Estimating heterogeneous treatment effects to inform targeting of national health insurance programmes

Noemi Kreif, Karla Diaz-Ordaz*

*Karla.Diaz-Ordaz@lshtm.ac.uk

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

1. Introduction

- Since the late '90s, Indonesia has introduced health insurance policy, aiming to 100% coverage
- Health insurance expected to improve health through increasing health care utilisation
- Interest in heterogeneity in treatment effects to inform targeted programme expansion
- **Aim:** explore drivers of heterogeneity in health insurance effect on assisted birth by:
  1. Targeted Maximum Likelihood Estimation (TMLE) with stacked machine learning (*Super Learner, SL*)
  2. Causal Forests
- **Data:** Indonesian Family Life Survey Data (2002-2014):
- Complete cases Birth level dataset ($n = 10985$), 34 baseline variables (denoted X), linked to mother's characteristics, household and community characteristics,

2. Causal estimands

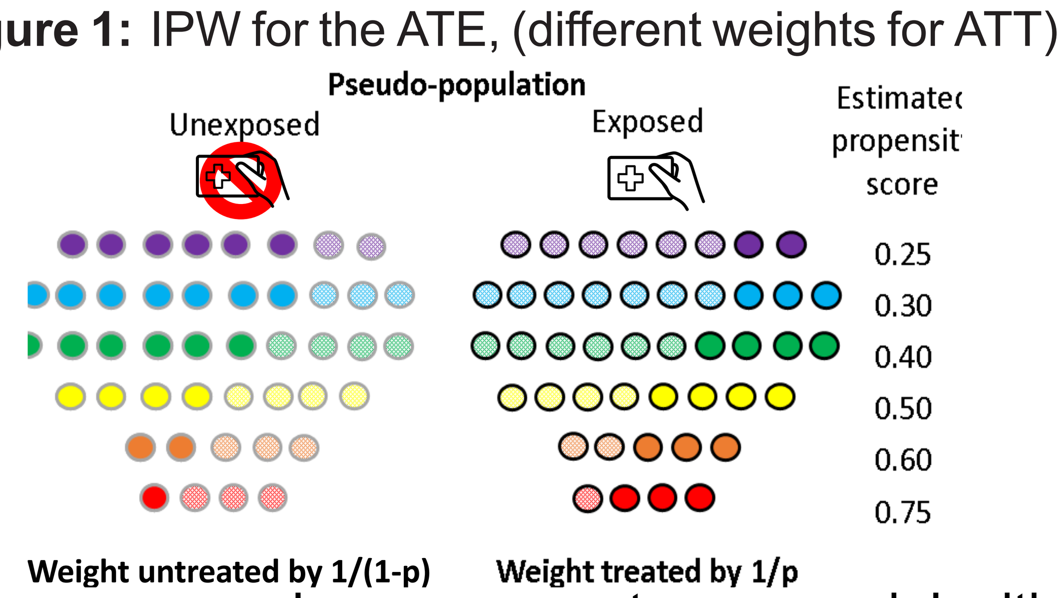
outcome Y = assisted birth,coded 1, n=8574
treatment A=insurance, coded 1, n =1053
Under assumptions **A1**: no interference, consistency and no unobserved confounding , we can identify and estimate from data:

$$\text{ATE} = E\left(Y^1 - Y^0\right),$$
$$\text{ATT} = E\left(Y^1 - Y^0 | A = 1\right),$$
$$\text{ATC} = E\left(Y^1 - Y^0 | A = 0\right),$$
$$\text{CATC}(x) = E\left(Y^1 - Y^0 | A = 0, X = x\right).$$

3.IPW and DR

- ATE can be estimated by outcome model adjusting for all confounders $E(Y|A,X)$, assuming that: **(A2)** the regression model is **correctly specified**
- alternatively, use *propensity score* $p(X) = E(A|X)$, and estimate via a simple model on exposure using Inverse probability of treatment weighting, assuming **A3** PS is correctly specified

