Using data-adaptive methods to investigate causal conditional treatment effects: towards personalised treatment regimes

Karla DiazOrdaz

@karlado

supported by wellcome

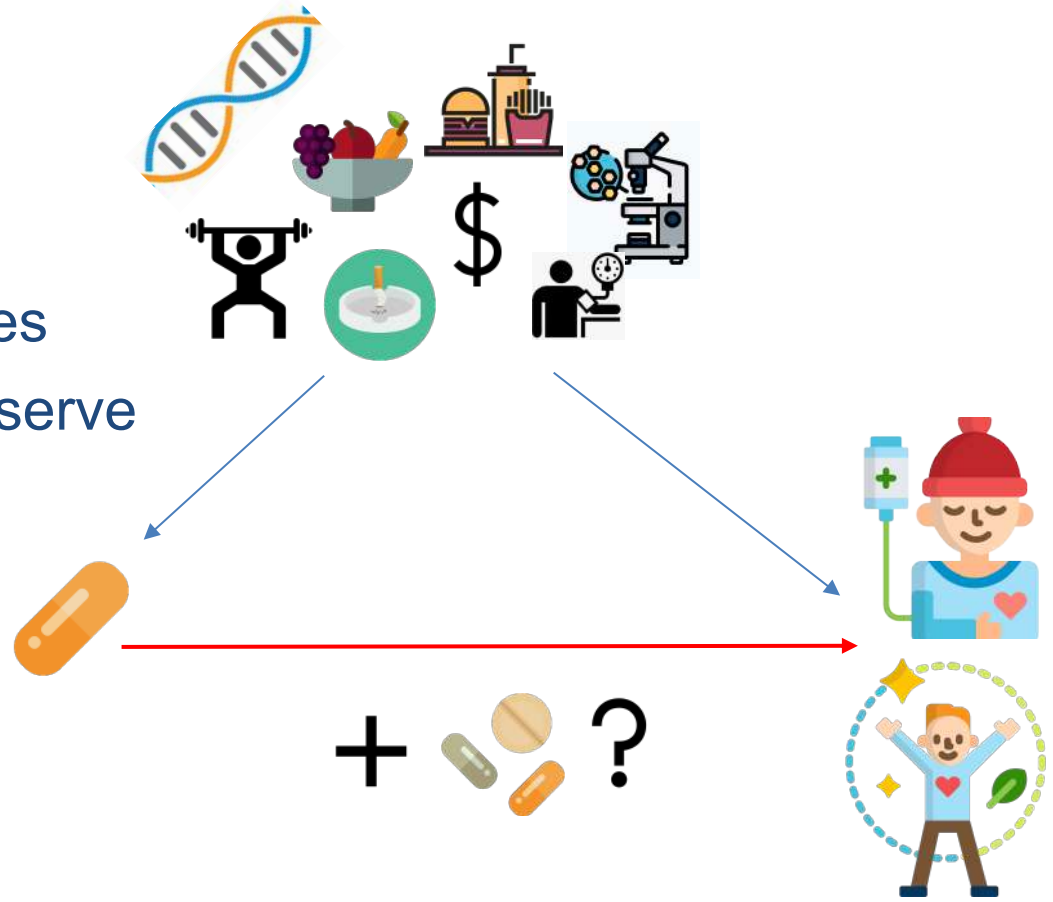
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

CENTRE for STATISTICAL METHODOLOGY

The Alan Turing Institute

Motivation:

- Big questions in Comparative effectiveness/ policy evaluation:
  - *which treatment works?*
  - *for whom?*
  - Can we optimise benefit by adding other treatments?
- often using observational data: confounding
- Understanding effect modifiers: personalised regimes
- Fundamental problem of causal inference: can't observe the same unit with and without the treatment
- So, we aim at estimating:
  - the average causal effect
  - conditional average treatment effect



Motivating example: Treatment intensification of T2DM patients

aged ≥18 between Jan 2000 and July 2017, with a minimum of 12 months of prior registration in CPRD

Potential outcomes

Y^1 if exposed

Y^0 if unexposed

Outcome:
Hba1c at 12, 24 48, 60 weeks

Causal estimand
(involving counterfactuals)

$$ATE = E[Y^1 - Y^0],$$
or

$$CATE(x) = E[Y^1 - Y^0 | X = x],$$
for baseline (vector) X
e.g. X baseline Hba1c and eGFR

Identifying assumptions

- No interference
- Consistency
- conditional exchangeability given C

Statistical estimand
(function of the observed data)

$$ATE = E[Y|A = 1, C] - E[Y|A = 0, C],$$

and/or

$$CATE(x) = E[Y|A = 1, C, X = x] - E[Y|A = 0, C, X = x]$$

- 3 possible treatments considered (after metformin)
  - sulphonylureas (SUs),
  - sodium-glucose co-transporter-2 inhibitors (SGLT2is) or
  - dipeptidyl peptidase-4 inhibitors (DPP4is)
- 31 baseline variables as controls for confounding: age, sex, BMI, renal function, hba1c, statin use

Estimation:

- outcome regression
- Inverse probability of treatment weighting
- Doubly robust methods

