

tehtuner: An R package to fit and tune models for the conditional average treatment effect

Jack M. Wolf ¹

¹ University of Minnesota Division of Biostatistics, USA

DOI: [10.21105/joss.05453](https://doi.org/10.21105/joss.05453)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Chris Hartgerink](#)  

Reviewers:

- [@elimillera](#)
- [@wwangstat](#)

Submitted: 02 April 2023

Published: 19 June 2023

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Randomized clinical trials (RCTs) often test and describe the average treatment effect, or how the candidate intervention is expected to increase or decrease an outcome of interest for all patients in the population of interest. However, secondary analyses may seek to identify ways in which underlying subject characteristics such as age or health status may modify the expected treatment effect, resulting in *treatment effect heterogeneity*. This phenomenon is expressed through the *conditional average treatment effect* (CATE) which is defined as

$$\tau(x) = E(Y^1 - Y^0 | X = x)$$

where x is a vector of covariate measurements and Y^1 and Y^0 are the potential outcomes that would be observed under the treatment and control arms, respectively. Information about the CATE can then be used to determine the optimal treatment on a subject-to-subject basis (i.e., “personalized medicine”) or identify sub-populations for which additional interventions or support are needed.

tehtuner fits models to estimate the CATE using the Virtual Twins method ([Foster et al., 2011](#)) while controlling the method’s probability of falsely detecting treatment modifiers when all subjects would respond to treatment the same by implementing the permutation procedure proposed in Wolf et al. (2022). A key feature of Virtual Twins is that it estimates a simple model such as a regression tree which can be easily interpreted to understand the CATE as opposed to other popular data-adaptive methods which trade in interpretability for model flexibility. This is accomplished through a two-step procedure which first uses a flexible method such as random forests to estimate each subject’s anticipated response under each treatment (Step 1) and then models the difference in these response estimates through a simple model such as a regression tree (Step 2).

Statement of need

Although there are several readily available R packages that can estimate the CATE ([Hill, 2011](#); [J. Tibshirani et al., 2022](#); [Vieille & Foster, 2018](#)); there are few at the time of writing which both estimate an interpretable model *and* guarantee controlled behavior when there is no treatment effect heterogeneity. While Vieille & Foster (2018) does provide an implementation of the original Virtual Twins manuscript, it does not support the methods in Steps 1 and 2 evaluated and recommended in Deng et al. (2023) such as super learner ([van der Laan et al., 2007](#)) and is prone to overfitting when there is no effect heterogeneity. Fan & Hong (2022) can estimate the effect of a given covariate on the treatment effect after marginalizing across all other covariates. This approach yields interpretable and well-behaved models but is best used for confirmatory rather than exploratory analyses as it is unable to analyze the CATE in a multivariate manner and requires covariate pre-specification to avoid Type I error inflation.

At this current time, `tehtuner` supports linear models fit via the LASSO (R. Tibshirani, 1996), MARS (Friedman, 1991), random forests (Breiman, 2001), and super learner (van der Laan et al., 2007) in Step 1 and linear models tuned via the LASSO (R. Tibshirani, 1996), regression and classification trees (Breiman et al., 2017), and conditional inference trees (Hothorn et al., 2006) in Step 2. Comparative evaluations of these methods can be found in Wolf et al. (2022) and Deng et al. (2023).

Example Usage

The primary function is `tunevt()`, which first fits the Step 1 model and then fits the Step 2 model with an appropriate penalty parameter to ensure that the probability of incorrectly detecting any treatment effect modifiers when there is no treatment effect heterogeneity is a user-specified value.

The following code fits a tuned Virtual Twins model using a random forest in Step 1 and a regression tree in Step 2 while setting the probability of falsely detecting treatment effect heterogeneity at 20%. In this example, the fitted model for the CATE can be used to identify three distinct subgroups with differential estimated treatment effects.

```
library(tehtuner)

vt_mod <- tunevt(
  data = tehtuner_example, Y = "Y", Trt = "Trt",
  step1 = "randomforest", step2 = "rtree", alpha0 = 0.2,
  p_reps = 100, ntree = 50
)

vt_mod
#> Call:
#> tunevt(data = tehtuner_example, Y = "Y", Trt = "Trt", step1 = "randomforest",
#>       step2 = "rtree", alpha0 = 0.2, p_reps = 100, ntree = 50)
#>
#> Step 2 "rtree" model:
#> n = 1000
#>
#> node), split, n, deviance, yval
#>      * denotes terminal node
#>
#> 1) root 1000 18372.4300  1.6830340
#>   2) V1< -3.008541 511  5643.7030 -0.3942647 *
#>   3) V1>=-3.008541 489  8219.4140  3.8537890
#>     6) V3>=0.000282894 19  448.7299 -5.3806930 *
#>     7) V3< 0.000282894 470  6084.9480  4.2270980 *
#>
#> Approximate 0.8 quantile of the MNPP null distribution: 0.07695
#> Observed MNPP: 0.2454, p-value: < 2.22e-16
```

Acknowledgments

The author would like to thank their advisors, Drs. David M. Vock and Joseph S. Koopmeiners, for their support in developing the methods implemented in this software. This work was supported by the National Institute on Drug Abuse (Award Number R01DA046320). The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Drug Abuse.

References

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification And Regression Trees*. Routledge. <https://doi.org/10.1201/9781315139470>
- Deng, C., Wolf, J. M., Vock, D. M., Carroll, D. M., Hatsukami, D. K., Leng, N., & Koopmeiners, J. S. (2023). Practical guidance on modeling choices for the virtual twins method. *Journal of Biopharmaceutical Statistics*. <https://doi.org/10.1080/10543406.2023.2170404>
- Fan, Q., & Hong, H. (2022). *Hdcate: Estimation of Conditional Average Treatment Effects with High-Dimensional Data*. <https://CRAN.R-project.org/package=hdcate>
- Foster, J. C., Taylor, J. M. G., & Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, 30(24), 2867–2880. <https://doi.org/10.1002/sim.4322>
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1), 1–67. <https://doi.org/10.1214/aos/1176347963>
- Hill, J. L. (2011). Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*, 20(1), 217–240. <https://doi.org/10.1198/jcgs.2010.08162>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Tibshirani, J., Athey, S., Sverdrup, E., & Wager, S. (2022). *grf: Generalized Random Forests*. <https://CRAN.R-project.org/package=grf>
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288. <https://www.jstor.org/stable/2346178>
- van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super Learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1). <https://doi.org/10.2202/1544-6115.1309>
- Vieille, F., & Foster, J. (2018). *aVirtualTwins: Adaptation of Virtual Twins Method from Jared Foster*. <https://CRAN.R-project.org/package=aVirtualTwins>
- Wolf, J. M., Koopmeiners, J. S., & Vock, D. M. (2022). A permutation procedure to detect heterogeneous treatment effects in randomized clinical trials while controlling the type I error rate. *Clinical Trials*, 19(5), 512–521. <https://doi.org/10.1177/17407745221095855>