Figure 1: IPW for the ATE, (different weights for ATT)



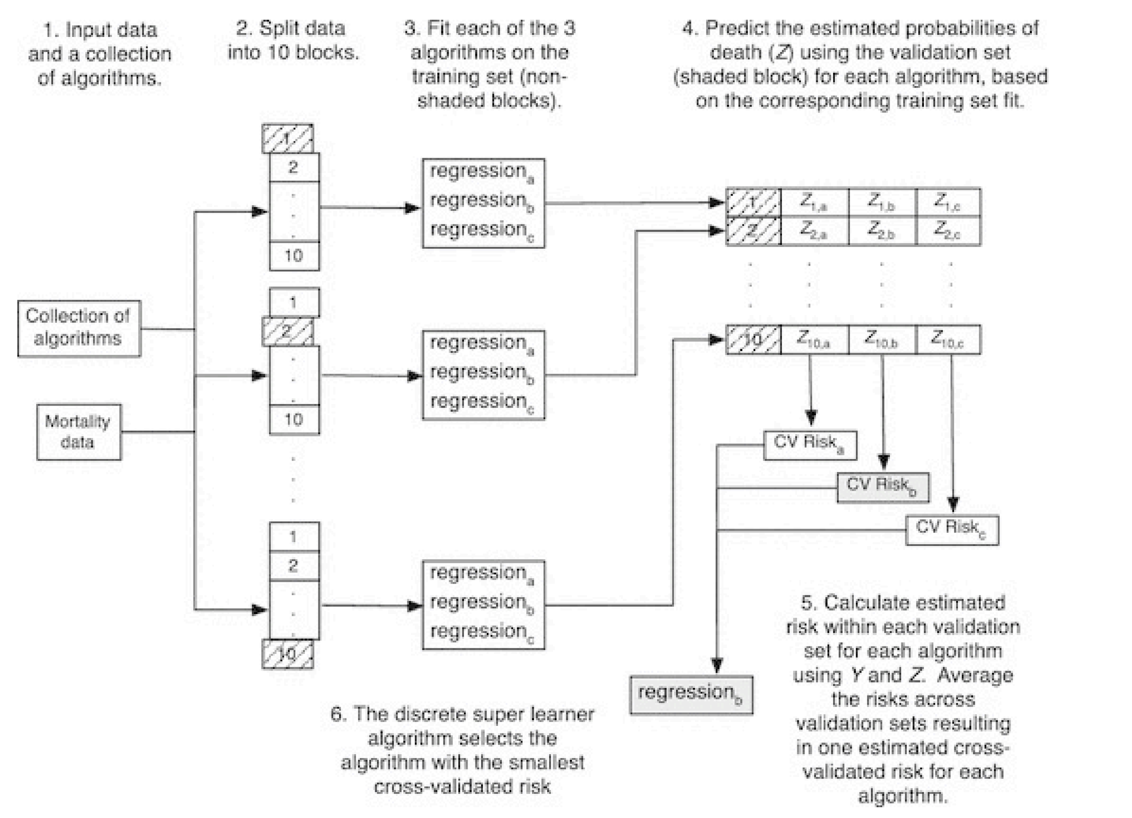
- we can also use an outcome model with IPW and get **doubly-robust estimators (DR)**, consistent if either model is correct.