⇒ Modelling assumptions

Modelling assumptions

1. Outcome regression estimates the ATE by β:

$$\mu(C) = E[Y|A,C] = \alpha + \beta A + \gamma^T C$$

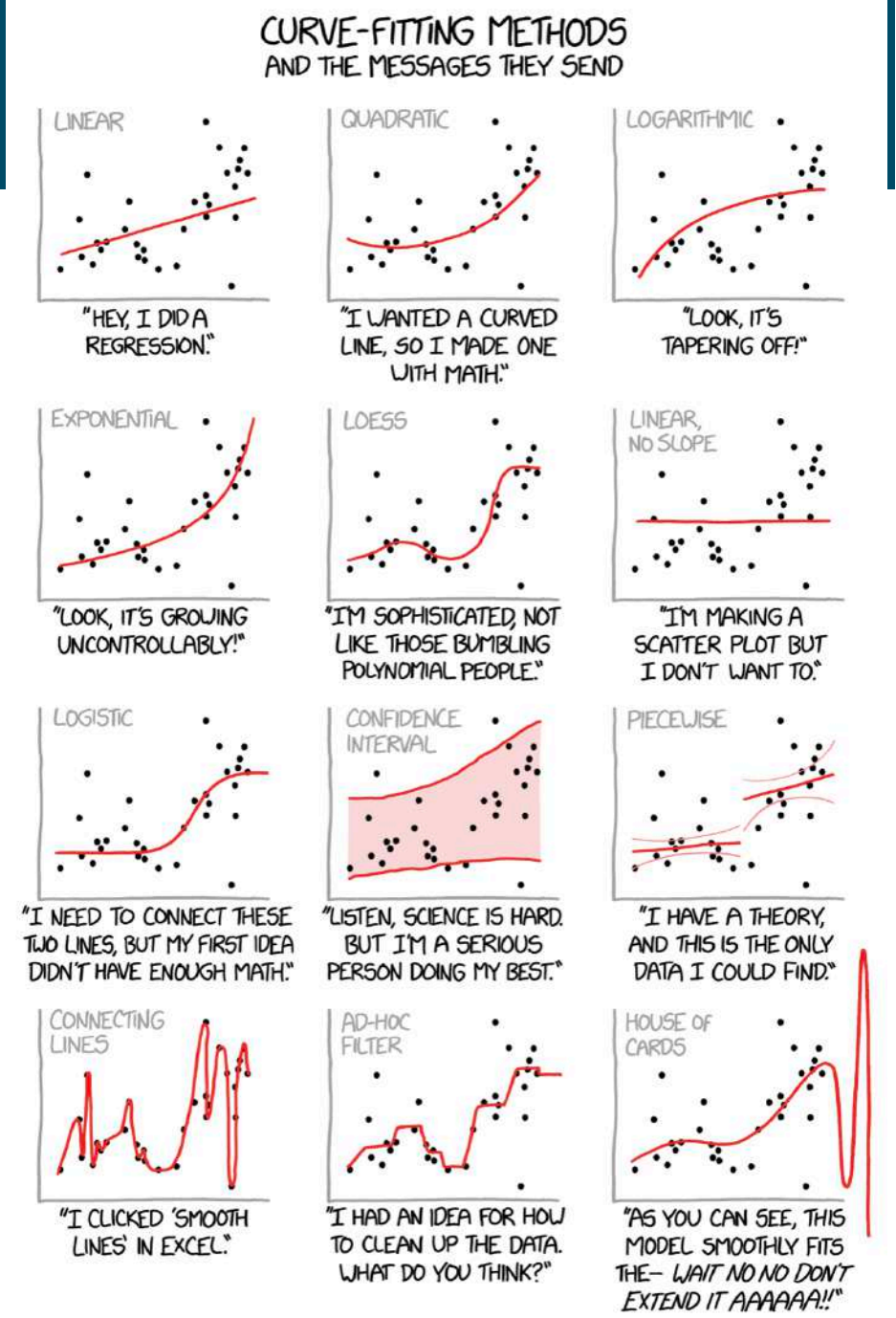
- outcome regression model is correctly specified
- can be checked from the data:
  - Overfitting
  - extrapolation

2. Propensity scores: we model $p(C) = Pr[A = a|C]$

- Central result : if conditional exchangeability holds given C, then conditional exchangeability also holds given $p(C)$.
- Positivity of the treatment assignment

