

Philippe Boileau

✉ philippe_boileau@berkeley.edu • 🌐 www.pboileau.ca
📧 PhilBoileau • 📞 0000-0002-4850-2507 • 📄 KTVhL_IAAAAJ

Education

- University of California, Berkeley** **Berkeley, CA**
PhD in Biostatistics *2020–2023*
Thesis: Nonparametric Methods for High-Dimensional Data Analysis
Minor: Computational and Genomic Biology
Committee: Sandrine Dudoit, Mark van der Laan, Alan Hubbard, Andres Cardenas
- University of California, Berkeley** **Berkeley, CA**
MA in Biostatistics *2018–2020*
Thesis: Exploring High-Dimensional Biological Data with Sparse Contrastive
Principal Component Analysis
Committee: Sandrine Dudoit, Elizabeth Purdom, Haiyan Huang
- Concordia University** **Montréal, QC**
BSc in Honours Statistics *2016–2018*
Thesis: Modelling the Obesity Epidemic with Networks
Advisors: Lisa Kakinami, Lea Popovic

Selected Experience

- Associate, Healthcare Economics and Outcomes Research** **Montréal, QC**
Analysis Group *2023–Pres.*
Provide data-driven strategies for clients in the life sciences industry across all phases of product development and commercialization, with a particular emphasis on the management, design, and conduct of pharmacoepidemiology studies using electronic health record and administrative claims data.
- Data Science Intern** **South San Francisco, CA**
Genentech and Roche Canada *2021–2023*
Developed assumption-lean statistical inference methods for predictive biomarker discovery, benchmarked them against existing procedures, and applied them to historical clinical oncology trial data. Implemented visualization functions in R for effective summaries of clinical trial data. Compiled and compared designs for phase I oncology trials and produced an educational document for internal use. Performed a comprehensive simulation study of mixed models for repeated measures software.
- Graduate Student Researcher** **Berkeley, CA**
University of California, Berkeley Superfund *2020–2022*
Supported the organization's environmental health scientists and epidemiologists by analyzing epigenetic data, developing statistical analysis pipelines, consulting on experimental design, and preparing manuscripts.
- Statistician** **San Francisco, CA**
Ha Lab, University of California, San Francisco *2020–2021*
Designed experiments, implemented bioinformatic pipelines, and applied biostatistical methods for the analysis of bulk transcriptomic data of salivary gland tissue collected from adenoid cystic carcinoma patients.
- Undergraduate Research Assistant** **Montreal, QC**
Canadian Longitudinal Study on Aging *2015–2018*
Worked with biostatisticians, data curators, and data access managers on various projects, including the development of an R package to facilitate analysis of proprietary data and the creation of a public-use dataset.

Awards and Honours

- 2022** Student & Early Career Travel Award, Symposium on Data Science & Statistics
Environmental and Molecular Mutagenesis Editor's Choice Award for "In vitro relationships of galactic cosmic radiation and epigenetic clocks in human bronchial epithelial cells" by Nwanaji-Enwerem et al.
- 2021** Alexander Graham Bell Canada Graduate Scholarship – Doctoral (Declined: not tenable outside of Canada), Natural Sciences and Engineering Research Council of Canada
Postgraduate Scholarship – Doctoral, Natural Sciences and Engineering Research Council of Canada
Bourse de doctorat en recherche, Fonds de recherche du Québec, Nature et technologies
Extraordinary Teaching in Extraordinary Times Award: DATA 8, Summer 2020, University of California, Berkeley
- 2020** Biostatistics Block Grant and Non-Resident Student Tuition Award, University of California, Berkeley
- 2019** Bourse de maîtrise en recherche, Fonds de recherche du Québec, Nature et technologies
- 2018** Biostatistics Block Grant and Non-Resident Student Tuition Award, University of California, Berkeley
Best Poster Presentation, Québec Society for Lipid, Nutrition and Metabolism Scientific Meeting
Best Undergraduate Poster Presentation, Canadian Statistics Student Conference
- 2017** Undergraduate Summer Scholarship, Institut des sciences mathématiques
Alumni Association Scholarship, Concordia University
Faculty of Arts and Science Scholar, Concordia University