4.TMLE for ATE and CATC

- **better still** we can use machine-learning (SL, see Panel 5→) to estimate the PS and outcome models, reducing model misspecification ⇒ Double Machine learning or **TMLE**
- Step 1 – Estimate PS and mean potential outcome $\mu^0(x)$ and $\mu^1(x)$ using SL with library:
  - logistic regression (pair-wise interactions), GAMs, random forests, boosting, BARTs
- Step 2 - Calculate individual-level treatment effects $\widehat{\tau}(x) = \widehat{\mu}^1(x) - \widehat{\mu}^0(x)$
- Step 3- use Random Forest for variable importance of effect modifiers; results:
  - Age at child birth
  - Receipt of cash transfer
  - Year of birth (of baby)
  - Can write in Indonesian
- Step 3 - Estimate ATE(T/C)s and CATCs conditional on some of these variables
- $\widehat{\mu}(A, X)$ predictions updated with SL $\widehat{p}(X)$

5. SuperLearner (SL)

- Stacking learner using cross-validation to train multiple machine learners
- SL creates an optimal weighted average of the predictions obtained by each learner
- asymptotically as accurate as the best possible prediction algorithm considered.
- rates of convergence: depend on the individual learners



6. Causal Forests for ATE and CATC

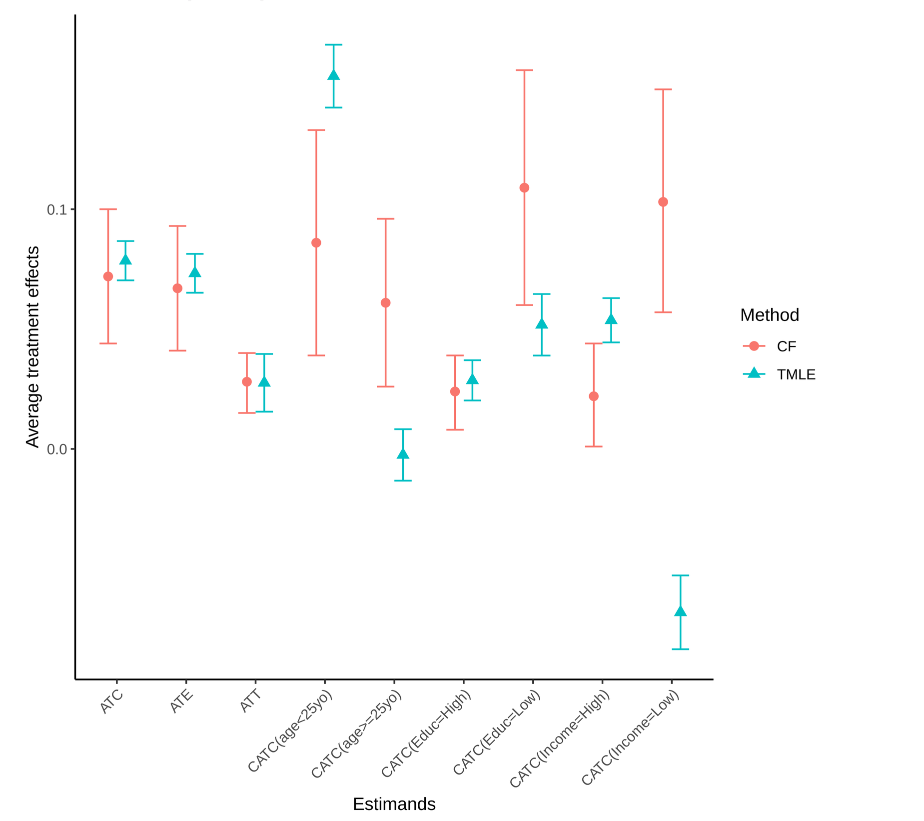
- Causal Forests: DR estimator by (Weighted) estimating equation targets $\tau(x)$, estimation using random forests with splitting rule to maximise heterogeneity in estimated treatment effect
- uses sample splitting: in one tree, an observation is either used to select splits or estimate $\tau(x)$, keeping the inference *honest*
- Forests are formed using subsample aggregation with estimated weights
- Step 1 – Use regression forests to obtain estimates of $p(X)$ and $\mu(x) = E(Y|X)$
- obtain out-of-bag predictions from these, and plug in into

$$\tau(x) = \frac{\sum_i w_i(x)\left\{Y_i - \mu(x)\right\}\left\{A_i - p(X_i)\right\}}{\sum_i w_i(x)\left\{A_i - p(X_i)\right\}}$$