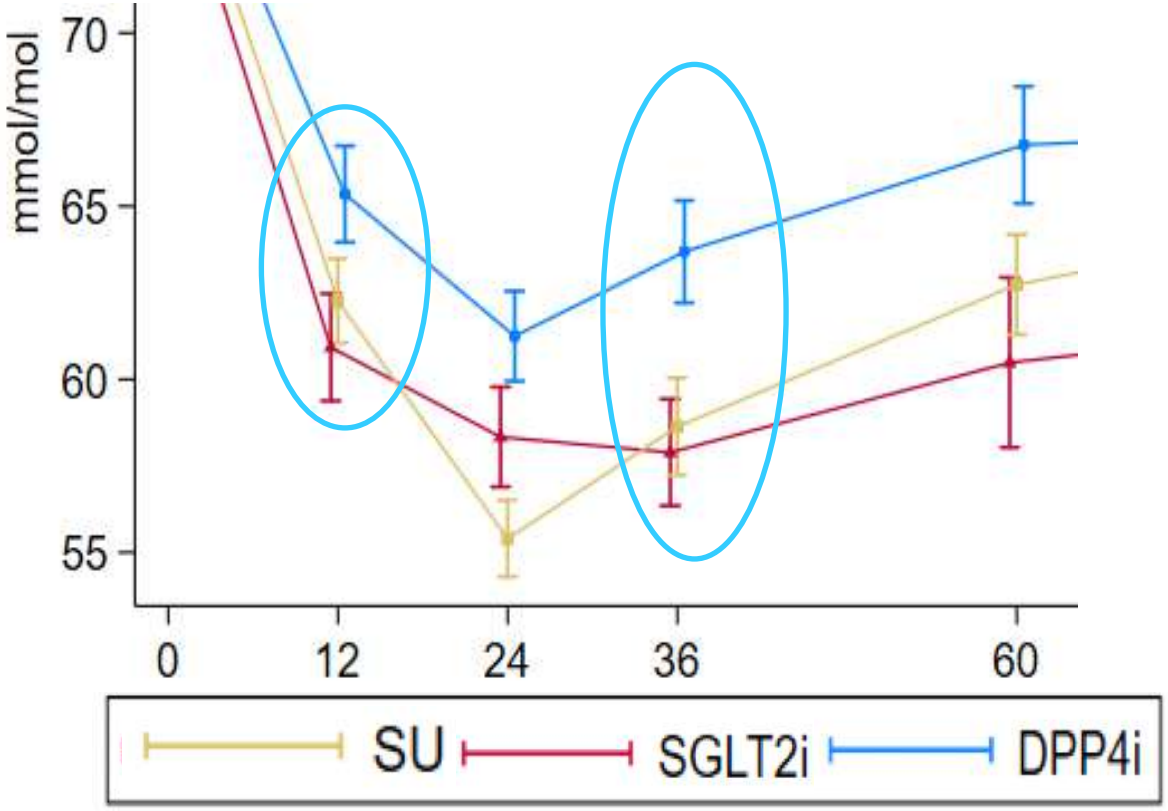
$$0 < Pr[A = a|C] < 1$$

- $p(C)$ must be correctly specified
- Model misspecification is likely and difficult to diagnose
- Especially with poor overlap

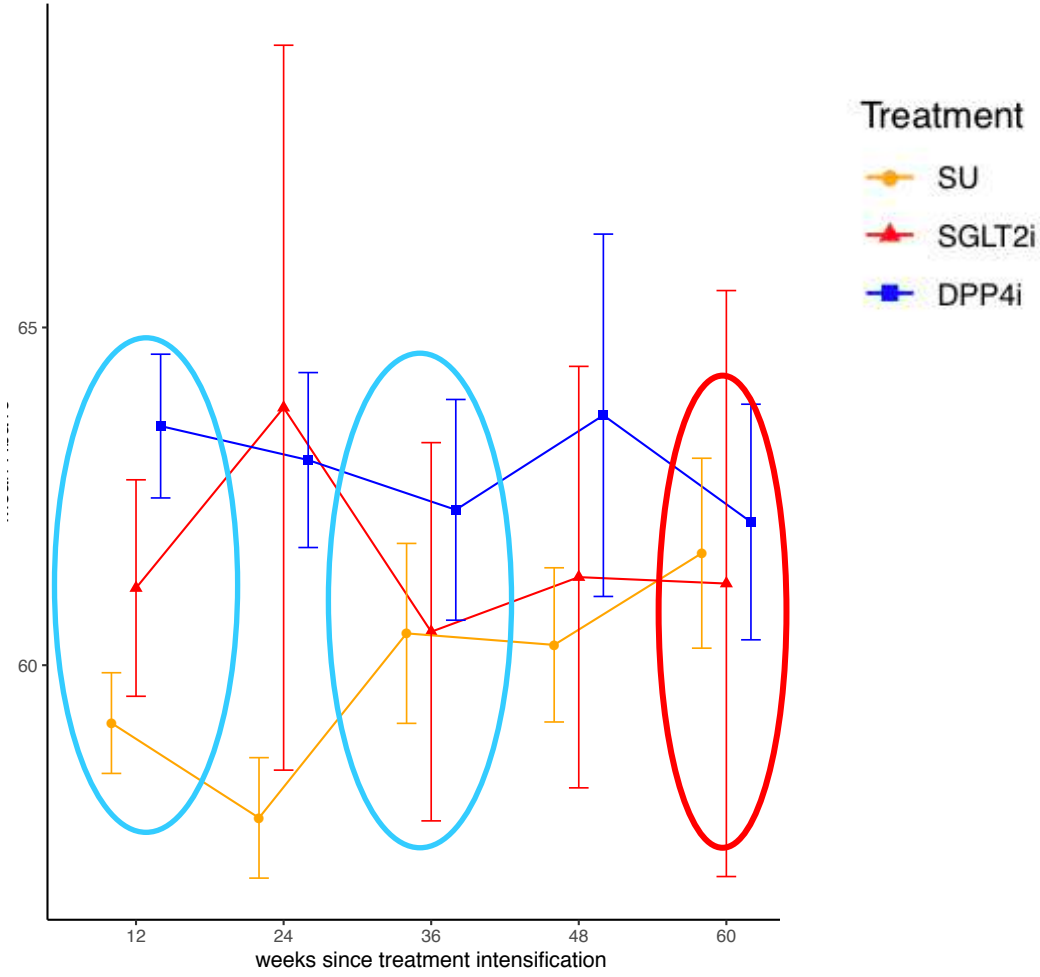


T2DM cohort for treatment intensification

- Original multinomial PS matching found:



- DR with multinomial PS found:



Treatment effect heterogeneity

- Perhaps the ATE is masking effects in certain populations
- We can estimate treatment effect modification curves: conditional ATE, given a predetermined variable X
- Very often we want to know how effects vary with covariates
    - to explore mechanisms
    - to target a specific population
- Understanding this can be important for finding optimal treatment regimes
- Recent interest in using machine learning to find drivers of treatment effect heterogeneity
- **Problem: We don't know how to do uniformly valid confidence bands for CATE(x).**

Two possible solutions

- Wager and Athey (JASA 2018) make progress by focusing on random forest with assumptions guaranteeing consistency **: causal forest**
- Chernozhukov et al. (arXiv:1712.04802) build on **Double Machine Learning**:
- Key: be less ambitious and focus on:
    - is there heterogeneity?
    - what are the characteristics of those with the largest treatment effect?
- Many other novel CATE with ML (e.g. Kunzel, Sekhon and Bickel, Luedtke and van der Laan, 2016)
- Next: we use causal forests to explore drivers of heterogeneity of T2DM treatments

Estimating *heterogeneous* treatment effects via Causal Forest

- Ideally we would like to learn individual treatment effect:

$$\text{ITE} = Y_i^1 - Y_i^0$$

- but not feasible (Fundamental problem of causal inference)
- Instead, we target the CATE function for a given X:

$$\tau(x) = E\left[Y_i^1 - Y_i^0 | X_i = x\right]$$

- **Causal Forests** "customises" the random forest algorithm (Breiman 2001) to predict CATE instead of the observed outcomes