Publications and Preprints

* indicates shared first-authorship

Statistical Methodology

- [1] **P. Boileau**, N. Leng, and S. Dudoit. Guidance on individualized treatment rule estimation in high dimensions. *arXiv; response invited at The International Journal of Biostatistics*, 2024+. URL <https://arxiv.org/abs/2306.16402v1>
- [2] **P. Boileau**, N. Leng, N. S. Hejazi, M. van der Laan, and S. Dudoit. A nonparametric framework for treatment effect modifier discovery in high dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkae084, 2024. ISSN 1369-7412. doi: 10.1093/jrsssb/qkae084. URL <https://doi.org/10.1093/jrsssb/qkae084>
- [3] N. S. Hejazi, **P. Boileau**, M. J. van der Laan, and A. E. Hubbard. A generalization of moderated statistics to data adaptive semiparametric estimation in high-dimensional biology. *Statistical*

Methods in Medical Research, 32(3):539–554, 2023. URL <https://journals.sagepub.com/doi/full/10.1177/09622802221146313>

- [4] **P. Boileau**, N. S. Hejazi, M. J. van der Laan, and S. Dudoit. Cross-validated loss-based covariance matrix estimator selection in high dimensions. *Journal of Computational and Graphical Statistics*, 0(0):1–12, 2022. doi: 10.1080/10618600.2022.2110883. URL <https://doi.org/10.1080/10618600.2022.2110883>
- [5] **P. Boileau**, L. Kakinami, T. A. Barnett, M. Henderson, and L. Popovic. Heatmaps and consensus clustering for ego network exploration [version 1; peer review: awaiting peer review]. *F1000Research*, 11(771), 2022. doi: 10.12688/f1000research.108964.1
- [6] **P. Boileau**, N. T. Qi, M. J. van der Laan, S. Dudoit, and N. Leng. A flexible approach for predictive biomarker discovery. *Biostatistics*, 07 2022. ISSN 1465-4644. doi: 10.1093/biostatistics/kxac029. URL <https://doi.org/10.1093/biostatistics/kxac029>. kxac029
- [7] **P. Boileau**, N. S. Hejazi, and S. Dudoit. Exploring high-dimensional biological data with sparse contrastive principal component analysis. *Bioinformatics*, 36(11):3422–3430, 03 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa176. URL <https://doi.org/10.1093/bioinformatics/btaa176>

Applications

- [8] J. O. Humtsoe, H.-S. Kim, L. Jones, J. Cevallos, **P. Boileau**, F. Kuo, L. G. T. Morris, and P. Ha. Development and characterization of MYB-NFIB fusion expression in adenoid cystic carcinoma. *Cancers*, 14(9), 2022
- [9] A. K. Bozack, **P. Boileau**, A. E. Hubbard, F. C. M. Sillé, C. Ferreccio, C. M. Steinmaus, M. T. Smith, and A. Cardenas. The impact of prenatal and early life arsenic exposure on epigenetic age acceleration among adults in northern chile. *Environmental Epigenetics*, 06 2022. ISSN 2058-5888. doi: 10.1093/eep/dvac014. URL <https://doi.org/10.1093/eep/dvac014>. dvac014
- [10] J. C. Nwanaji-Enwerem, **P. Boileau**, J. M. Galazka, and A. Cardenas. In vitro relationships of galactic cosmic radiation and epigenetic clocks in human bronchial epithelial cells. *Environmental and Molecular Mutagenesis*, 63(4):184–189, 2022
- [11] A. K. Bozack, **P. Boileau**, L. Wei, A. E. Hubbard, F. C. M. Sillé, C. Ferreccio, J. Acevedo, L. Hou, V. Ilievski, C. M. Steinmaus, M. T. Smith, A. Navas-Acien, M. V. Gamble, and A. Cardenas. Exposure to arsenic at different life-stages and DNA methylation meta-analysis in buccal cells and leukocytes. *Environmental Health*, 20(1):79, 2021

