# WHO CAN BENEFIT FROM OVARIAN CANCER SCREENING?

## A counterfactual prediction modelling / CATE modelling / HTE study of an ovarian cancer screening trial

Keling Wang
02 April 2025

Causal Inference Group Meeting
Dept. Epidemiology, Erasmus MC

Erasmus MC
Universitair Medisch Centrum Rotterdam

# CONTENTS

1. Basic idea, aims, and data revisited
2. Target trial and deviations from the original trial
3. Statistical analysis: g-formula + random forest + re-fitting
4. Questions and discussion

# BACKGROUND, AIMS AND DATA

# BACKGROUND

- Ovarian cancer is hard to be early detected.
- The previous screening procedure consists of transvaginal ultrasound (TVU) and CA-125 screening.
- Two large-scale trials reported disappointing results.
  - PLCO (70k randomized, closure at 2009; using single CA125 cutoff at 28 U/ml)
  - UKCTOCS (200k randomized, closure at 2020; using an algorithm of CA125)
- Following them, OC screening is not recommended anymore in the meantime.

# BACKGROUND

- However, some subgroups of participants could still benefit from the screening strategy, *e.g.* those with very high baseline risks.
- We aim to re-analyze the PLCO trial with 15-year all-cause mortality as primary outcome to:
    1. select a set of assumed effect measure modifiers (EMM) and estimate the **conditional average treatment effect** (CATE) conditioning on them;
    2. **predict the screening effects** for women in the trial based on selected predictors;
    3. find **characteristic (EMM) combinations** with which one can receive screening benefits.

# PLCO TRIAL DATA

- Population:
  1. ~ 70k women randomized;
  2. ~ 68k women eligible for analysis after applying our TTE criteria (discussed later).
- Treatment regime: deterministic dynamic
  1. Treated group: maximal 4 years of TVU (0,1,2,3) and 6 years of CA-125 (0,1,2,3,4,5) until (1) a diagnosis confirmed; (2) lost to follow-up; (3) death; (4) all screening visits completed.
  2. Control group: follow-up only until (1) lost to follow-up; (2) death.
- Outcome:
  1. All-cause mortality: obtained from trial records and from death registry after trial closure *(primary)*
  2. OC-specific mortality: classified according to death cert. codes and a trial committee.

# PLCO TRIAL DATA: SUMMARY

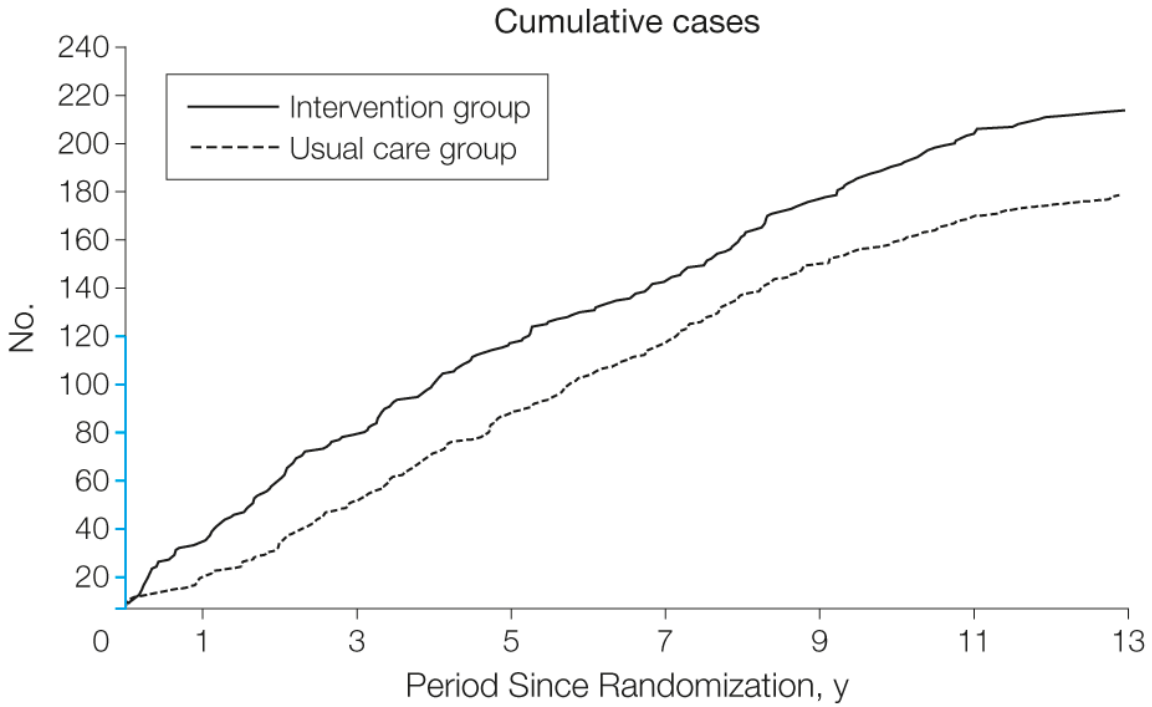
- Power of the trial is limited, especially for OC-specific mortality.

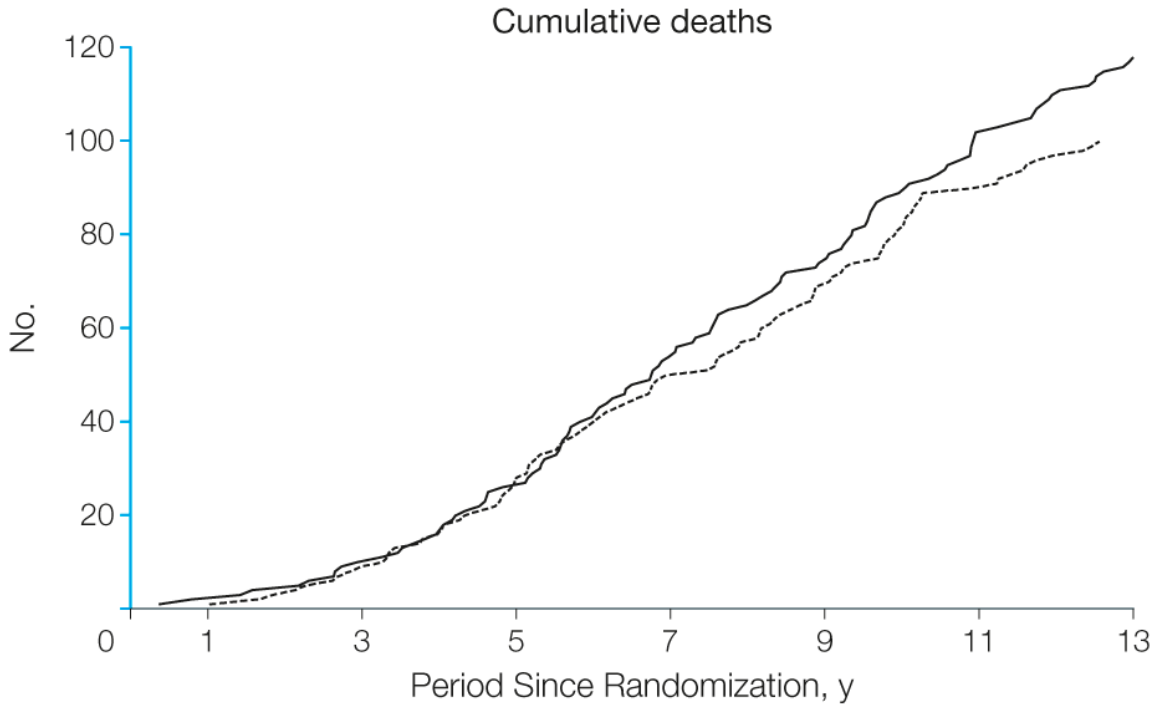
| Outcome (censored at 15th year) | # Event treated | # Event control | Crude 15-yr cumulative *RR* (ITT) | Crude *IRR* (ITT) |
|---|---|---|---|---|
| Ovarian cancer-specific mortality | 184 / 34238 *(0.54%)* | 177 / 34285 *(0.52%)* | 1.04 *(0.85 to 1.28)* | 1.04 (475,184 person-years) *(0.85 to 1.29)* |
| All-cause mortality | 5630 / 34238 *(16.4%)* | 5637 / 34285 *(16.4%)* | 1.00 *(0.97 to 1.03)* | 1.00 (477,484 person-years) *(0.97 to 1.04)* |

# PLCO TRIAL DATA: SUMMARY

- Censored observations and competing events are not balanced, especially when accounting for non-compliance in per-protocol analysis.

| | Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| All-cause death | Treated | 137 | 156 | 209 | 238 | 243 | 298 | 343 | 375 | 405 | 413 | 485 | 510 | 578 | 593 | 647 | 5630 |
| | Control | 133 | 174 | 204 | 235 | 263 | 287 | 333 | 355 | 413 | 468 | 489 | 558 | 508 | 619 | 598 | 5637 |
| OC death | Treated | 2 | 3 | 7 | 9 | 12 | 16 | 12 | 15 | 11 | 20 | 16 | 15 | 15 | 14 | 17 | 184 |
| | Control | 1 | 4 | 10 | 7 | 13 | 14 | 8 | 14 | 14 | 18 | 8 | 19 | 19 | 11 | 17 | 177 |
| Censoring | Treated | 2 | 2 | 3 | 1 | 0 | 0 | 4 | 3 | 7 | 14 | 230 | 552 | 636 | 632 | * | 2086 |
| | Control | 1 | 0 | 2 | 0 | 0 | 3 | 1 | 0 | 2 | 32 | 304 | 849 | 876 | 874 | * | 2944 |
| Competing event* | Treated | 135 | 153 | 202 | 229 | 231 | 282 | 331 | 360 | 394 | 393 | 469 | 495 | 563 | 579 | 630 | 5446 |
| | Control | 132 | 170 | 194 | 228 | 250 | 273 | 325 | 341 | 399 | 450 | 481 | 539 | 489 | 608 | 581 | 5460 |
| Censoring PP data | Treated | 1111 | 548 | 417 | 98 | 77 | 2 | 3 | 2 | 5 | 13 | 214 | 515 | 593 | 596 | * | 4194 |
| | Control | 1 | 0 | 2 | 0 | 0 | 3 | 1 | 0 | 2 | 32 | 304 | 849 | 876 | 874 | * | 2944 |




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Intervention group | | | | | | | |
| Cumulative cancers | 28 | 74 | 113 | 139 | 174 | 202 | 212 |
| Cumulative person-years | 33 908 | 100 777 | 166 273 | 230 393 | 292 223 | 341 975 | 371 833 |
| Usual care group | | | | | | | |
| Cumulative cancers | 13 | 45 | 83 | 113 | 146 | 167 | 176 |
| Cumulative person-years | 33 994 | 101 279 | 167 380 | 232 046 | 294 424 | 344 734 | 374 976 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Intervention group | | | | | | | |
| Cumulative cancers | 2 | 10 | 26 | 54 | 74 | 102 | 118 |
| Cumulative person-years | 34 210 | 102 191 | 169 354 | 235 475 | 299 372 | 350 870 | 381 574 |
| Usual care group | | | | | | | |
| Cumulative cancers | 0 | 9 | 28 | 50 | 69 | 90 | 100 |
| Cumulative person-years | 34 260 | 102 344 | 169 617 | 235 836 | 299 903 | 351 557 | 382 502 |

Erasmus MC
Universitair Medisch Centrum Rotterdam

*JAMA.* 2011;305(22):2295-2303. doi:10.1001/jama.2011.766

# WHAT'S DIFFICULT?

- Dynamic treatment regime (PP/AT)
- Inconsistent trial eligibility criteria and treatment plans.
- Imbalanced losses to follow-up and massive competing events.
- Too many (~40) potential EMMs with limited power.

# WORKFLOW

1. Target trial emulation to “correct” the original dataset
2. Data cleaning and covariate imputation
3. g-formula + random forest to model counterfactuals under treated and under control, and calculate CATEs.
4. Re-fitting a simple and explainable model for CATEs.

# 1. TARGET TRIAL

(Only the most important deviations are shown here)

# TTE COMPONENTS AND DEVIATIONS

## 1. Causal contrast / estimand

| Estimand type | Causal contrasts | Outcome and horizon |
|---|---|---|
| ITT effect: effect of treatment assignment on… | $\psi = \Pr[Y^{z=1}=0\|t,\boldsymbol{X}] - \Pr[Y^{z=0}=0\|t,\boldsymbol{X}]$ (survival difference)<br>$\psi = \Pr[Y^{z=1}=0\|t,\boldsymbol{X}] \,/\, \Pr[Y^{z=0}=0\|t,\boldsymbol{X}]$ (survival ratio) | Primary outcome: All-cause mortality / overall survival<br><br>Secondary outcome: OC-specific mortality / OC-specific survival<br><br>Horizon: $t = 6, 10, 15$ yr |
| Effect among compliers: effect of treatment received among those complied with assignment on… | $\psi = \Pr[Y^{g}=0\|t,\boldsymbol{X},Z=1] - \Pr[Y^{g'}=0\|t,\boldsymbol{X},Z=0]$<br>$\psi = \Pr[Y^{g}=0\|t,\boldsymbol{X},Z=1] \,/\, \Pr[Y^{g'}=0\|t,\boldsymbol{X},Z=0]$ | |
| Treatment effect: effect of treatment received among all eligible women | $\psi = \Pr[Y^{g}=0\|t,\boldsymbol{X}] - \Pr[Y^{g'}=0\|t,\boldsymbol{X}]$<br>$\psi = \Pr[Y^{g}=0\|t,\boldsymbol{X}] \,/\, \Pr[Y^{g'}=0\|t,\boldsymbol{X}]$ | |

# TTE COMPONENTS AND DEVIATIONS

## 2. Eligibility Criteria

| Target trial specification | Reference trial (Original trial) specification |
|---|---|
| (1) Adult women, as registered (in term of sex), aged 55-74 at the start of follow-up.<br>(2) **Have at least one side of ovaries or fallopian tubes NOT removed, and are at risk of dying from ovarian cancer.**<br>Other same as (2), (5), and (6) in original trial.<br><br>**Deviation from the reference trial:**<br>All women with two ovaries removed are considered ineligible regardless of the trial inception calendar date;<br>All women with current tamoxifen use are considered eligible regardless of the trial inception calendar date. | (1) Adult women aged 55-74 at the start of follow-up.<br>(2) No history of lung, colorectal, or ovarian cancer.<br>(3) *** Between 1993-1996: No previous oophorectomy; After 1996: This criterion dropped.<br>(4) *** Between 1993-1999: No tamoxifen use at trial inception; After 1999: this criterion dropped.<br>(5) No current treatment for cancer.<br>(6) No previous surgical removal of the entire prostate, one lung, or the entire colon. |

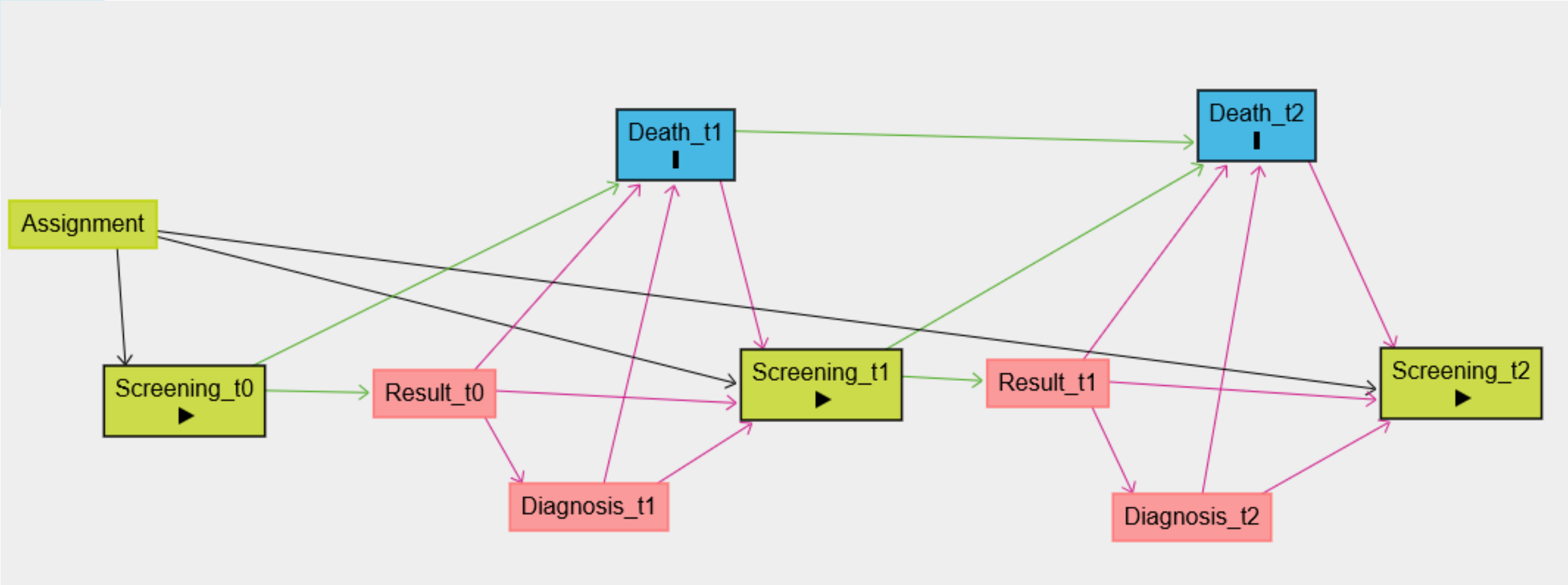
Erasmus MC
Universitair Medisch Centrum Rotterdam