Causal Forests (Wager and Athey 2018, Athey et al. 2019)

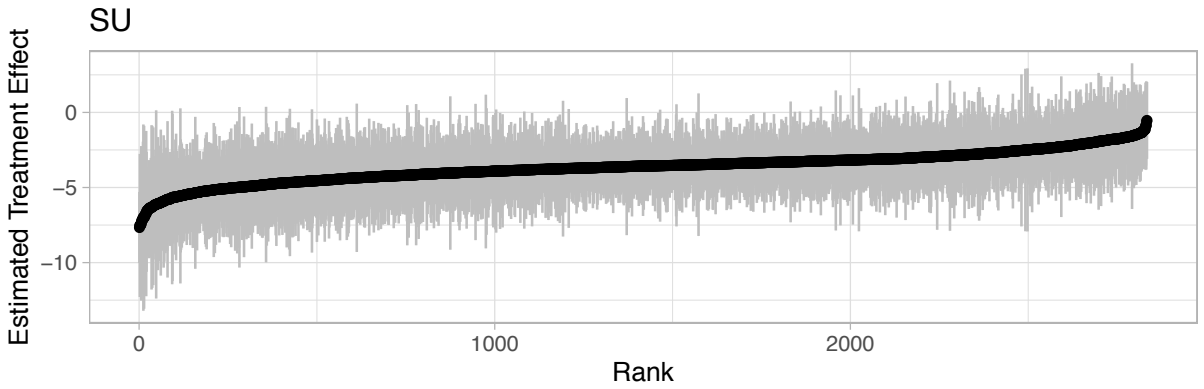
- Builds "causal trees": **maximise heterogeneity in estimated treatment effect** as opposed to minimise RMSE in outcome prediction
- ***"Honesty"*** = Sample splitting:
    - in one tree, i is either used to select splits or estimate $\tau(x)$
- Forests are formed using weighted aggregation
- weights chosen to minimise bias in $\tau(x)$
- Estimator is consistent asymptotically normal, inference is possible
- Implements an omnibus test of heterogeneity

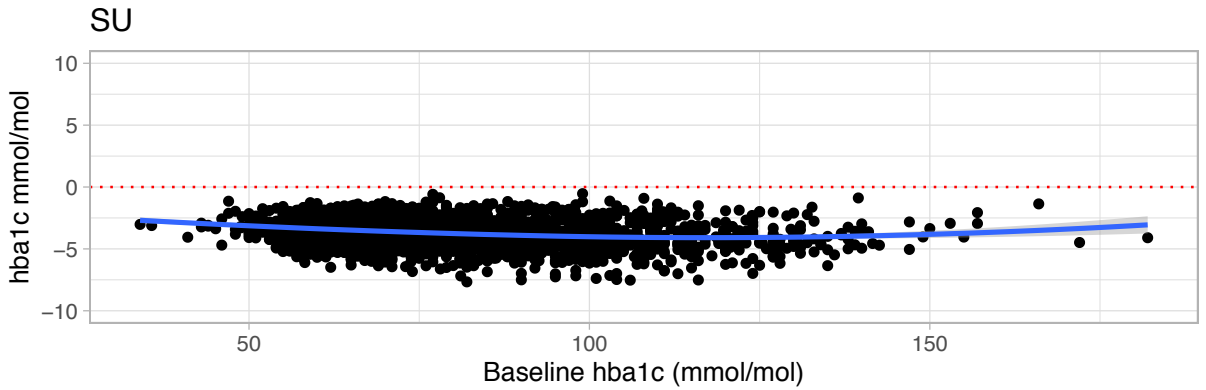
Implementation of “Causal Forests” for an observational study (package grf Athey, Tibshiriani and Wager 2019)

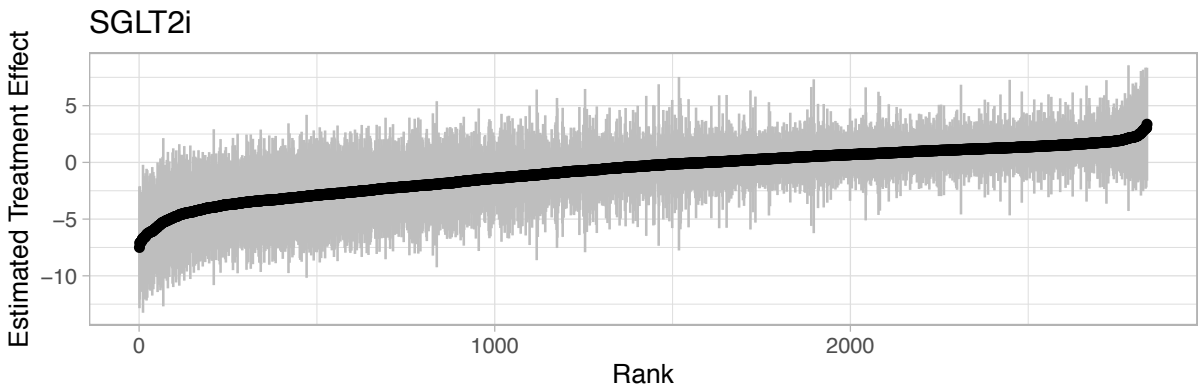
- Step 1: **Deal with confounding**
  - Use regression forests to obtain estimates of:
    - outcome model $\mu(C)$ and
    - PS model $p(C)$
  - Use the residuals (outcome and exposure minus corresponding mean) as “transformed outcomes” (unconfounded if models correct)
  - “double robust” property
- Step 2: **Estimate “Raw” Causal Forests on the transformed outcomes**
  - Select most important effect modifiers using variable importance
- Step 3**: Re-Estimate Causal Forest** with most influential variables as effect modifiers
  - Obtain estimates of individual treatment effects with CIs
  - obtain omnibus test for heterogeneity
- Step 4**: Estimate ATE and CATE**