Refereed Software

- [12] J. Duncan, T. Tang, C. F. Elliott, **P. Boileau**, and B. Yu. simChef: High-quality data science simulations in R. *Journal of Open Source Software*, 9(95):6156, 2024. doi: 10.21105/joss.06156. URL <https://doi.org/10.21105/joss.06156>
- [13] **P. Boileau**, N. S. Hejazi, B. Collica, M. J. van der Laan, and S. Dudoit. cvCovEst: Cross-validated covariance matrix estimator selection and evaluation in R. *Journal of Open Source Software*, 6(63):3273, 2021. doi: 10.21105/joss.03273. URL <https://doi.org/10.21105/joss.03273>

- [14] **P. Boileau**, N. S. Hejazi, and S. Dudoit. scPCA: A toolbox for sparse contrastive principal component analysis in R. *Journal of Open Source Software*, 5(46):2079, 2020. doi: 10.21105/joss.02079. URL <https://doi.org/10.21105/joss.02079>

Non-Refereed Software

- [15] **P. Boileau**. *unihtee: Univariate heterogeneous treatment effect estimation*, 2023. URL <https://github.com/insightsengineering/unihtee>. R package version 0.1.0
- [16] **P. Boileau**. *uniCATE: Univariate Conditional Average Treatment Effect Estimation*, 2022. URL <https://github.com/insightsengineering/uniCATE>. R package version 0.4.0
- [17] **P. Boileau**. *neatmaps: Heatmaps for Multiple Network Data*, 2019. URL <https://CRAN.R-project.org/package=neatmaps>. R package version 2.1.0

Book Chapters

- [18] A. Lun and **P. Boileau**. *Case Study: Messmer Human ESC*, chapter 13. Bioconductor, 2020. URL <http://bioconductor.org/books/3.16/OSCA.workflows/messmer-hesc.html>

Technical Reports

- [19] **P. Boileau**, L. Popovic, T. A. Barnett, M. Henderson, and L. Kakinami. Exploration of multi-network data with heatmaps. In *ISM Internship Proceedings, Summer 2017*, 2017. URL <http://ism.uqam.ca/~ism/publications/proceedings/>

Theses

- [20] **P. Boileau**. *Nonparametric Methods for High-Dimensional Data Analysis*. PhD thesis, University of California, Berkeley, Berkeley, CA, 2023
- [21] **P. Boileau**. Exploring high-dimensional biological data with sparse contrastive principal component analysis. Master's thesis, University of California, Berkeley, Berkeley, CA, May 2020
- [22] **P. Boileau**. Modelling the obesity epidemic with networks. Bachelor of honour's science thesis, Concordia University, Montréal, QC, May 2018

In Preparation

- [1] **P. Boileau***, N. S. Hejazi*, I. Malenica*, P. G. Gilber, S. Dudoit, and M. J. van der Laan. Identifying direct causal effects under unmeasured confounding, applied to vaccine efficacy evaluation. 2024+
- [2] **P. Boileau**, A. Cardenas, N. Ghildayal, M. Doyon, P. Perron, L. Bouchard, and M.-F. Hivert. Cell-type specific DNA methylation signatures of prenatal smoking on the human placenta. 2024+

Presentations

* indicates shared first-authorship

Contributed Talks

- [1] G. Duran-Pacheco, J. Dedic, and **P. Boileau**. Comparing R libraries with SAS's PROC MIXED for the analysis of longitudinal continuous endpoints using MMRM. 5th Conference of the Central European Network, September 2023
- [2] **P. Boileau**, N. Leng, M. J. van der Laan, and S. Dudoit. A nonparametric framework for treatment effect modifier discovery in high dimensions. American Causal Inference Conference, Austin, TX, May 2023. URL https://github.com/PhilBoileau/ACIC-2023_unihtee
- [3] **P. Boileau**, N. T. Qi, M. J. van der Laan, S. Dudoit, and N. Leng. unicate: Flexible predictive biomarker discovery. UC Berkeley Center for Computational Biology Retreat 2022, November 2022. URL <https://github.com/PhilBoileau/UCB-CCB-retreat-2022-talk>
- [4] **P. Boileau**, N. S. