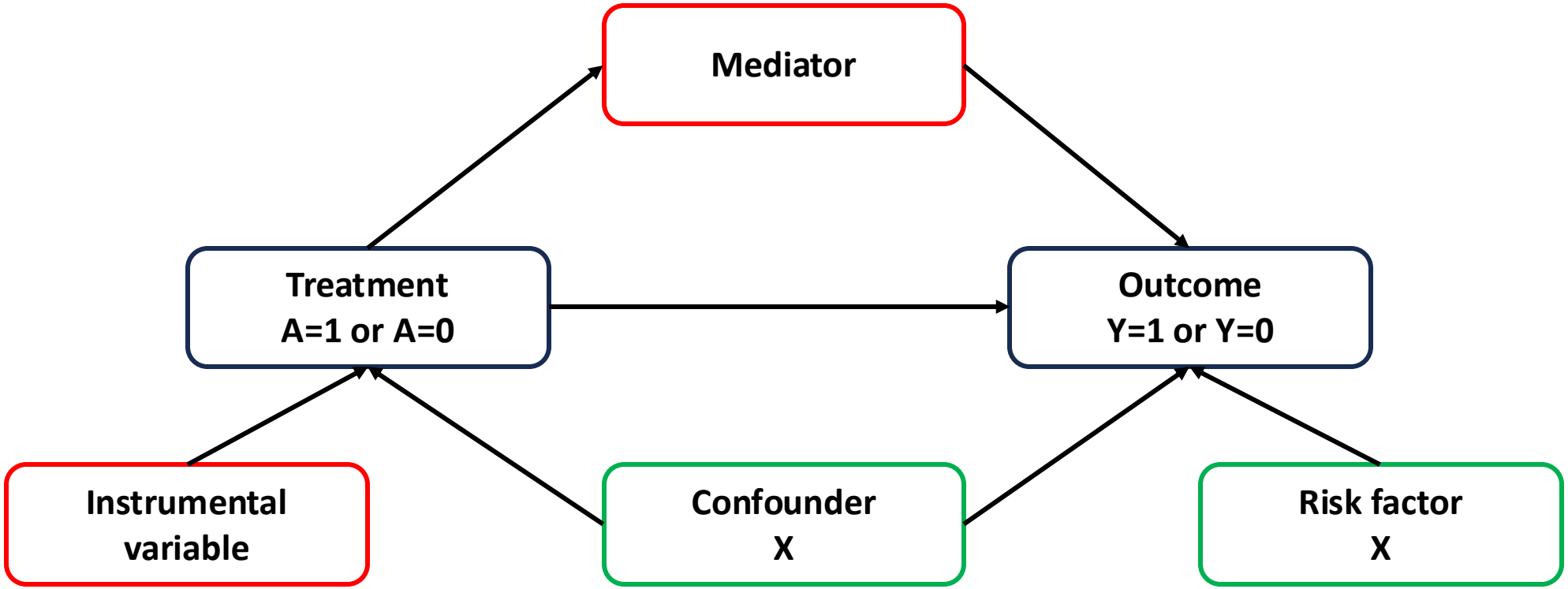
# II. Causal inference

- Causal inference aims to study the cause-and-effect between variables
- Potential confounders may influence outcomes, leading to associations not reflecting true causation
- First, one can minimize confounders by the trial design: RCT
  - Small sample size RCT reflects sample to sample fluctuation
- Secondly, statistical analyses can be applied to infer a causal relation

# II. Causal inference

Red – Do not include in adjusted analysis
Green – Include in adjusted analysis

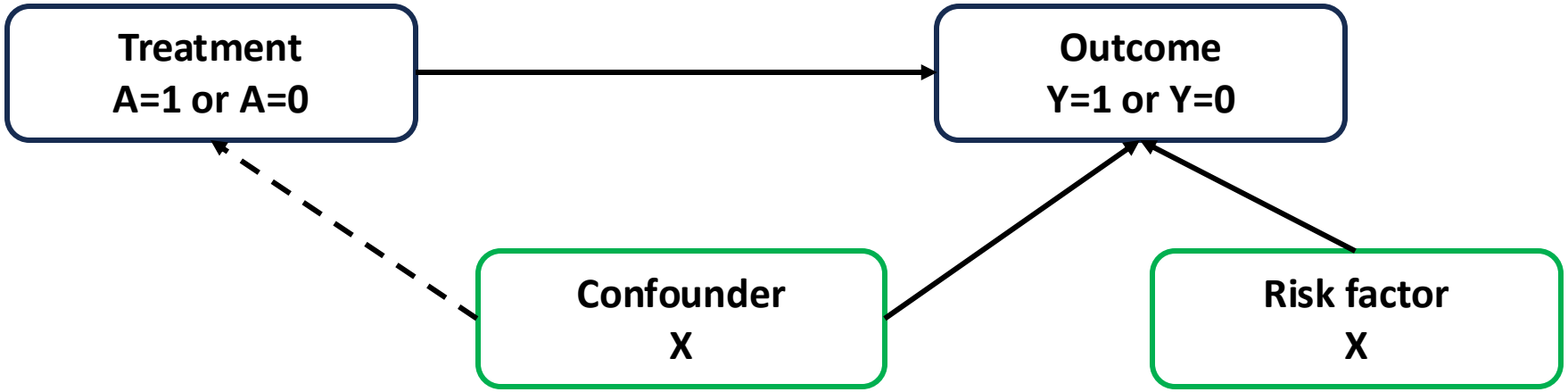
- Type of variables in real word data (RWD)




Ref : *G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets : a comparative simulation study,* Chatton A. et al.

# II. Causal inference

- Theoretical RCT scenario




Ref : *G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets : a comparative simulation study,* Chatton A. et al.

# II. Causal inference

- Causal inference infers simulating counterfactual scenarios to estimate what would have happened in the absence of the treatment or exposure.

- Causal inference statistical analyses such as Propensity Score Matching / Weighting, G-computation, double-robust methods…
  - ➢ Propensity score: Predicts treatment, creating similar counterfactual groups
  - ➢ G-computation: Simulates outcomes under different treatments using observed data
  - ➢ Double robust methods: Incorporates both model-based and data-driven approaches

- RCT theoretically positivity : positive probability of receiving each treatment

# II. Causal inference – Estimators

- Propensity score : Balances treatment groups based on the likelihood of receiving a treatment given covariates

- G-computation : Estimates treatment effect by modeling the outcome under all treatment levels

- Double robust : Combines the treatment allocation model and the outcome model to estimate the treatment effect

# III. G-computation (GC) – Pros and cons

- ✚ GC uses the same counterfactual population for each outcome

  Propensity score methods create a new counterfactual population for each outcome if the covariates in model are linked to the outcome

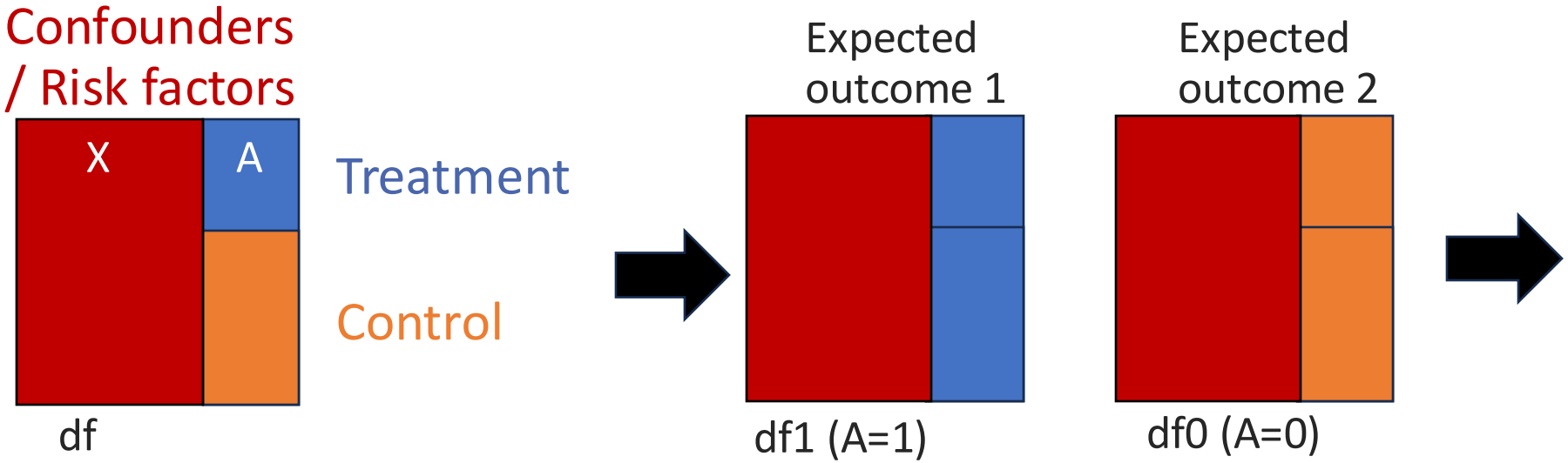
- ✚ GC based on risk factors, easier to select, avoids instrumental variables, easier to create the outcome model than the PS model

- ✚ GC had the lowest bias and variance in previous simulation studies (RWD)

- ➖ GC is not double robust (model misspecification)

Ref : *G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets : a comparative simulation study,* Chatton A. et al.

# III. G-computation (GC) – Presentation




ATE = Mean (
$\hat{E}[Y|A=1,X]$ –
$\hat{E}[Y|A=0,X]$ )

1) Outcome model / Q-model

```
model <- glm(Y ~ A + X, data = df)
```

E[Y|A,X]

2) Predicted outcomes if all observations considered on Treatment and on Control

```
predict(model,
  newdata = df1)
```

$\hat{E}[Y|A=1,X]$

```
predict(model,
  newdata = df0)
```

$\hat{E}[Y|A=0,X]$

3) Average difference between the expected outcomes 1 and 2

# III. G-computation (GC) – Application R code

```r
### Load example data
data("ToothGrowth")
# supp: Supplement type (VC or OJ)
# len: Tooth length
# dose: numeric Dose in milligrams/day

### Creating an outcome y for the example
mydata <- ToothGrowth
mydata$len <- scale(mydata$len)
mydata$supp <- ifelse(mydata$supp == "VC",1,0)
bx <- -1 + 0.2 * mydata$dose + 0.3 * mydata$len + 1.2 * mydata$supp
probs <- plogis(bx)
mydata$y <- rbinom(nrow(mydata), 1, prob = probs)
```

---

```r
### Outcome model
mod <- glm(y ~ supp + len + dose, data = mydata, family = "binomial")

### New data and expected outcomes
mydata0 <- mydata1 <- mydata
mydata0$supp <- 0
mydata1$supp <- 1
p0 <- mean(predict(mod, newdata = mydata0, type = "response")) ; print(p0)
p1 <- mean(predict(mod, newdata = mydata1, type = "response")) ; print(p1)

### ATE
print(p1 - p0)
```

# IV. Application High-Wean - Context

**Non-invasive ventilation versus high-flow nasal oxygen for postextubation respiratory failure in ICU**

In intensive care units (ICUs), patients experiencing post-extubation respiratory failure have poor outcomes. This study aims at comparing mortality between patients treated with NIV alternating with high-flow nasal oxygen or high-flow nasal oxygen alone.

**Design:**

Post-hoc analysis of a multicenter, randomized, controlled trial focusing on patients who experienced post-extubation respiratory failure within the 7 days following extubation.

**Outcomes:**

- mortality at day 28
- reintubation at 48h
- mortality at day 90

Ref : *Thille AW, Monseau G, Coudroy R, et al. Non-invasive ventilation versus high-flow nasal oxygen for postextubation respiratory failure in ICU: a post-hoc analysis of a randomized clinical trial. Critical Care (London, England). 2021;25(1):221. doi: 10.1186/s13054-021-03621-6*

# IV. Application High-Wean - Description

| | High-flow nasal oxygen (n=62) | Non-invasive ventilation (n=84) |
|---|---|---|
| *Characteristics of the patients at admission* | | |
| **Age, years** | 71 ± 9 | 70 ± 9 |
| **Male sex, n (%)** | 40 (65%) | 53 (63%) |
| **Body-mass index, kg/m2** | 29 ± 7 | 28 ± 7 |
| **SAPS II at admission, points** | 60 ± 18 | 56 ± 17 |
| **Underlying chronic cardiac disease, n (%)** | 27 (44%) | 38 (45%) |
| **Underlying chronic lung disease, n (%)** | 23 (37%) | 34 (40%) |
| **Chronic obstructive pulmonary disease, n (%)** | 18 (29%) | 25 (30%) |
| **Acute respiratory failure as reason for intubation, n (%)** | 35 (56%) | 54 (64%) |
| *Characteristics of the patients the day of extubation* | 4.5 ± 2.7 | 4.4 ± 2.8 |
| **SOFA score, points** | | |
| Duration of mechanical ventilation, median (IQR), days | 5 [3–11] | 6 [4–10] |
| **Weaning difficulty, n (%)** | | |
| -Simple weaning | 36 (58%) | 49 (58%) |
| -Difficult or prolonged weaning | 26 (42%) | 35 (42%) |

…

*Selected risk factors in bold*

# IV. Application High-Wean - Description

| | High-flow nasal oxygen (n=62) | Non-invasive ventilation (n=84) |
|---|---|---|
| *Characteristics of the patients at admission* | | |
| **Age, years** | 71 ± 9 | 70 ± 9 |
| **Male sex, n (%)** | 40 (65%) | 53 (63%) |
| **Body-mass index, kg/m2** | 29 ± 7 | 28 ± 7 |
| **SAPS II at admission, points** | 60 ± 18 | 56 ± 17 |
| **Underlying chronic cardiac disease, n (%)** | 27 (44%) | 38 (45%) |
| **Underlying chronic lung disease, n (%)** | 23 (37%) | 34 (40%) |
| **Chronic obstructive pulmonary disease, n (%)** | 18 (29%) | 25 (30%) |
| **Acute respiratory failure as reason for intubation, n (%)** | 35 (56%) | 54 (64%) |
| *Characteristics of the patients the day of extubation* | 4.5 ± 2.7 | 4.4 ± 2.8 |
| **SOFA score, points** | | |
| Duration of mechanical ventilation, median (IQR), days | 5 [3–11] | 6 [4–10] |
| **Weaning difficulty, n (%)** | | |
| -Simple weaning | 36 (58%) | 49 (58%) |
| -Difficult or prolonged weaning | 26 (42%) | 35 (42%) |

Predicted mortality 68.1%

Predicted mortality 59.8%

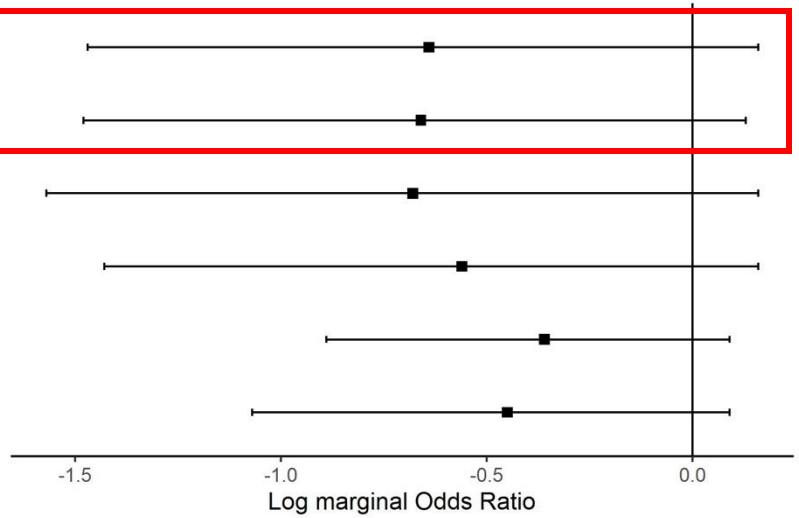
Differences due to randomness

…

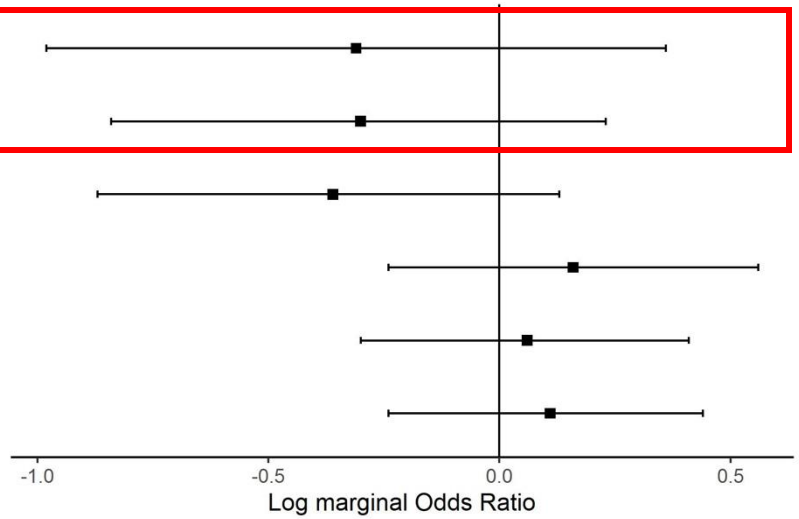
*Selected risk factors in bold*

# IV. Application High-Wean - Results

| Outcome | Method | log mOR (95%CI) |
|---|---|---|
| Mortality at day 28 | - Unadjusted | -0.64 [-1.47;0.16] |
| | - Elasticnet | -0.66 [-1.48;0.13] |
| | - Lasso | -0.68 [-1.57;0.16] |
| | - Neural network | -0.56 [-1.43;0.16] |
| | - Support vector machine | -0.36 [-0.89;0.09] |
| | - Super learner | -0.45 [-1.07;0.09] |




| Outcome | Method | log mOR (95%CI) |
|---|---|---|
| Reintubation at 48 h | - Unadjusted | -0.31 [-0.98;0.36] |
| | - Elasticnet | -0.3 [-0.84;0.23] |
| | - Lasso | -0.36 [-0.87;0.13] |
| | - Neural network | 0.16 [-0.24;0.56] |
| | - Support vector machine | 0.06 [-0.3;0.41] |
| | - Super learner | 0.11 [-0.24;0.44] |



# IV. Application High-Wean

- Sample size calculation parameters generally used :

Significance Level (α)

Power (1 - β)

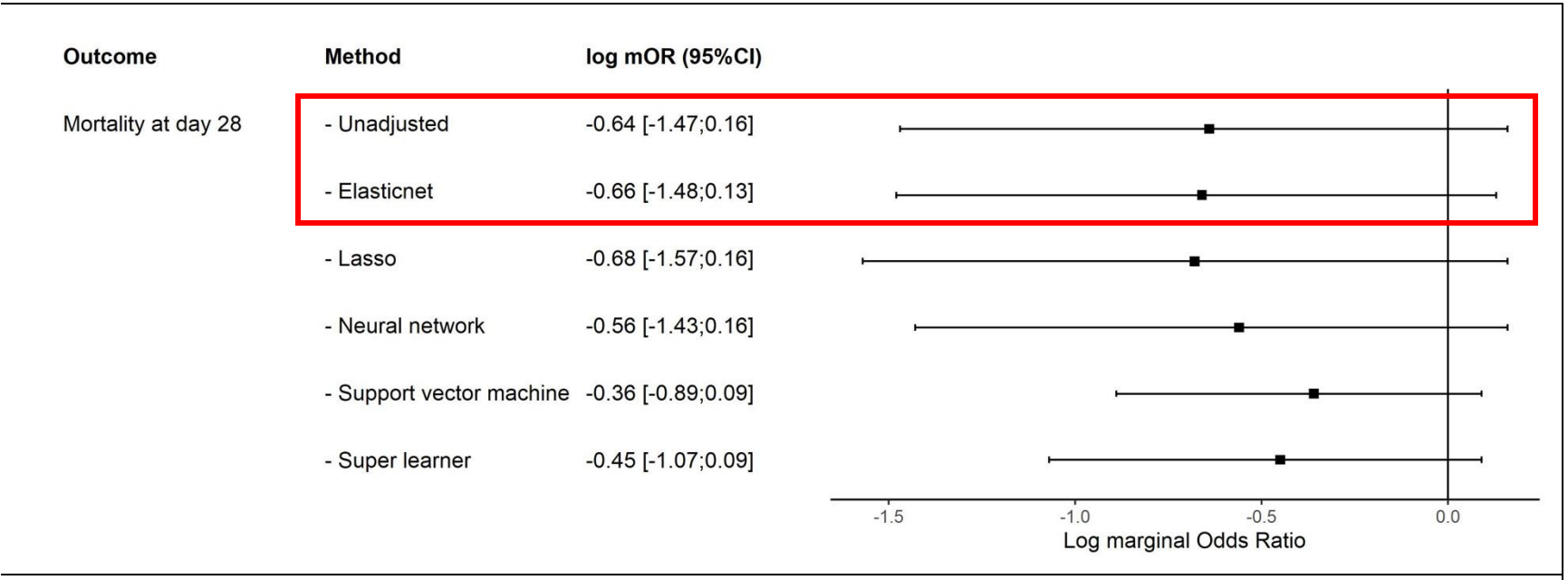
Standard Deviation (σ)

Effect Size (Δ)

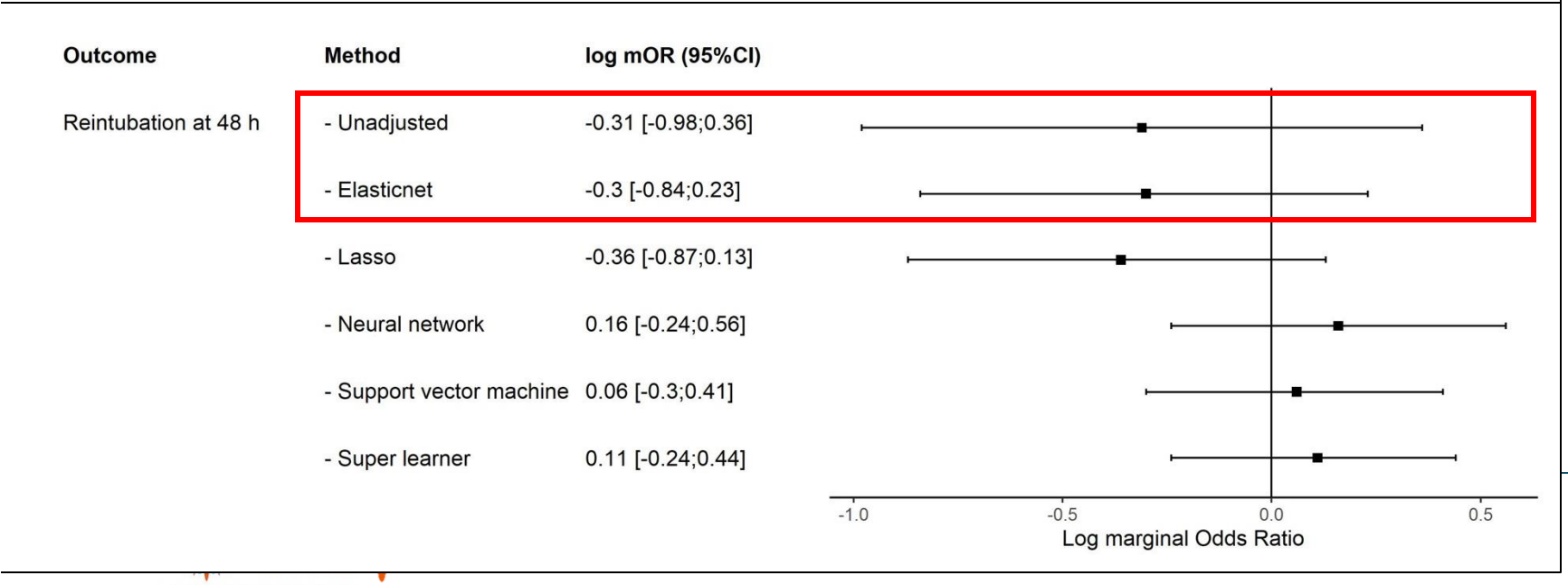
Sample size (n)

Linked between each other

# IV. Application High-Wean – AUC




- Mortality at day 28
  Elasticnet AUC of 0.60




- Reintubation at 48h
  Elasticnet AUC of 0.90

*Importance of risk factors and constructing the model correctly*

# V. Conclusion

- Design:
  - RCT
  - Anticipate the random variability
  - Include enough patients
- Statistical methods:
  - G-computation
  - Construct the outcome model correctly
- Protocol / Statistical analysis plan
  - Detail the planned analyses
  - Collect the right data for the multivariate analysis

# V. Conclusion – Protocole example

Ignorer les facteurs de stratification dans l'analyse conduit à des intervalles de confiance incorrects (1). Plus généralement, les analyses ajustées sur les facteurs de risque peuvent augmenter la puissance et donner des taux d'erreur de type I corrects (2). Nous utiliserons la G-computation pour estimer l'effet moyen du traitement (ATE).

Pour chaque critère, les analyses suivantes seront réalisées.

Premièrement, nous estimerons le modèle multivarié prédictif de critère de jugement (modèle de travail). Parmi les prédicteurs potentiels, nous prendrons en compte le facteur de stratification et les facteurs influençant potentiellement le pronostic des patients et listés dans l'analyse descriptive. Aucune interaction entre les bras de traitement et les prédicteurs ne sera testée. La régularisation élastique nette sera utilisée (3) : une régression linéaire pour la force des membres inférieurs ou le niveau d'activité physique à 6 mois, une régression logistique pour l'obésité sarcopénique à 6 mois, et des régressions linéaires mixtes pour les autres critères de jugements répétés. Les paramètres de pénalité seront estimés en minimisant les erreurs quadratiques moyennes en validation croisée 20 fois.

Deuxièmement, sur la base du modèle de travail, nous prédirons le critère de jugement si tous les participants avaient été assignés dans le bras expérimental et ensuite si tous les participants ont été assignés dans le bras contrôle. Pour chaque résultat, l'ATE sera la différence entre les deux moyennes des prédictions individuelles.

*1. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. BMJ*
*2. Morris TP, Walker AS, Williamson EJ, White IR. Planning a method for covariate adjustment in individually randomised trials: a practical guide. Trials*
*3. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. J R Stat Soc Ser B Stat Methodol*