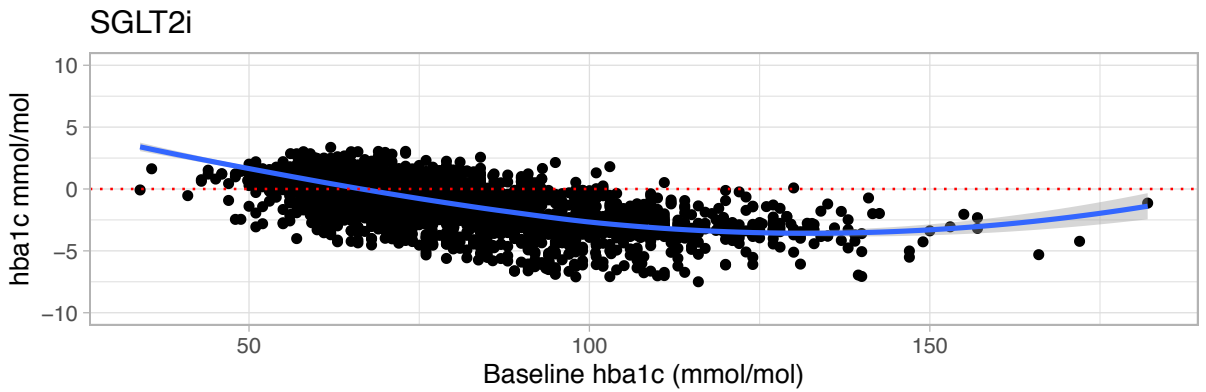
Heterogenous treatment effects and CATE for baseline Hba1c at 12w

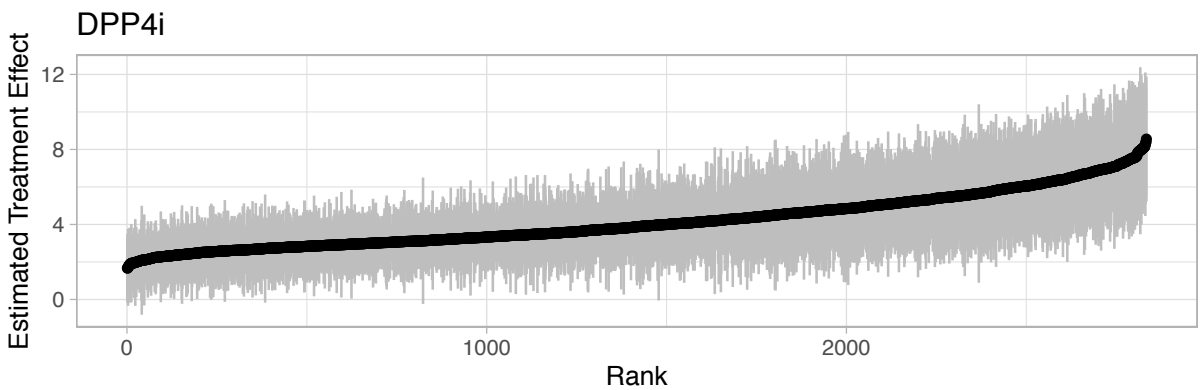
Omnibus test of heterogeneity : p=0.0003 (for all 3 treatments)

