

# CHARLES J. WOLOCK

cwolock@upenn.edu ♦ [cwolock.github.io](https://cwolock.github.io)

Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania  
417 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104

## EDUCATION

---

### University of Washington, Seattle, WA

Ph.D., Biostatistics

Advisors: Marco Carone, Ph.D. and Noah Simon, Ph.D.

*Sept. 2018 - Aug. 2023*

### Harvard University

B.A., Organismic and Evolutionary Biology

Language citation, Spanish

*Summa cum laude*, Highest Honors

*Sept. 2011 - May 2015*

## PROFESSIONAL POSITIONS

---

### Postdoctoral Researcher

Department of Biostatistics, Epidemiology and Informatics

University of Pennsylvania

Supervisor: Yong Chen, Ph.D.

*Sept. 2023 - present*

### Research Staff Associate

Institute for Genomic Medicine

Columbia University

Supervisors: Andrew S. Allen, Ph.D. and David B. Goldstein, Ph.D.

*Nov. 2016 - July 2018*

## TEACHING EXPERIENCE

---

### Instructor of Record

*University of Washington*

BIOST 311: Regression Methods in the Health Sciences

*Mar. 2022 - June 2022*

### Teaching Assistant

*University of Washington*

BIOST 511: Medical Biometry I (Lead TA)

*Sept. 2020 - Dec. 2020*

BIOST 310: Biostatistics for the Health Sciences (Lead TA)

*Sept. 2021 - Dec. 2021*

### Research Mentor

*Fred Hutchinson Cancer Center*

Pathways Undergraduate Research Program

*June 2021 - Aug. 2021*

*University of Washington*

Directed Reading Program: Introduction to Prediction

*Jan. 2023 - Mar. 2023*

### Guest Lecturer

*University of Washington*

BIOST 561: Computational Skills for Biostatistics

*May 2023*

## AWARDS, HONORS, FELLOWSHIPS

---

### University of Washington

NSF Graduate Research Fellowship	Sept. 2020 - Sept. 2023
Thomas R. Fleming Excellence in Biostatistics Award	June 2023
UW Biostatistics highest honor for Ph.D. students	
Donovan J. Thompson Award	Oct. 2020
Best combined performance on the Ph.D. Applied and Theory qualifying exams	
NIH T32 Statistical Genetics Training Grant	Sept. 2018 - Sept. 2020

### Harvard University

Phi Beta Kappa	May 2015
Herchel Smith Research Fellowship	June 2014 - Aug. 2014
Microbial Sciences Initiative Research Fellowship	June 2014 - Aug. 2014
John Harvard Scholar	May 2013

## PROFESSIONAL SERVICE

---

### Manuscript Reviewer

Bayesian Analysis  
Biometrics  
Biostatistics  
Journal of Machine Learning Research  
Journal of the Royal Statistical Society: Series B  
Scientific Reports  
Statistics and Computing  
Statistics in Biopharmaceutical Research

### Conference Session Chair

Next Generation Methods for Evidence Synthesis and Reproducibility (*JSM, Portland, Aug. 2024*)  
Heterogeneous Treatment Effects (*ACIC, Seattle, May 2024*)

## UNIVERSITY SERVICE

---

### University of Washington, Department of Biostatistics

Faculty Search Committee	Sept. 2022 - Mar. 2023
Student Seminar Coordinator	Sept. 2020 - June 2022
Peer mentor	June 2019 - June 2022
Admissions Committee	Sept. 2021 - Mar. 2022
Education Policy and Teaching Evaluation Committee	Sept. 2020 - Sept. 2021
Equity, Diversity, and Inclusion Committee	Sept. 2019 - June 2021

## BIBLIOGRAPHY

---

(\* equal contribution)

### Refereed Journal Publications

1. Sperling EA, **Wolock CJ**, Morgan AS, Gill BC, Kunzmann M, Halverson GP, Macdonald FA, Knoll AH, Johnston DT. (2015). Statistical analysis of iron geochemical data suggests limited late Proterozoic oxygenation. *Nature* 523: 451–454.