- Step 2: grow a "Raw" Causal Forest; allows to split trees on confounders
- calculate the weights $w_i(x)$ and
- Obtain variable importance measure to find important effect-modifiers
- Step 3: Re-estimate Causal Forest splitting trees on selected variables (from previous step)
- Obtain predictions using out of bag observations to estimate individual treatment effects and ATE, ATT, ATC and CATCs
- Use **omnibus "best linear predictor" test** for the presence of heterogeneity

7. Results: Effect of on

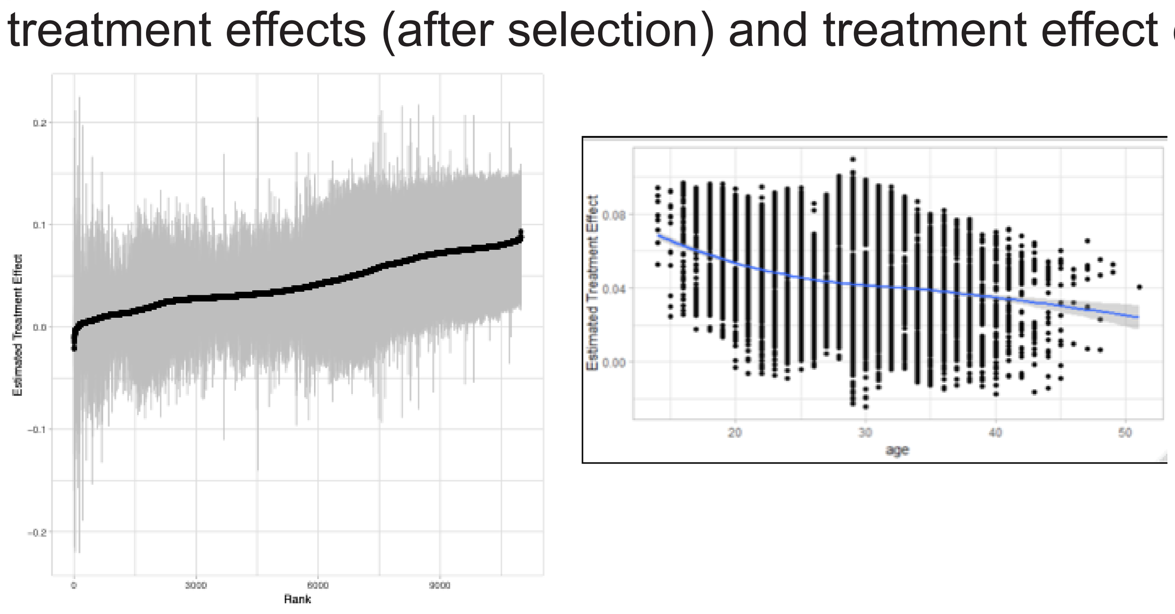
Figure 2: Estimates for ATE, ATT and CATC conditional on greater effect-modifiers, by TMLE and Causal Forest (CF)



8. Results: Heterogeneous Treatment effects

- The omnibus test found no evidence of heterogeneity ($p = 0.14$)
- However, it would appear that younger mothers have larger positive effect of insurance on the probability of having a health professional when given birth

Figure 3: heterogeneous treatment effects (after selection) and treatment effect curve as a function of age



9. Conclusions

- TMLE and Causal Forest report heterogeneity in effects of health insurance on assisted birth. TMLE has narrower CIs.
- Younger, poorer, less educated mothers benefit more. Larger estimated effect among controls
- **Limitation of subgroup TMLE:**
  - Nuisance parameters estimated to fit full sample (not subgroups)
  - only indirectly learnt about CATE, though subgroups
- **Next:** optimal treatment: allocation rule for next extension of health insurance, maximising benefits (subject to budget constraints and equity)