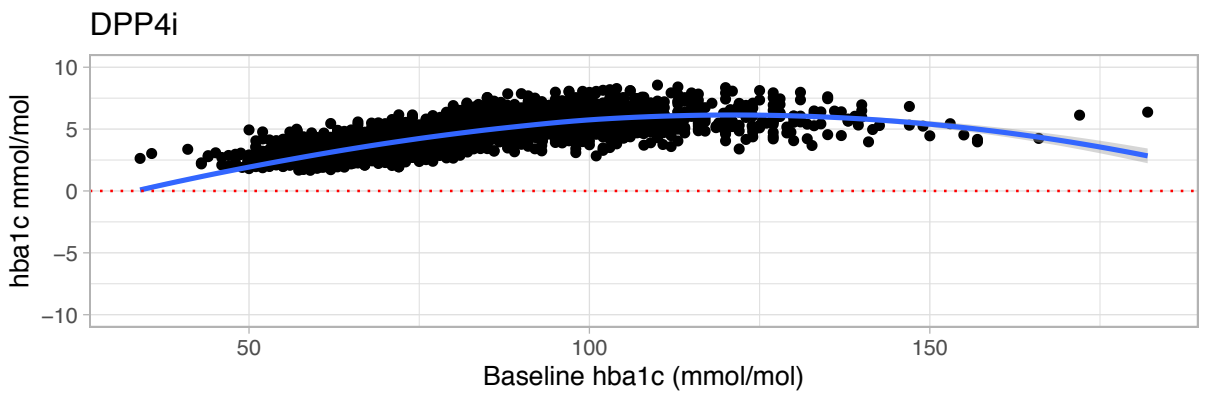






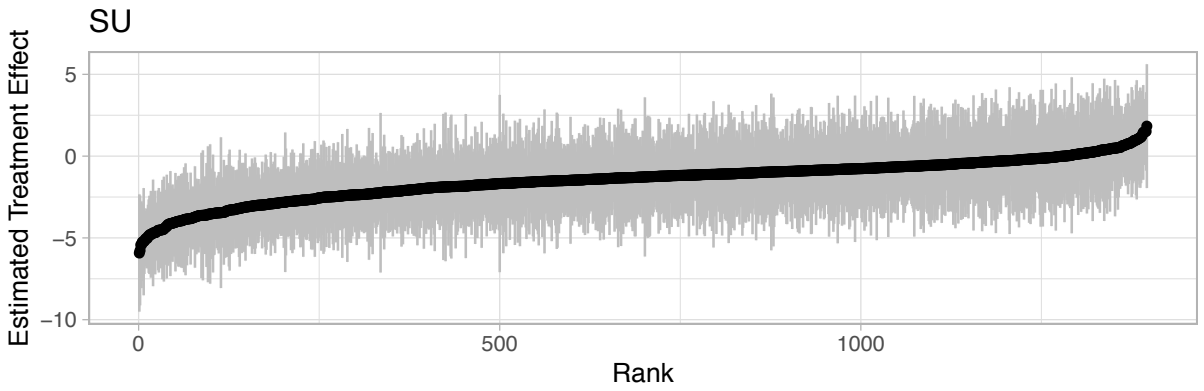


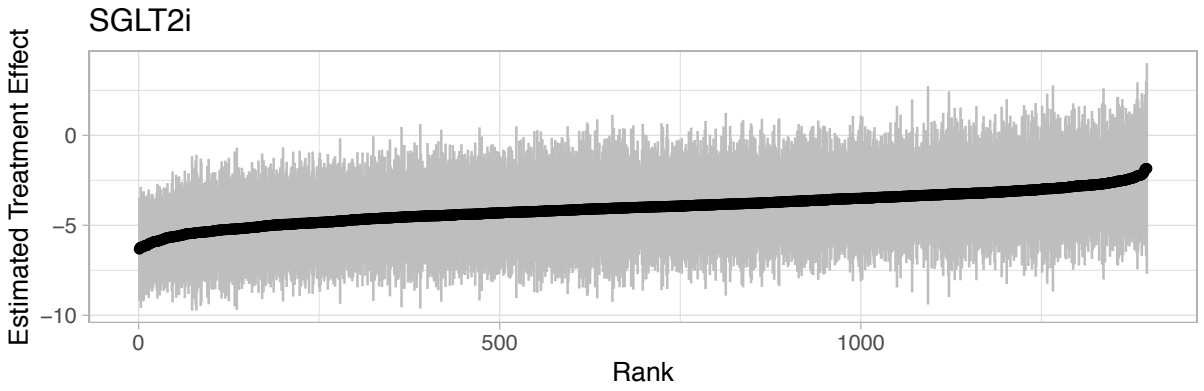


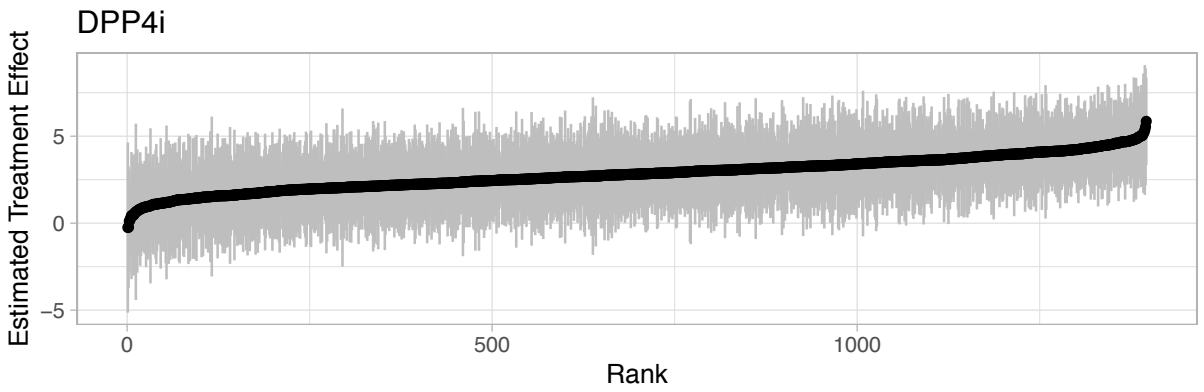


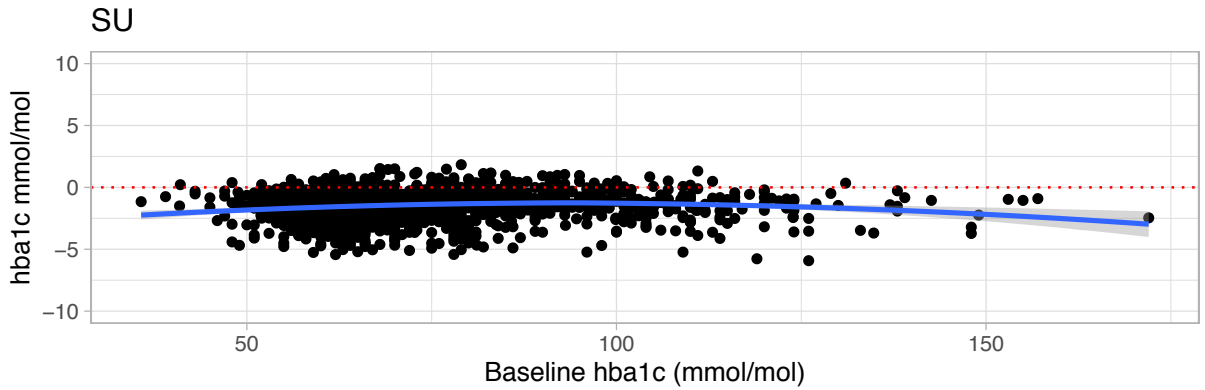
Heterogenous treatment effects and CATE for baseline Hba1c at 48w

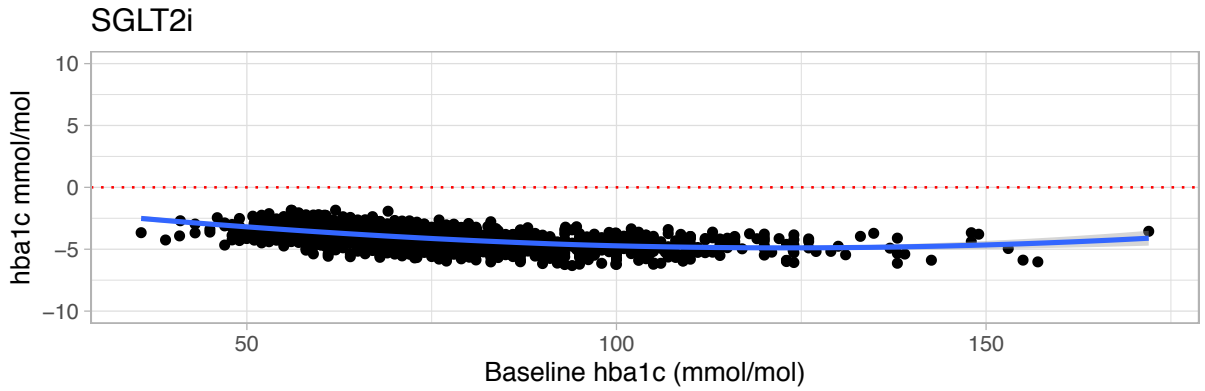
Omnibus test of heterogeneity : p=0.05 for SU vs the rest