2. Raghavan NS, Brickman AM, Andrew H, Manly JJ, Schupf N, Lantigua R, The Alzheimer's Disease Sequencing Project, **Wollock CJ**, Kamalakaran S, Petrovski S, Tosto G, Vardarajan BN, Goldstein DB, Mayeux R. (2018). Whole exome sequencing in 20,197 individuals identifies ultra-rare SORL1 loss-of-function variants in late-onset Alzheimer's disease. *Annals of Clinical and Translational Neurology* 5(7): 832–842.
3. Bittleston LS, **Wollock CJ**, Bakhtiar EY, Chan XY, Chan KG, Pierce NE, Pringle A. (2018). Convergence between the microcosms of Southeast Asian and North American pitcher plants. *eLife* 7.
4. Hayeck TJ, Stong N, **Wollock CJ**, Copeland B, Kamalakaran S, Goldstein DB, Allen AS. (2019). Improved Pathogenic Variant Localization using a Hierarchical Model of Sub-regional Intolerance. *American Journal of Human Genetics* 104(2): 299–309.
5. **Wollock CJ**, Stong N, Ma F, Nagasaki T, Lee W, Tsang SH, Kamalakaran S, Goldstein DB, Allikmets R. (2019). A case-control collapsing analysis identifies retinal dystrophy genes associated with ophthalmic disease in patients with no pathogenic ABCA4 mutations. *Genetics in Medicine* 21: 2336–2344.
6. Gelfman S, Dugger SA, Moreno CAM, Ren Z, **Wollock CJ**, Shneider NA, Phatnani H, Cirulli ET, Lasseigne BN, Harris T, Maniatis T, Rouleau GA, Brown RH, Gitler AD, Myers RM, Petrovski S, Allen AS, Harms MB, Goldstein DB. (2019). A new approach for rare variation collapsing on functional protein domains implicates specific genic regions in ALS. *Genome Research* 29(5): 809–818.
7. Cameron-Christie S, **Wollock CJ**, Groopman E, Petrovski S, Kamalakaran S, Povysil G, Zhang M, Fleckner J, Mar. RE, Gelfman S, Marasa M, Li Y, Sanna-Cherchi S, Kiryluk K, Allen AS, Fellström B, Haefliger C, Platt A, Goldstein DB, Gharavi A. (2019). Exome-based rare-variant analyses in chronic kidney disease. *Journal of the American Society of Nephrology* 30(6): 1109–1122.
8. Ma CJ, **Wollock CJ**, Stong N, Nagasaki T, Lee W, Goldstein DB, Allikmets R. (2019). Case-control collapsing analysis identifies genes mimicking Stargardt/ABCA4 disease. *Investigative Ophthalmology & Visual Science* 60(9): 2935–2935.
9. Eade K, Gantner ML, Hostyk JA, Nagasaki T, Giles S, Harkins-Perry S, Fallon R, Baldini M, Scheppke L, Dorrell MI, Cai C, Baugh EH, **Wollock CJ**, Wallace M, Berlow RB, Goldstein DB, Metallo CM, Friedlander M, Allikmets R. (2021). Serine biosynthesis defect due to haploinsufficiency of phosphoglycerate dehydrogenase (PHDGH) causes retinal disease. *Nature Metabolism* 3(3): 366–377.
10. Bansal A, Heagerty PJ, Inoue LYT, Veenstra DL, **Wollock CJ**, Basu A. (2021). A Value of Information Framework for Personalizing the Timing of Surveillance Testing. *Medical Decision Making* 42(4): 474–486.
11. Wasser SK, **Wollock CJ**, Kuhner MK, Brown JE, Morris C, Horowitz R, Wong A, Fernandez CJ, Otiende MY, Hoareau Y, Kaliszewska ZA, Jeon E, Han KL, Weir BS. (2022). Elephant genotypes reveal the size and connectivity of transnational ivory traffickers. *Nature Human Behaviour* 6(3): 371–382.
12. Heil J, **Wollock CJ**, Pierce NE, Pringle A, Bittleston LS. (2022). Pitcher plant-associated microbial communities differ primarily by host species across a longitudinal gradient. *Environmental Microbiology* 24(8): 3500–3516.
13. Alfieri J, Coble MD, Conroy C, Dahl A, Hares D, Weir BS, **Wollock CJ**, Zhao E, Kingston H, Zolandz T. (2022). A new implementation of a semi-continuous method for DNA mixture interpretation. *Forensic Science International: Reports* 6.
14. Chen YT, Williamson BD, Okonek T, **Wollock CJ**, Spieker AJ, Hee Wai TY, Hughes JP, Emerson SS, Willis AD. (2022). rigr: Regression, Inference, and General Data Analysis Tools in R. *The Journal of*

*Open Source Software* 7(80): 4847.

15. Kohn M, **Wolock CJ**, Poulson I, Fernando N. (2023). A meta-analysis of outcomes of patients with chronic hepatitis C vs. patients without chronic hepatitis C undergoing total hip or total knee arthroplasty. *EFORT Open Reviews* 8(4): 180–188.
16. Bittleston LS\*, **Wolock CJ\***, Maeda J, Infante V, Ané JM, Pierce NE, Pringle, A. (2023). Carnivorous *Nepenthes* pitchers with less acidic fluid house nitrogen-fixing bacteria. *Applied and Environmental Microbiology* 89(7): e00812-23.
17. **Wolock CJ**, Gilbert PB, Simon N, Carone M. (2024). A framework for leveraging machine learning tools to estimate personalized survival curves. *Journal of Computational and Graphical Statistics* 33(3): 1098–1108.
18. Kennedy CJ, Kearns JC, Geraci JC, Gildea SM, Hwang IH, King AJ, Liu H, Luedtke A, Marx BP, Panini S, Petukhova MV, Sampson NA, Smoller JW, **Wolock CJ**, Zainal HN, Stein MB, Ursano RJ, Kessler RC. (2024). Predicting Suicides Among US Army Soldiers After Leaving Active Service. *JAMA Psychiatry*.
19. Wu Q, Pajor NM, Lu Y, **Wolock CJ**, Tong J, Lorman V, Johnson KB, Moore JH, Forrest CB, Asch DA, Chen Y. (2024). A latent transfer learning method for estimating hospital-specific post-acute healthcare demands following SARS-CoV-2 infection. *Patterns* 5(11): 101079.
20. **Wolock CJ**, Gilbert PB, Simon N, Carone M. (2024). Assessing variable importance in survival analysis using machine learning. *Biometrika*.
21. **Wolock CJ**, Williamson BD, Shortreed SM, Simon GE, Coleman KJ, Yeargans R, Ahmedani BK, Daida Y, Lynch FL, Rossom RC, Ziebell RA, Cruz M, Wellman RD, Coley RY. (2024). Importance of variables from different time frames for predicting self-harm using health system data. *Journal of Biomedical Informatics* 160: 104750.