# TTE COMPONENTS AND DEVIATIONS

## 3. Treatment strategy: dynamic treatment regime

- *Intervention: one follows the assigned screening plan* ***until*** (1) *a positive screening result is detected* or (2) definitive diagnosis or (3) death.
- *Control:* follow-ups until death.
- (Under PP and AT) we evaluate the joint effect and assume no synergy.

| | Year | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7+ |
|---|---|---|---|---|---|---|---|---|---|---|
| Treated | TVU | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| | CA125 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| Control | TVU | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | CA125 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |




***Note:*** **Confounding, censoring and competing risk not shown above.**

# 2. STATISTICAL ANALYSIS

***Establishing a random forest plugged g-formula (MSM) model***

Variable selection through importance

Re-fitting an effect ~ EMM model to explain the heterogeneity

Bootstrapping to get confidence intervals

# G-FORMULA TO ESTIMATE CATE

- G-formula allows to flexibly estimate **ATE**s while accounting for competing events and censoring under dynamic treatment regimes and time-varying confounding.
- Is there a way to use g-formula for CATE estimation (this way we will solve most difficulties we have here)?
- -- Of course!

# G-FORMULA TO ESTIMATE CATE

- What g-formula (an MSM) offers for an ATE / marginal mean of counterfactuals:

$$\Pr\left[Y^{g}_{k+1}=1\right] = \sum_{\bar{l}}\sum_{i=0}^{k}\Pr\left[Y_{i+1}=1 \mid \cdots\right]\prod_{j=0}^{i}\left\{\Pr\left[Y_{j}=0 \mid \cdots\right]f(l_j \mid \bar{l}_{j-1}, \bar{a}_{j-1}, \cdots)\right\}$$

- What g-formula implements to estimate this ATE:

**3. Predict (simulate) individual histories using this prob.dist.**

**4. Assign treatment received according to regime**

**5. Predict (simulate) Y's**

$$\Pr\left[Y^{g}_{k+1}=1\right] = \sum_{\bar{l}}\sum_{i=0}^{k}\Pr\left[Y_{i+1}=1 \mid \cdots\right]\prod_{j=0}^{i}\left\{\Pr\left[Y_{j}=0 \mid \cdots\right]f(l_j \mid \bar{l}_{j-1}, \bar{a}_{j-1}, \cdots)\right\}$$

**6. Sum over the big soup and get a sample mean**

**2. Model Y's conditioning on histories**

(calculated from last time point Y)

**1. Make covariate prob.dens. histories according to model and empirical dist. of baseline ($j = 0$)**

# G-FORMULA TO ESTIMATE CATE

- What we want g-formula to offer: CATE / conditional mean of counterfactuals, conditioning on baseline (time-fixed) EMMs:

$$\Pr\left[Y^g_{k+1} = 1 \mid X\right] = \sum_{\bar{l}} \sum_{i=0}^{k} \Pr\left[Y_{i+1} = 1 \mid \cdot, X\right] \prod_{j=0}^{i} \left\{ \Pr\left[Y_j = 0 \mid \cdot, X\right] f(l_j \mid \bar{l}_{j-1}, \bar{a}_{j-1}, \cdot, X) f(X) \right\}$$

- There exists a valid modification to g-formula for CATE estimation with time-varying EMMs [1], but we need to model this:

$$f\left(l_j \mid \bar{l}_{j-1}, \bar{a}^g_{j-1}, \bar{x}_{j-1}\right) f\left(x_j \mid \bar{x}_{j-1}, \bar{l}_j, \bar{a}^g_{j-1}\right) f\left(a^g_j \mid \bar{a}^g_{j-1}, \bar{l}_j, \bar{x}_j\right)$$

The practice is (in principle) to model all $x$, $l$, $a$ as variables as time-varying components for which histories needs to be made.

# OUR APPROACH AS A SPECIAL CASE

- While the original approach has several limitations (and implementation issues for us)*…
- we propose to consider $\langle \bar{L}, \bar{A} \rangle \perp \boldsymbol{X}$ . Since we are in an RCT the independence should be valid. This reduces the things to:

$$\Pr\left[Y^{g}_{k+1} = 1 \mid X\right] = \sum_{\bar{l}} \sum_{i=0}^{k} \Pr\left[Y_{i+1} = 1 \mid \cdot, X\right] \prod_{j=0}^{i} \left\{ \Pr\left[Y_j = 0 \mid \cdot, X\right] f(l_j \mid \bar{l}_{j-1}, \bar{a}_{j-1}, \cdot, \right.$$

Not modelled

$$\Pr\left[Y_{i+1} = 1 | \bar{L}_i = l_i,\ Y_i = C_i = 0,\ \bar{A}_i = \bar{a}^{g}_i,\ \mathbf{X} = \mathbf{x_0}\right]$$

$$= \Pr\left[Y_{i+1} = 1 | \bar{L}_i = l_i,\ Y_i = C_i = 0\right] \Pr\left[Y_{i+1} = 1 | \bar{A}_i = \bar{a}^{g}_i,\ \mathbf{X} = \mathbf{x_0}\right]$$

$$Y_{i+1} \sim A_i + A_i \mathbf{X_0} + \mathbf{X_0}$$

Erasmus MC
Universitair Medisch Centrum Rotterdam

# PLUGGING RANDOM FOREST

- Therefore, inserting ("plug") a parametric $Y$ model conditioning on $\langle \boldsymbol{X_0}, A \rangle$ would allow us to estimate $\Pr[Y_k^g = 1|\boldsymbol{X}]$: $Y_{i+1} \sim A_i + A_i\mathbf{X_0} + \mathbf{X_0}$

- For flexibility and variable selection, we use a (probability) random forest to model the outcome, including all treatment variables, all potential effect measure modifiers, and their interaction terms.
- What's nice about random forest: (a lot)
- When predicting Y (step 5), the node splitting for Y(0) and for Y(1) should be separated. This gives our model more flexibility.

# RF-PLUGGED G-FORMULA

- The final model takes these specifications:
- - **Y** model (random forest)
- - **C** model (glm, with prior knowledge)
- - **D** model (glm, with prior knowledge)
- - **A** and **L** model (lagged glm)

# MODEL SPECIFICATION DETAILS

- **Y model:** (random forest) treatment * EMMs + year + year-of-rand
- **C model:** Censoring ~ [basic demographic and smoking/alcohol] + [menstruation and pregnancy history] + [current and lagged screening result]
- **D model:** Compevent ~ [basic] + [lagged result]  + [medical history] + [family history]
- **Time varying** covariate and treatment model:
- - Treatment ~ lagged result + lagged treatment + screening history
- - Result ~ lagged result + lagged treatment + screening history

Establishing a random forest plugged g-formula (MSM) model

↓

***Variable selection through importance***

↓

Re-fitting an effect ~ EMM model to explain the heterogeneity?

↓

Bootstrapping to get confidence intervals

# VARIABLE SELECTION THROUGH RF

- We start with manually selected ~40 potential EMMs
- Corrected impurity importance is used for variable selection (*how much "impurity" could be reduced by splitting on this variable, corrected for variable information amount (levels and type))*
- Variable selection goal: not for precision, but for interpretability[1]* (the goal is CATEs but **not** *individual prediction*, so we can pick whatever variable we want to condition on.)
- Variable selection strategy: Recursively picked the most important variables and calculate the importance within the new model. [1]

[1] Genuer R et al. Variable selection using random forests. 10.1016/j.patrec.2010.03.014

Fit ITT RF's for both outcomes on all variables and selected the most important 15 variables for each outcome

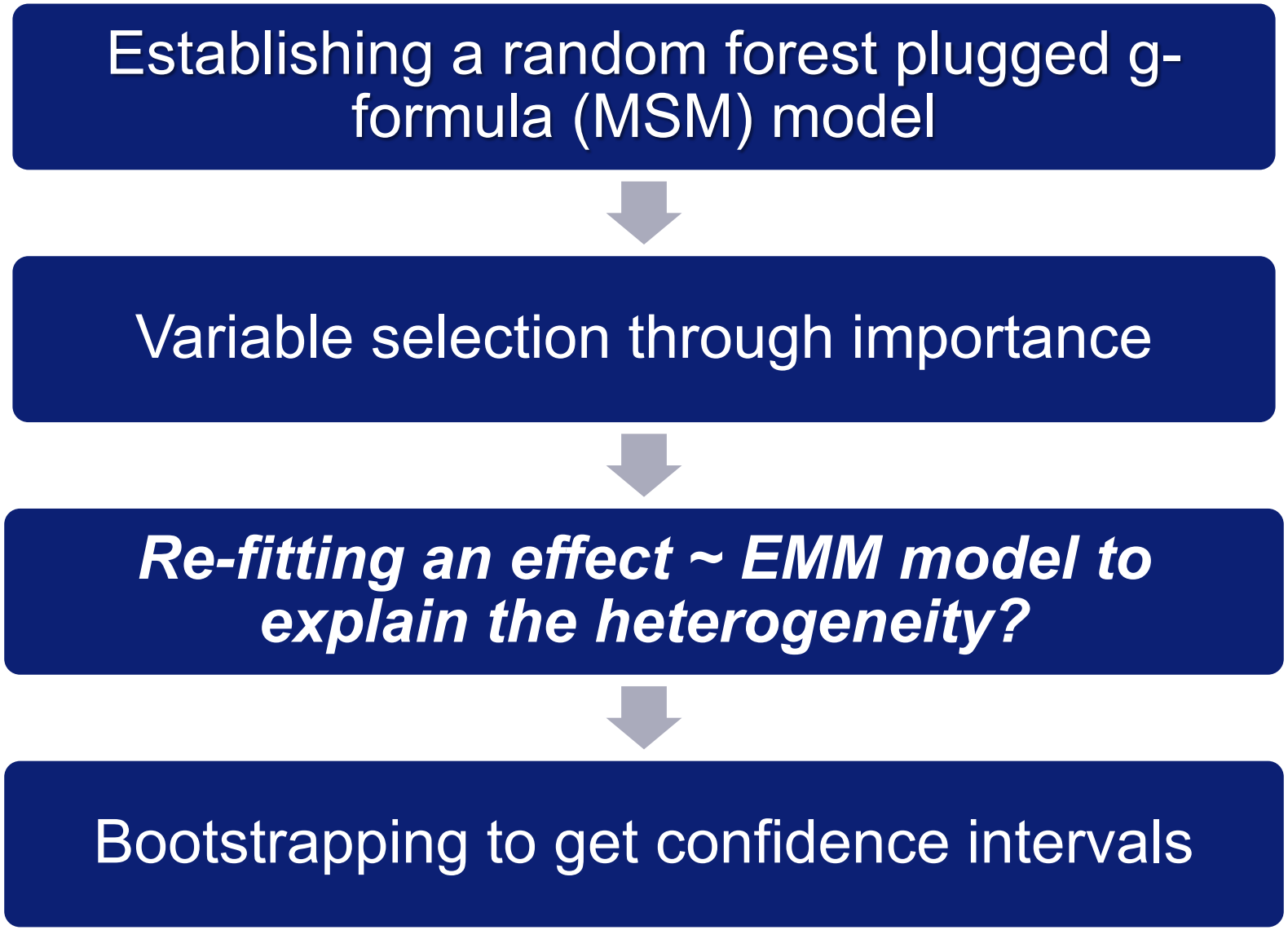
The importance is ranked based on both the importance of $\mathbf{1}(Z = 0)$: $\boldsymbol{X}$ and $\mathbf{1}(Z = 1)$: $\boldsymbol{X}$

Fit ITT RF's on these 15 variables and selected the most important 7-8 variables for each outcome.

Exclude variables with similar clinical significance (maybe collinearity)

Selected the most important and clinical interpretable 15 variables

The final RF model used in g-formula is based on the 15 variables



# RE-FITTING AN INTERPRETABLE MODEL

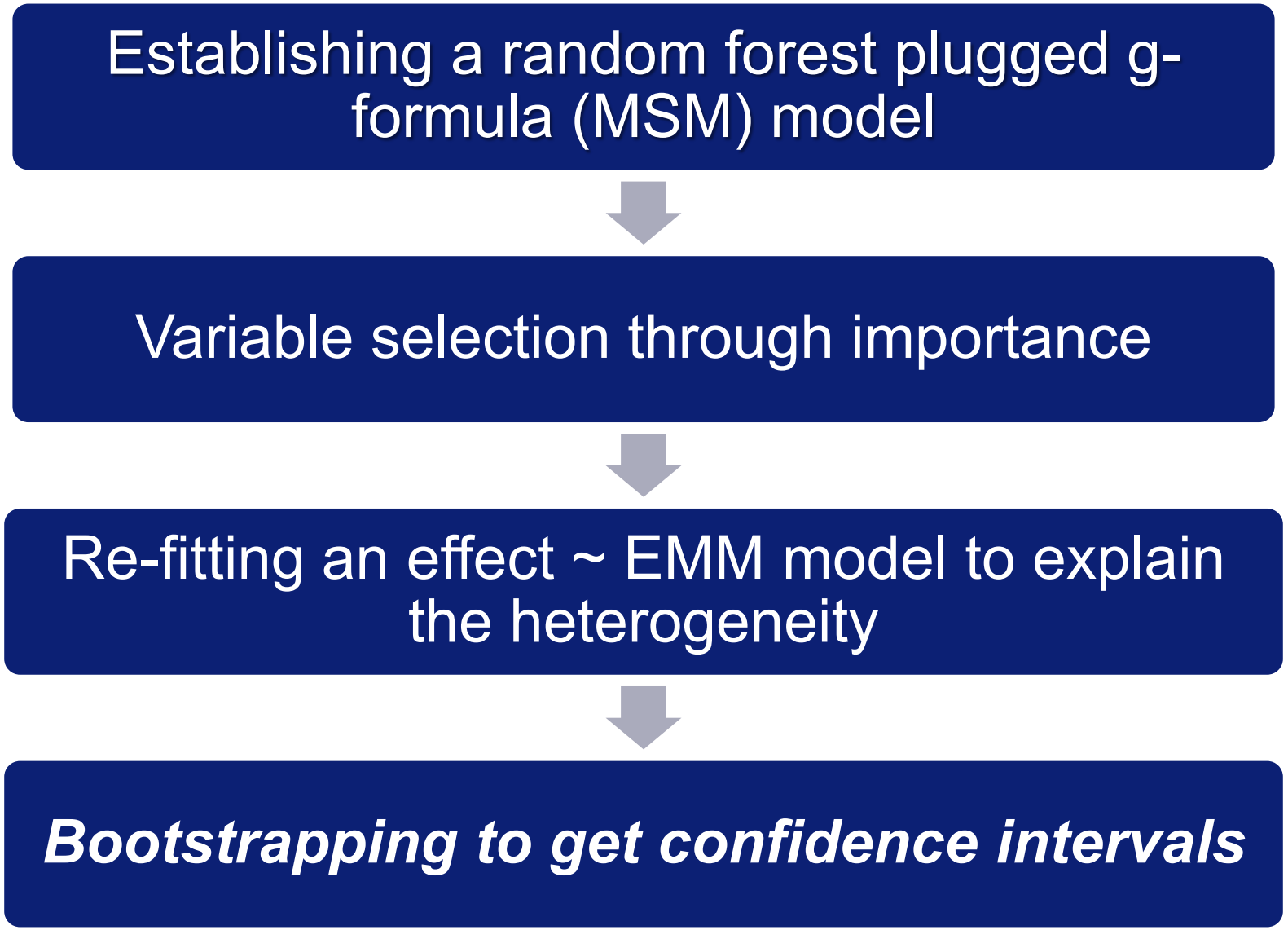
- G-formula + RF outputs predicted conditional counterfactual outcomes (cumulative risk / survival) under treatment, control and natural course.
- We can calculate conditional effects as desired and predict the effect based on EMMs.
- However, which ones of the 15 variables really work? – It's hard to tell since we use RF.
- How to make the use of our outcomes for clinical purpose?
- We need an interpretable and simple model that summarize the effect measure modifications.

# RE-FITTING AN INTERPRETABLE MODEL

- We re-fit a simple (G)LM:

$$\mathbb{E}\left[\Pr\left[Y^{z=1}=0\right]-\Pr\left[Y^{z=0}=0\right]\middle|\mathbf{X}\right] = \mathbf{X}\beta$$

$$\mathbb{E}\left[\Pr\left[Y^{z=1}=0\right]/\Pr\left[Y^{z=0}=0\right]\middle|\mathbf{X}\right] = \exp\left(\mathbf{X}\beta\right)$$



# BOOTSTRAPPING

- We used 100 bootstrapped samples and calculated the 95% CI's using normal method for:
- - ATE, ATT, ATC;
- - magnitude of effect measure modification;
- - CATE predictions;
- - refitted model parameters.

# 3. SOME RESULTS

# TABLE 1

*Participant characteristics among those eligible*

| | Total (n = 68,523) | Assigned to intervention (n = 34,238) | Assigned to control (n = 34,285) |
|---|---|---|---|
| Age | 62.5 (±5.4) | 62.5 (±5.4) | 62.5 (±5.4) |
| **Race** | | | |
| White | 58,385 (85.2%) | 29,276 (85.5%) | 29,109 (84.9%) |
| Black | 4,229 (6.2%) | 2,094 (6.1%) | 2,135 (6.2%) |
| Hispanic | 2,055 (3.0%) | 1,009 (2.9%) | 1,046 (3.1%) |
| Asian and other | 3,854 (5.6%) | 1,859 (5.4%) | 1,995 (5.8%) |
| **Occupation** | | | |
| working | 40,338 (58.9%) | 20,235 (59.1%) | 20,103 (58.6%) |
| not working | 2,694 (3.9%) | 1,283 (3.7%) | 1,411 (4.1%) |
| retired | 23,982 (35.0%) | 11,993 (35.0%) | 11,989 (35.0%) |
| other | 1,509 (2.2%) | 727 (2.1%) | 782 (2.3%) |
| **Education** | | | |
| lower | 4,291 (6.3%) | 2,160 (6.3%) | 2,131 (6.2%) |
| high school | 26,538 (38.7%) | 13,222 (38.6%) | 13,316 (38.8%) |
| some college | 15,059 (22.0%) | 7,634 (22.3%) | 7,425 (21.7%) |
| college or higher | 22,635 (33.0%) | 11,222 (32.8%) | 11,413 (33.3%) |

# TABLE 1 (CONT'D)

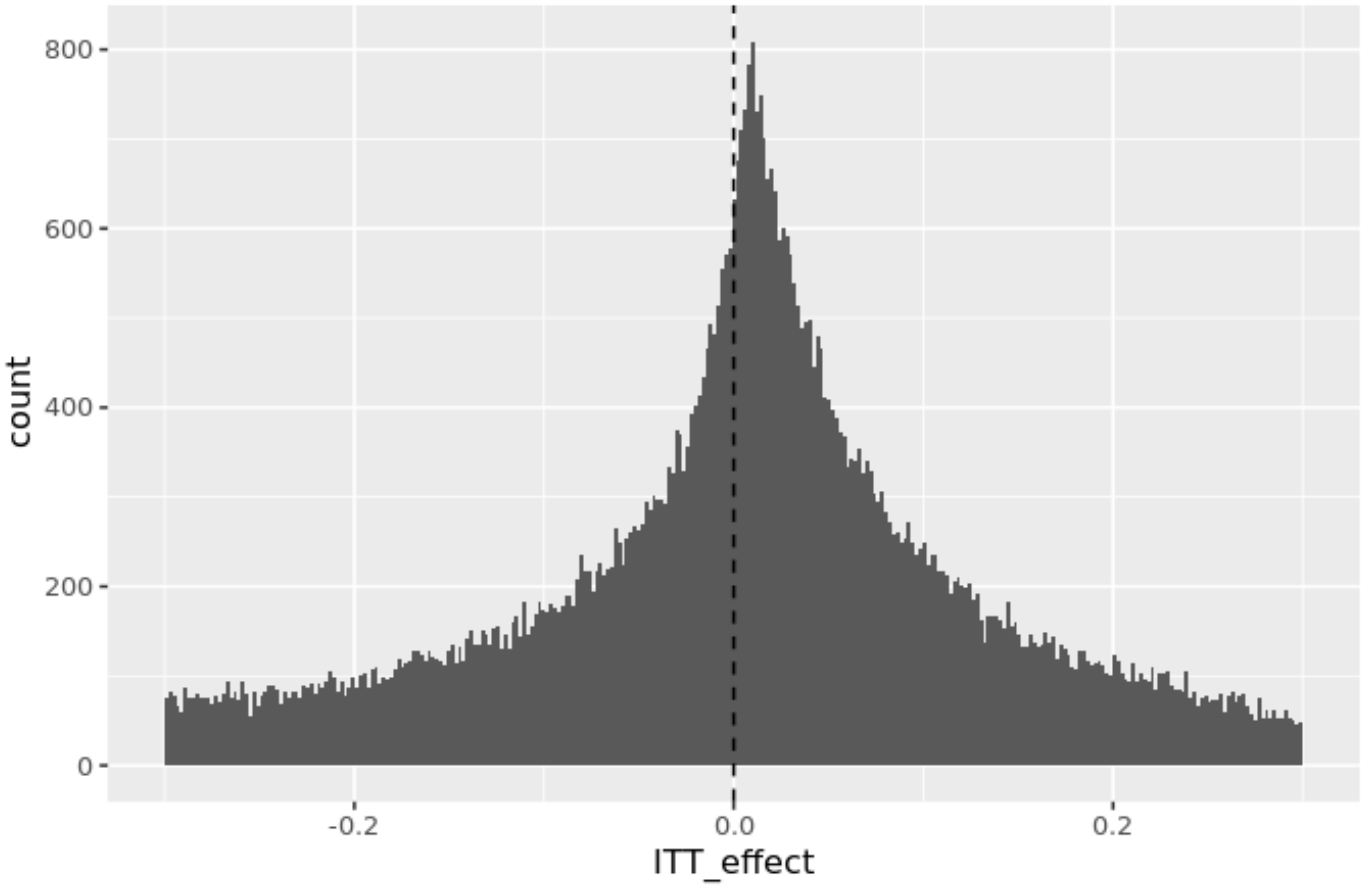
| | Total (n = 68,523) | Assigned to intervention (n = 34,238) | Assigned to control (n = 34,285) |
|---|---|---|---|
| History of diabetes | 6,094 (8.9%) | 2,930 (8.6%) | 3,164 (9.2%) |
| History of colorectal diseases | 6,136 (9.0%) | 3,023 (8.8%) | 3,113 (9.1%) |
| **History of any cancers** | | | |
| No | 62,159 (90.7%) | 31,118 (90.9%) | 31,041 (90.5%) |
| Yes | 4,273 (6.2%) | 2,178 (6.4%) | 2,095 (6.1%) |
| Unclear | 2,091 (3.1%) | 942 (2.8%) | 1,149 (3.4%) |
| History of benign ovarian tumor/cyst | 11,058 (16.1%) | 5,209 (15.2%) | 5,849 (17.1%) |
| **Family history of ovarian-related cancers** | | | |
| No | 64,068 (93.5%) | 32,031 (93.6%) | 32,037 (93.4%) |
| Yes | 3,328 (4.9%) | 1,642 (4.8%) | 1,686 (4.9%) |
| Possibly | 1,127 (1.6%) | 565 (1.7%) | 562 (1.6%) |
| Smoking pack-years | 15.4 (±25.8) | 15.1 (±25.6) | 15.6 (±26.0) |
| Alcohol use | 44,442 (64.9%) | 22,214 (64.9%) | 22,228 (64.8%) |
| Daily total energy intake | 1442.8 (±600.8) | 1443.2 (±599.4) | 1442.5 (±602.2) |

# TABLE 1 (CONT'D)

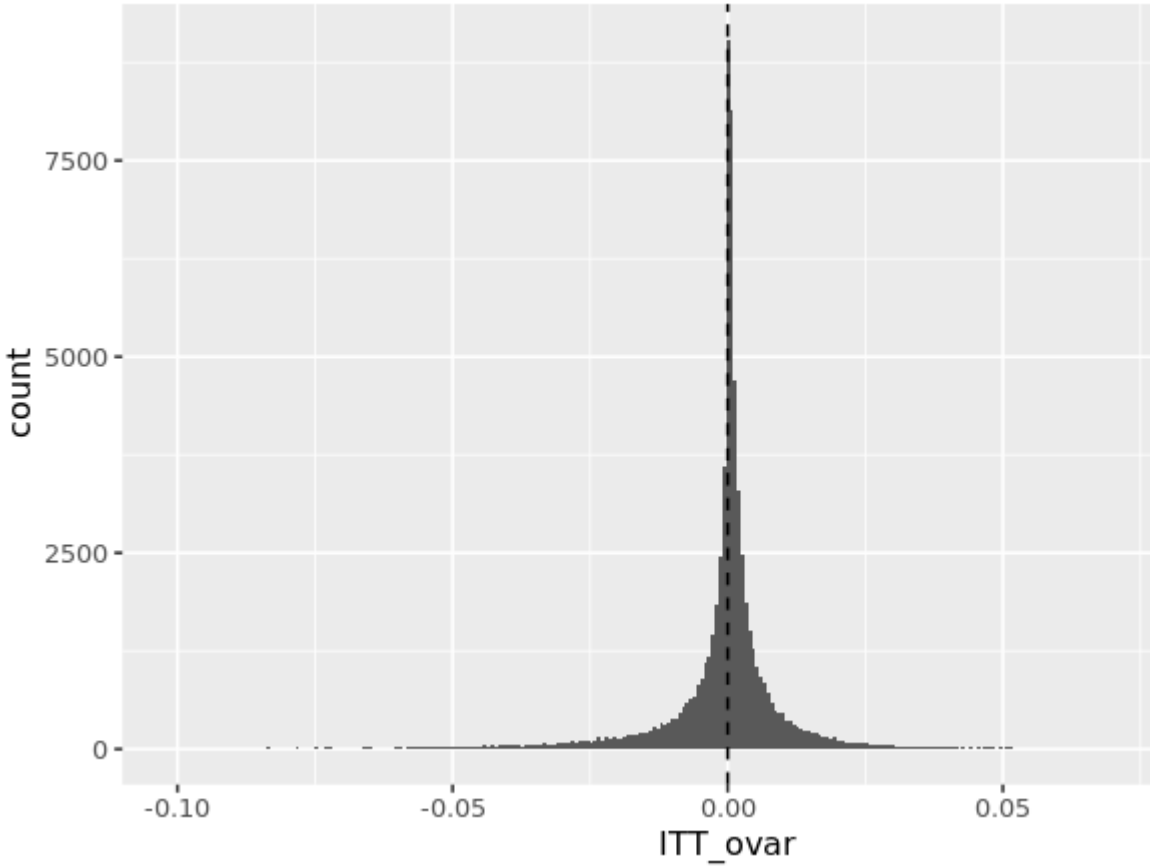
| | Total (n = 68,523) | Assigned to intervention (n = 34,238) | Assigned to control (n = 34,285) |
|---|---|---|---|
| Current sex hormone use | 30,846 (45.0%) | 15,486 (45.2%) | 15,360 (44.8%) |
| **Age at menopause** | | | |
| <40 | 11,952 (17.4%) | 5,936 (17.3%) | 6,016 (17.5%) |
| 40-44 | 8,532 (12.5%) | 4,298 (12.6%) | 4,234 (12.3%) |
| 45-49 | 15,028 (21.9%) | 7,489 (21.9%) | 7,539 (22.0%) |
| 50-54 | 25,119 (36.7%) | 12,597 (36.8%) | 12,522 (36.5%) |
| >= 55 | 7,892 (11.5%) | 3,918 (11.4%) | 3,974 (11.6%) |
| **Number of pregnancies** | | | |
| never | 4,936 (7.2%) | 2,455 (7.2%) | 2,481 (7.2%) |
| 1-2 | 15,364 (22.4%) | 7,708 (22.5%) | 7,656 (22.3%) |
| 3-4 | 45,282 (66.1%) | 22,673 (66.2%) | 22,609 (65.9%) |
| >=5 | 2,941 (4.3%) | 1,402 (4.1%) | 1,539 (4.5%) |

# ITT: EFFECT OF ASSIGNMENT

| | 15-year effect | | 10-year effect | | 6-year effect | |
|---|---|---|---|---|---|---|
| | ATE | ATT + ATC | ATE | ATT + ATC | ATE | ATT + ATC |
| Overall survival | | | | | | |
| - survival difference | –.025 (–.028, –.016) | –.004 (–.012, . –.001) –.038 (–.044, –.032) | –.016 (–.019, –.011) | –.004 (–.009, –.0005) –.025 (–.029, –.020) | –.009 (–.011, –.005) | –.003 (–.006, .0004) –.014 (–.017, –.011) |
| - survival ratio | .970 (.966, .979) | .993 (.985, .998) .954 (.947, .961) | .982 (.979, .988) | .994 (.989, .999) .973 (.968, .978) | .991 (.988, .994) | .997 (.993, 1.000) .986 (.982, .988) |
| OC-specific survival | | | | | | |
| - survival difference | –.002 (–.003, –.000) | –.000 (–.002, .001) –.003 (–.005, –.001) | –.001 (–.002, –.000) | –.000 (–.001, .000) –.002 (–.003, –.000) | –.001 (–.001, –.000) | –.000 (–.001, .000) –.001 (–.001, –.000) |
| - survival ratio | .997 (.996, 1.000) | 1.000 (.998,1.001) .997 (.995, .999) | .999 (.998, 1.000) | .999 (.999, 1.000) .998 (.997, 1.000) | .999 (.998,1.000) | .999 (.999, 1.000) .999 (.999, 1.000) |

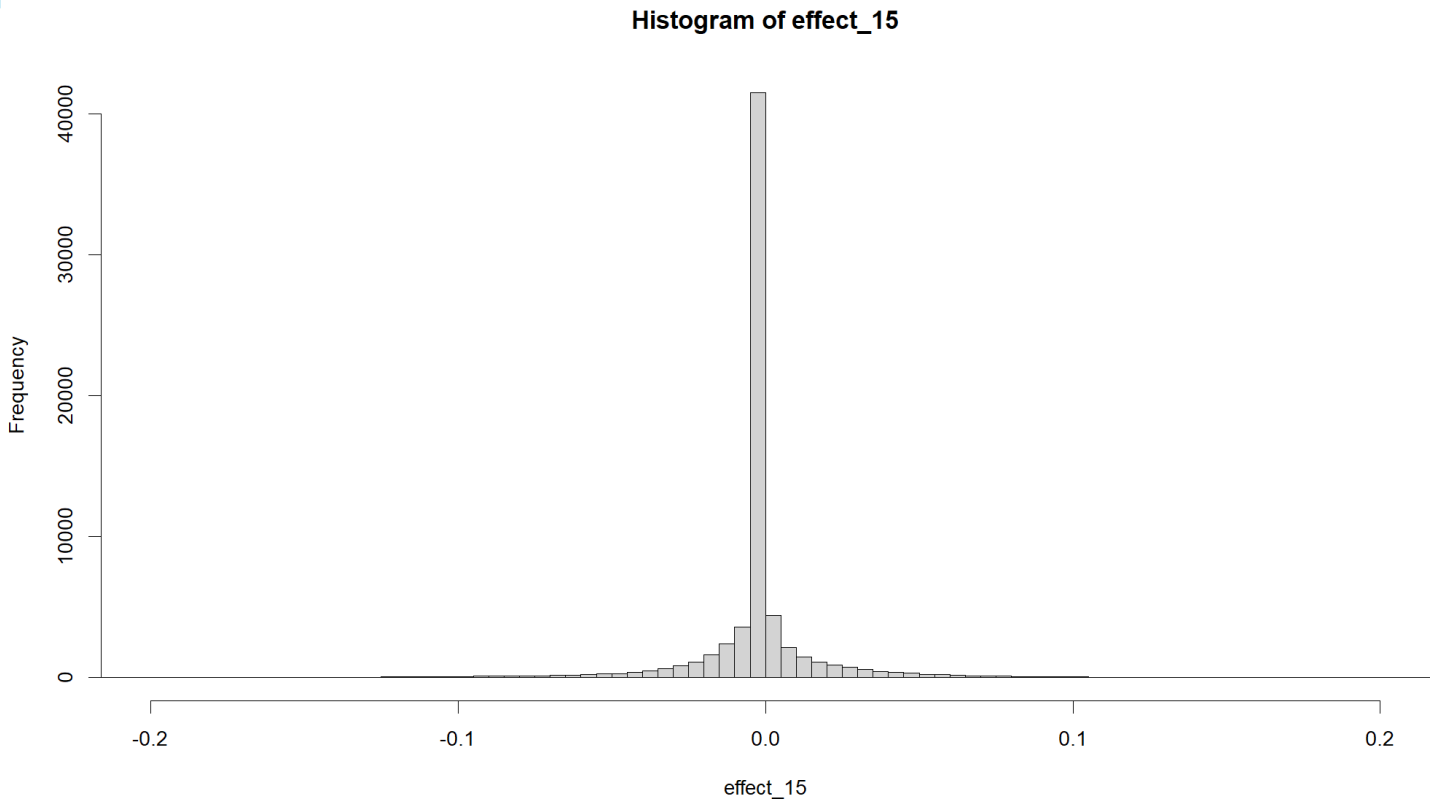



15-year overall survival difference (ITT)




15-year OC-specific survival difference

Erasmus MC
Universitair Medisch Centrum Rotterdam




t.b.d.