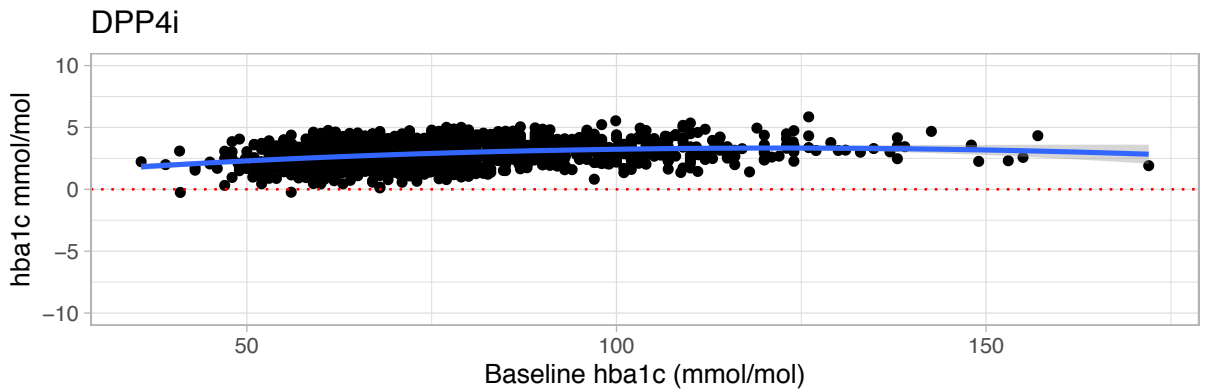




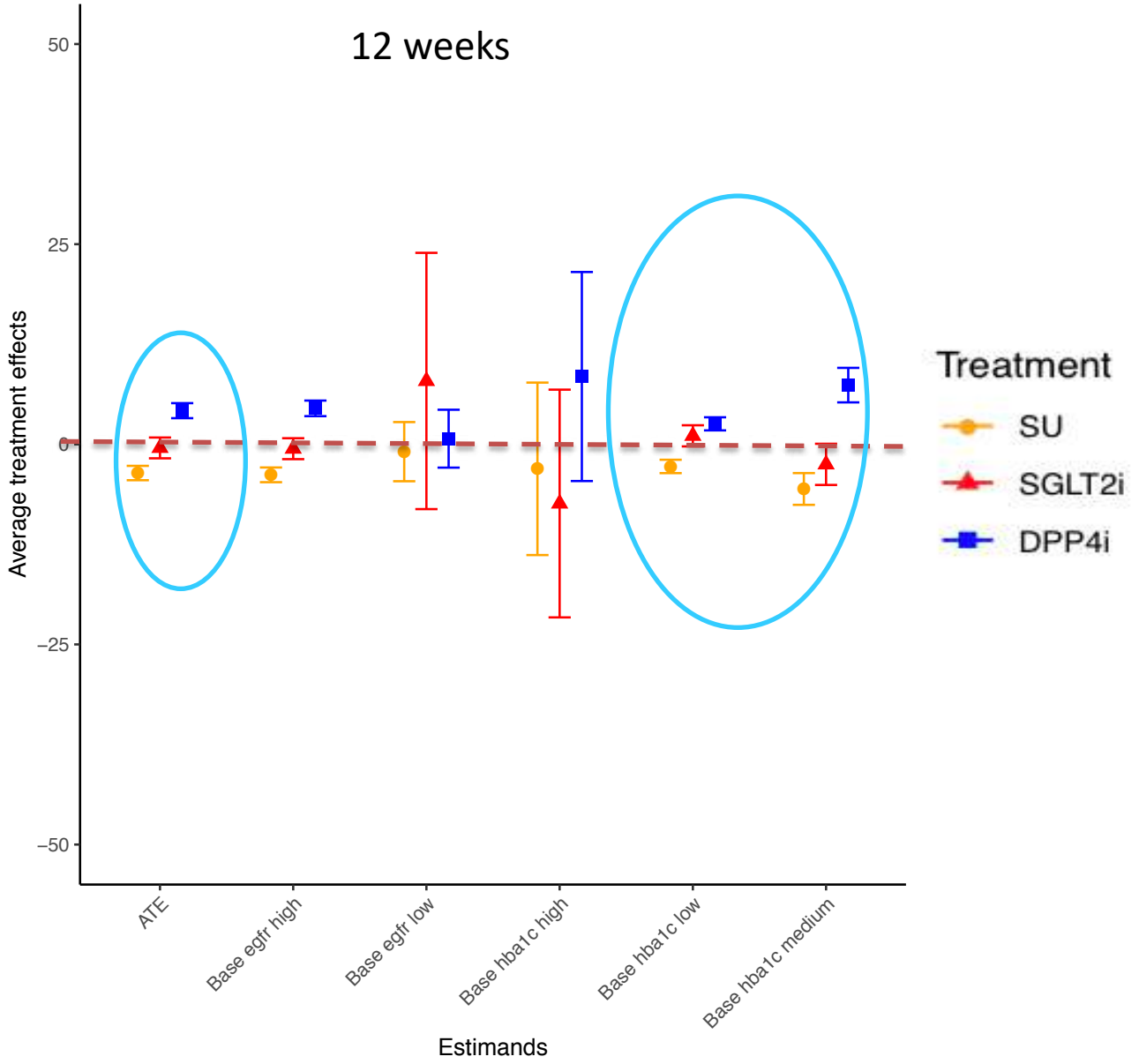


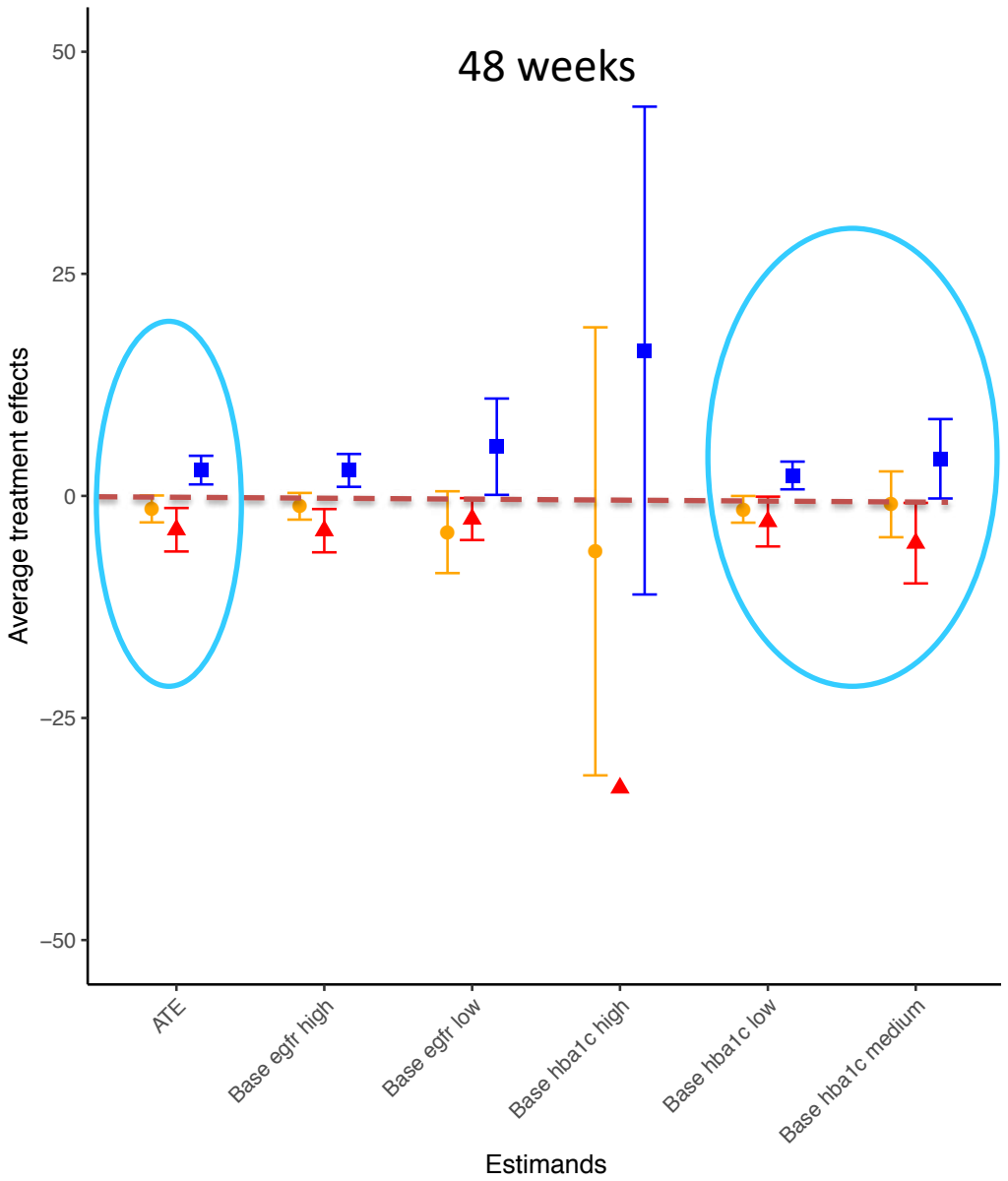






ATE and Conditional treatment effects (each drug vs the others)





Discussion

- Interpret in light of small numbers

N	12w	48 w
SU	1402	763
SGLT2i	294	109
DPP4i	1138	524
total	2834	1396

- DR (non-ML) and CF (ML) approaches gave similar results for ATE
- CF helps detect heterogeneity in treatment effects
- Next step: double-robust machine learning for CATEs via g-estimation
- Maybe time-updated covariates explain better the treatment effect heterogeneity. Need to consider further treatment intensification
- Potential for remaining unobserved confounding
    - ML developed for IV
        - ATE: Belloni et al. 2012,
        - for CATE via g-estimator: DiazOrdaz et al. 2018 (arXiv)

References

1. Wager S & Athey S (2018) Estimation and Inference of Heterogeneous Treatment Effects using Random Forests, Journal of the American Statistical Association, 113:523, 1228-1242, DOI: 10.1080/01621459.2017.1319839
2. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J. Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, Volume 21, Issue 1, 1 February 2018, Pages C1–C68
3. Athey S, Tibshirani J, Wager S. Generalized random forests. The Annals of Statistics. 2019;47(2):1148-78
4. Belloni A, Chen D, Chernozhukov V, Hansen C. Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica. 2012 Nov;80(6):2369-429.
5. DiazOrdaz K, Daniel Rh, Kreif, N Data-adaptive doubly robust instrumental variable methods for treatment effect heterogeneity Pre-print on arXiv https://arxiv.org/abs/1802.02821