#### Under Revision/Under Review/Preprint

1. **Wolock CJ\***, Jacob S\*, Weil AA, Bennett JC, Elias-Warren A, O’Hanlon J, Kenny A, Jewell NP, Rotnitzky A, Chu HY, Carone M. (2024+). Investigating symptom duration using current status data: a case study of post-acute COVID-19 syndrome. [arXiv:2407.04214](https://arxiv.org/abs/2407.04214). Invited revision, *Epidemiology*.
2. Kenny A, **Wolock CJ**. (2024+). SimEngine: A Modular Framework for Statistical Simulations in R. [arXiv:2403.05698](https://arxiv.org/abs/2403.05698). Under review, *Journal of Statistical Software*.
3. Zhang D, Zhang B, Wang H, Lu Y, **Wolock CJ**, Wang L, Schuemie M, Hripcsak G, Chen Y. (2024+) Negative control-calibrated difference-in-difference analyses: addressing unmeasured confounding in real-world data with application to quantifying the impact of pandemic on racial/ethnic differences. Under review, *The Lancet Digital Health*.
4. Wollum A, Ralph L, Grossman D, **Wolock CJ**, Moseson H. (2024+) Evaluating a double list experiment to measure the population proportion of people who have attempted a self-managed abortion in the United States: analysis of a nationally representative cross-sectional survey in 2017. Under review, *American Journal of Epidemiology*.
5. Morenz ER\*, **Wolock CJ\***, Carone M. (2024+) Debiased machine learning for survival functionals based on left-truncated right-censored data. [arXiv:2411.09017](https://arxiv.org/abs/2411.09017).

#### In Preparation

1. **Wolock CJ\***, Yan J\*, Ning Y, Chen Y. (2024+) Transfer learning for model-free variable importance.

2. **Wolock CJ**, Zhang B, Zhang D, Wu Q, Chen Y. (2024+) Estimating causal effects from electronic health records data with underreported exposure.
3. Hoffman KL, Marshall G, Frivold C, . . . , **Wolock CJ**, . . . , Gilbert PB, Naleway AL, Chu HY, Carone M. (2024+) Correlates of Protection Against SARS-CoV-2 Infection in Children.

## SOFTWARE

---

<a href="#">survML</a>	Tools for flexible survival analysis using machine learning ( <code>cre</code> , <code>aut</code> )
<a href="#">SimEngine</a>	Framework for reproducible statistical simulations in R ( <code>aut</code> )
<a href="#">rigr</a>	Regression, inference, and general data analysis tools for R ( <code>aut</code> )
<a href="#">vimp</a>	Inference on algorithm-agnostic variable importance ( <code>ctb</code> )

## PRESENTATIONS

---

(\*invited)

1. Nonparametric approaches to assessing variable importance using health data.  
*\*Department of Biostatistics, Columbia University, Jan. 2025.*  
*\*Department of Biostatistics and Computational Biology, University of Rochester, Dec. 2024.*  
*\*Division of Biostatistics and Health Data Science, University of Minnesota, Nov. 2024.*
2. Estimating causal effects from electronic health records data with underreported exposure.  
*JSM, Portland, Aug. 2024.*
3. Nonparametric variable importance for time-to-event outcomes with application to prediction of HIV infection.  
*\*JSM, Toronto, Aug. 2023.*  
*ENAR, Baltimore, Mar. 2024.*
4. Flexible estimation of the conditional survival function via observable regression models.  
*ENAR, Nashville, Mar. 2023.*  
*JSM, Washington, D.C., Aug. 2022.*  
*WNAR, virtual, June 2022.*

## PROFESSIONAL AFFILIATIONS

---

American Statistical Association  
 International Biometric Society (ENAR)

## SKILLS

---

<b>Programming</b>	Python, R, Bash, SQL
<b>Other applications</b>	$\text{\LaTeX}$ , Git
<b>Languages</b>	English (native), Spanish (proficient)