## 15-year overall survival difference (treatment)

## 15-year OC-specific survival difference

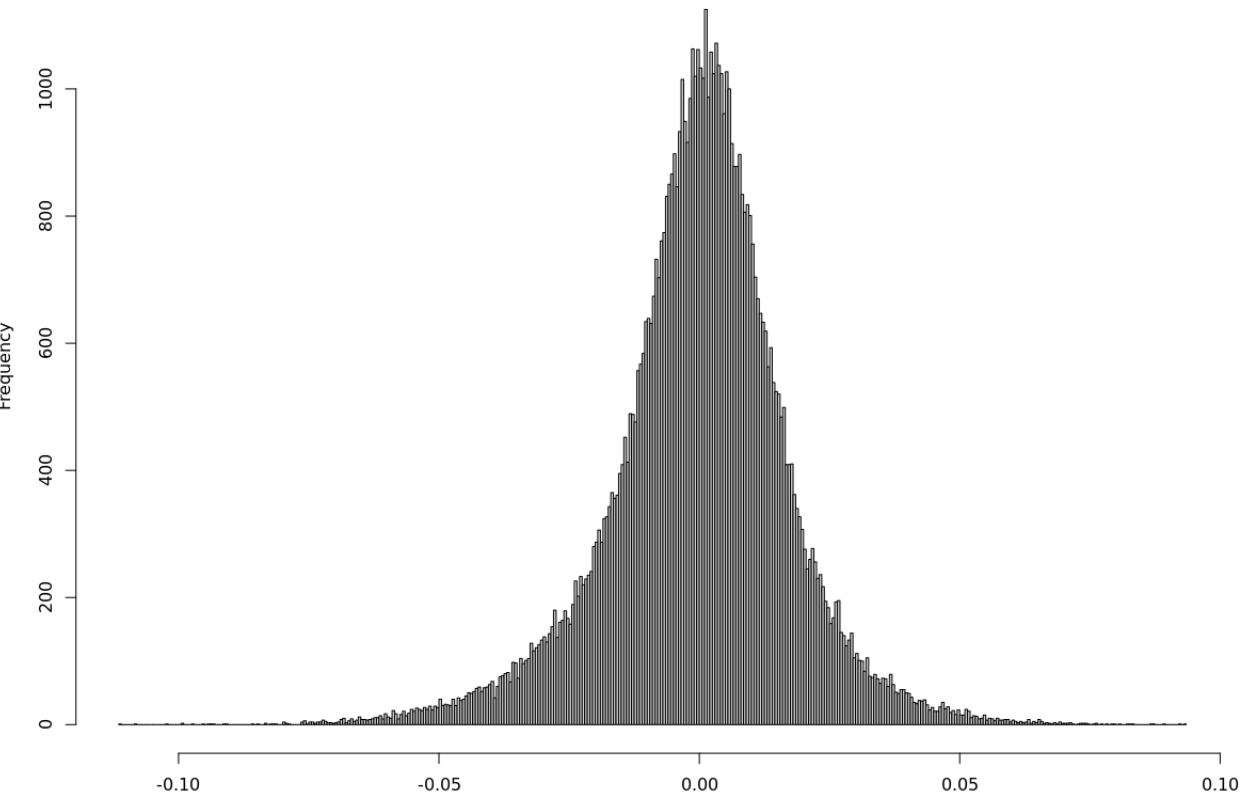
# ITT: MAGNITUDE OF EMM (PART)

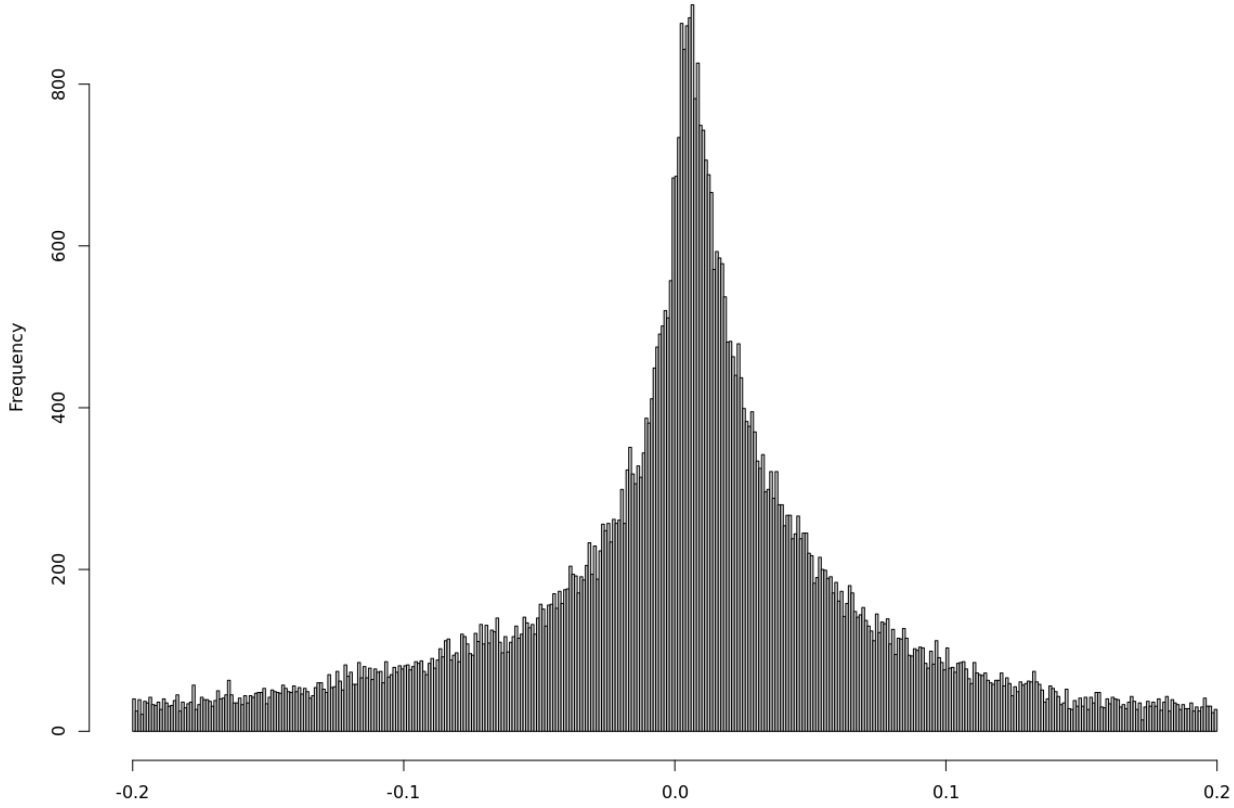
| | Diabetes H. | Colorec. H. | Hormone | Benign Ovar/cyst H. | Pack-years | age | Calorie intake |
|---|---|---|---|---|---|---|---|
| Overall survival | | | | | | | |
| 15-year DoSD | -.033 (-.053, -.011) | -.024 (-.037, -.012) | .012 (.004, .020) | -.019 (-.028, -.009) | -.035 (-.047, -.023) | -.023 (-.038, -.011) | -.002 (-.014, .010) |
| RoSR | .967 (.949, .989) | .965 (.948, .981) | 1.018 (1.007, 1.029) | .973 (.962, .988) | .946 (.931, .964) | .958 (.940, .979) | 1.000 (.984, 1.015) |
| OC-specific survival | | | | | | | |
| 15-year DoSD | -.002 (-.005, .002) | -.001 (-.005, .002) | .000 (-.002, .002) | -.002 (-.004, .001) | -.003 (-.005, .001) | -.005 (-.008, -.001) | .003 (-.001, .006) |
| RoSR | .998 (.995, 1.002) | .998 (.995, 1.003) | 1.000 (.998, 1.002) | .998 (.996, 1.001) | .997 (.995, 1.001) | .995 (.992, .999) | 1.003 (.999, 1.006) |

# POST HOC ROBUSTNESS AND FIT CHECK

- We (plan to) check the following issues to ensure robustness and that the model runs correctly:
- - (1) robustness of gformula+rf: compare results from gformula and from causal forest (grf)
- - (2) performance of gformula+rf: compare the (observable part) of outcome and censoring with gformula+rf predictions: e.g. check if the actually treated individuals are predicted correctly by g-formula under “treated” treatment regime
- - (3) performance of random forest itself: compare rf predictions with actually observed outcomes
- - (4) fit the same gformula+rf model on other imputed datasets to check the robustness of imputation.

