

# tidyhte: Tidy Estimation of Heterogeneous Treatment Effects in R

Drew Dimmery<sup>1</sup> and Edward Kennedy<sup>2</sup>

<sup>1</sup> Hertie School Data Science Lab <sup>2</sup> Department of Statistics and Data Science, Carnegie Mellon University, USA

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

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

Editor: [Open Journals](#)

## Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

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

## Summary

Heterogeneous treatment effects (HTE) describe how intervention impacts vary across individuals or subgroups. Understanding this variation is crucial for targeting interventions, optimizing resource allocation, and identifying mechanisms of action. `tidyhte` provides a principled framework for estimating heterogeneous treatment effects using modern machine learning methods. The package implements the doubly-robust learner of Kennedy (2023), combining outcome modeling and propensity score estimation with cross-validation to produce valid statistical inference on treatment effect heterogeneity.

The package uses a “recipe” design that scales naturally from single to multiple outcomes and moderators. Users specify machine learning algorithms for nuisance function estimation (treatment propensity and outcome models) and define moderators of interest; `tidyhte` handles cross-validation, model selection, diagnostics, and construction of quantities of interest. The tidy design integrates seamlessly with the broader R data science ecosystem and supports common empirical practices including weighting for population average effects and clustered treatment assignment.

## Statement of Need

Existing tools for heterogeneous treatment effect estimation often require substantial statistical expertise and involve navigating complex decisions about cross-validation, model selection, doubly-robust estimation, and valid inference. Implementations are scattered across packages with inconsistent interfaces, making reliable application difficult.

`tidyhte` addresses these challenges by providing a unified interface that: (1) implements state-of-the-art doubly-robust methods with automatic cross-validation, (2) uses intuitive “recipe” semantics familiar from packages like `recipes` (Kuhn et al., 2025) and `parsnip` (Kuhn & Vaughan, 2024), (3) handles both experimental and observational data, (4) scales from single to multiple outcomes and moderators, (5) provides built-in diagnostics for model quality and effect heterogeneity, and (6) returns tidy data formats for downstream analysis. By automating technical details while maintaining statistical rigor, the package makes modern HTE methods accessible to applied researchers who need to understand treatment effect variation.

## Research Applications

`tidyhte` supports research across clinical trials (identifying patient subgroups), policy evaluation (understanding differential population impacts), technology (optimizing interventions for user segments), economics (studying policy effects across demographics), and education (evaluating differential intervention impacts). The package was stress-tested as the HTE estimation

39 software for publications on Facebook and Instagram's effects on the 2020 US Presidential  
40 election, involving estimation across approximately ten moderators and dozens of outcomes  
41 (Guess et al., 2023a, 2023b; Nyhan et al., 2023).

## 42 Related Work

43 The package builds on the theoretical framework of Kennedy (2023) for doubly-robust HTE  
44 estimation. It leverages the SuperLearner ensemble learning framework (Van der Laan et al.,  
45 2007) for flexible machine learning. The tidy design philosophy follows Wickham (2014) and  
46 integrates with the broader tidyverse ecosystem (Wickham et al., 2019).

47 Other R packages for HTE estimation include grf (Athey et al., 2019; Tibshirani et al., 2024)  
48 for generalized random forests, bcf (Hahn et al., 2020; Murray et al., 2023) for Bayesian causal  
49 forests, and FindIt (Imai & Ratkovic, 2013, 2018). tidyhte differentiates itself through its  
50 doubly-robust methodology (Kennedy, 2023), recipe interface, support for scaling to multiple  
51 outcomes and moderators, and native handling of weights and clustering.

## 52 Acknowledgements

53 We gratefully acknowledge the collaboration with the US 2020 Facebook and Instagram  
54 Election Study team for being a testbed for the initial versions of tidyhte, particularly Pablo  
55 Barberá for testing this package during development and providing valuable feedback on the  
56 design and functionality. We received no financial support for this software project.

## 57 References

- 58 Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of*  
59 *Statistics*, 47(2), 1148–1178. <https://doi.org/10.1214/18-AOS1709>
- 60 Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A.,  
61 Dimmery, D., Freelon, D., Gentzkow, M., & others. (2023a). How do social media feed  
62 algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656),  
63 398–404.
- 64 Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A.,  
65 Dimmery, D., Freelon, D., Gentzkow, M., & others. (2023b). Reshares on social media  
66 amplify political news but do not detectably affect beliefs or opinions. *Science*, 381(6656),  
67 404–408.
- 68 Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models  
69 for causal inference: Regularization, confounding, and heterogeneous effects. *Bayesian*  
70 *Analysis*, 15(3), 965–1056. <https://doi.org/10.1214/19-BA1195>
- 71 Imai, K., & Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized  
72 program evaluation. *The Annals of Applied Statistics*, 7(1), 443–470. [https://doi.org/10.](https://doi.org/10.1214/12-AOAS593)  
73 [1214/12-AOAS593](https://doi.org/10.1214/12-AOAS593)
- 74 Imai, K., & Ratkovic, M. (2018). *FindIt: Finding heterogeneous treatment effects*. <https://doi.org/10.32614/CRAN.package.FindIt>
- 75
- 76 Kennedy, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal  
77 effects. *Electronic Journal of Statistics*, 17(2), 3008–3049. [https://doi.org/10.1214/](https://doi.org/10.1214/23-EJS2157)  
78 [23-EJS2157](https://doi.org/10.1214/23-EJS2157)
- 79 Kuhn, M., & Vaughan, D. (2024). *Parsnip: A common API to modeling and analysis functions*.  
80 <https://doi.org/10.32614/CRAN.package.parsnip>

- 81 Kuhn, M., Wickham, H., & Hvitfeldt, E. (2025). *Recipes: Preprocessing and feature engineering*  
82 *steps for modeling*. <https://doi.org/10.32614/CRAN.package.recipes>
- 83 Murray, J., Hahn, P. R., & Carvalho, C. (2023). *Bcf: Causal inference using bayesian causal*  
84 *forests*. <https://doi.org/10.32614/CRAN.package.bcf>
- 85 Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H.,  
86 Brown, T., Crespo-Tenorio, A., Dimmery, D., & others. (2023). Like-minded sources on  
87 facebook are prevalent but not polarizing. *Nature*, 620(7972), 137–144.
- 88 Tibshirani, J., Athey, S., & Wager, S. (2024). *Grf: Generalized random forests*. <https://doi.org/10.32614/CRAN.package.grf>  
89
- 90 Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical*  
91 *Applications in Genetics and Molecular Biology*, 6(1). [https://doi.org/10.2202/1544-6115.](https://doi.org/10.2202/1544-6115.1309)  
92 1309
- 93 Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>  
94
- 95 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund,  
96 G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M.,  
97 Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome  
98 to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. [https://doi.org/10.21105/](https://doi.org/10.21105/joss.01686)  
99 [joss.01686](https://doi.org/10.21105/joss.01686)

DRAFT