# (1) single run comparison of gformula and causal forest




| 10-year overall survival difference (causal forest) | 10-year overall survival difference (gformula) |
|---|---|
| ATE: 0.000 (-0.005, 0.005) | ATE: - 0.016 (-0.019, -0.010) |
| Positive effect: 35785 | Positive effect: 38314 |
| Negative effect: 32738 | Negative effect: 30209 |

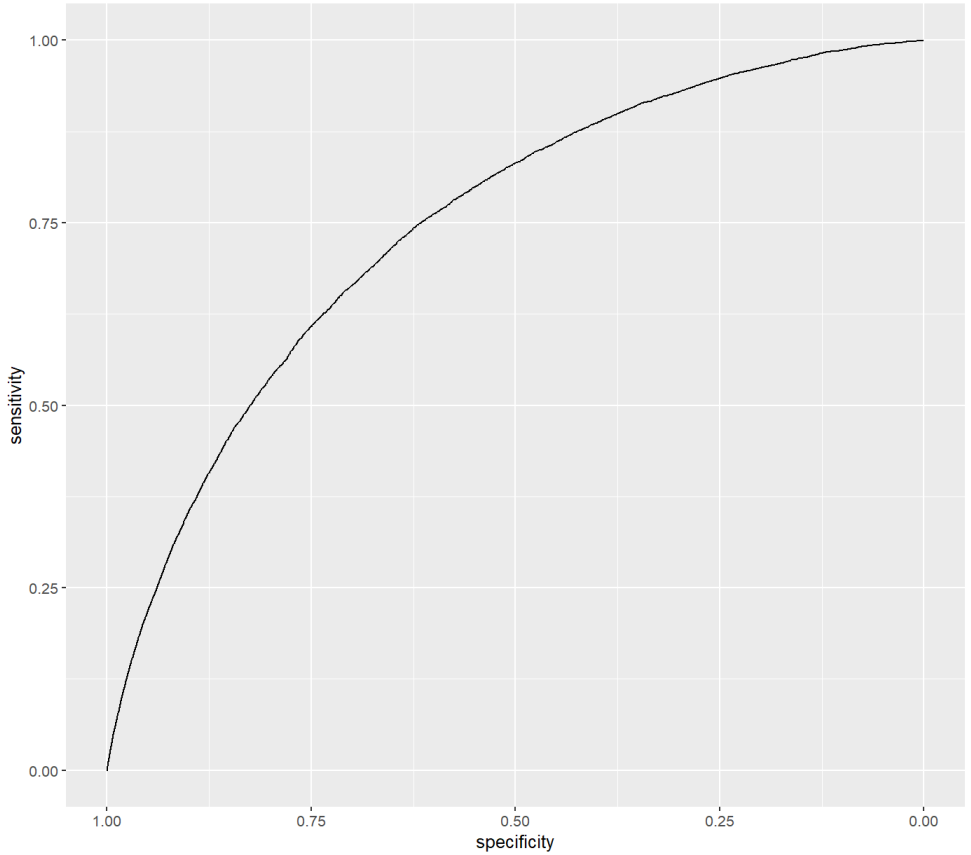
# (1) single run comparison of gformula and causal forest

| | Diabetes H. | Colorec. H. | Hormone | Benign Ovar/cyst H. | Pack-years ($4^{th}$ – $1^{st}$ q.) | age | Calorie intake |
|---|---|---|---|---|---|---|---|
| Overall difference of survival difference (DoSD) from g-formula | | | | | | | |
| 10-year DoSD | -.046 (-.063, -.027) | -.024 (-.033, -.014) | .012 (.005, .018) | -.015 (-.023, -.008) | -.027 (-.037, -.016) | -.018 (-.029, -.007) | .002 (-.007, .012) |
| Overall survival difference from causal forest (single run) | | | | | | | |
| 10-year DoSD | -.011 | -.006 | .002 | -.003 | -.006 | -.001 | .001 |

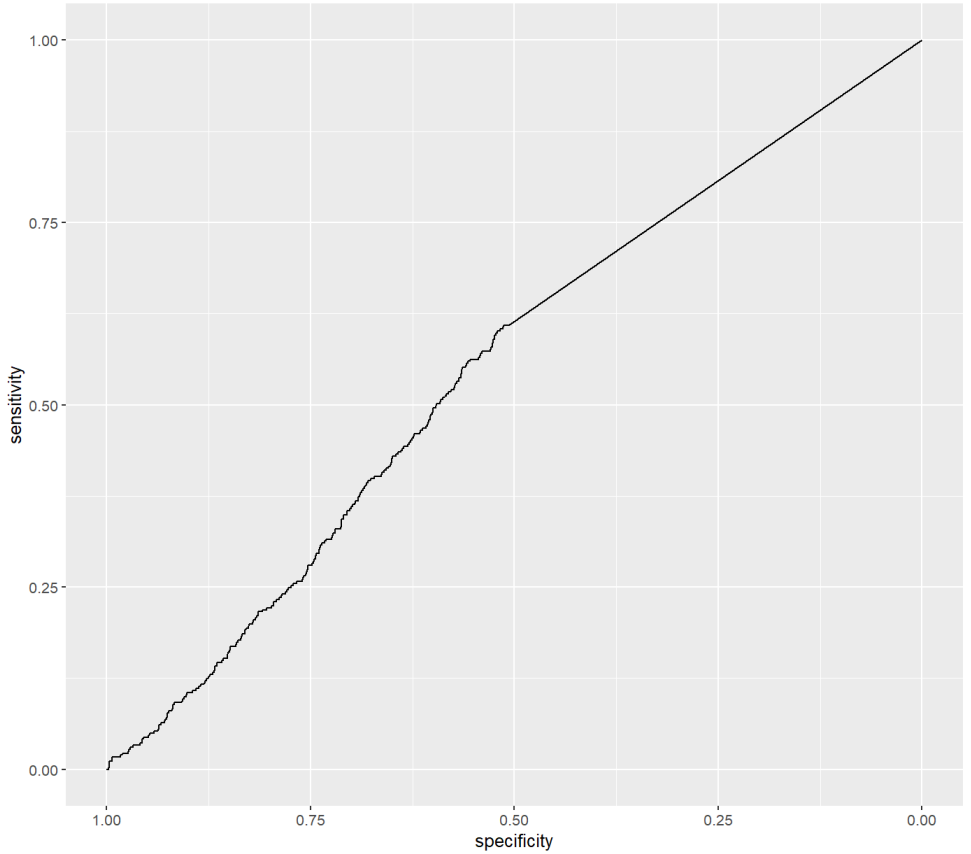
# (2) Performance on observed part of gformula+rf

| Metrics | Among actually treated | Among actually untreated |
|---|---|---|
| Overall survival (all-cause mortality) | | |
| AUC | .9991 | .9952 |
| Brier score | .0071 | .0089 |
| Log Loss | .0432 | .0525 |
| PR-AUC | .9171 | .7338 |
| Ovarian cancer specific survival (OC-specific mortality) | | |
| AUC | - | - |
| Brier score | .0002 | .0002 |
| Log Loss | .0016 | .0018 |
| PR-AUC | .9846 | .9832 |

# (3) Performance of random forest when predicting outcomes




**Predicting overall survival / all-cause mortality**
**AUC = 0.75**




**Predicting cancer specific survival / mortality**
**AUC = 0.56**

Erasmus MC
Universitair Medisch Centrum Rotterdam

# DISCUSSION

- 1. How to share our model? Is it necessary to re-fit a model?
- 2. The refitted GLM performs poorly. We can re-fit a more sophisticated model (RF, NN, etc.) and report their parameters. But then the interpretability is worse.
- 3. The scale of EMM on survival is difference from the scale on cumulative incidence.
- 4. Possible alternatives to model the heterogeneity (causal forest)

# QUESTIONS?

## And suggestions?

# STEP 2: SPECIFYING MODEL FOR EMM

- **We** have different choices of specifying EMM when modelling Y.

**Choice 3:** insert causal forest at each time point after predicting $Y^{a=0}$

$$\Pr[Y^{a=0}|\boldsymbol{X}] = \text{expit}\,(\boldsymbol{X\alpha});$$

$$\tau(\boldsymbol{X}) = \boldsymbol{X\beta};$$

$$\Pr[Y^{a=1}|\boldsymbol{X}] = \text{expit}\,(\boldsymbol{X\alpha}) + \boldsymbol{X\beta}.$$

This way the selection (and regularization) and splitting of predictors for tau(X) are based on the magnitude of effects rather than counterfactual Y's.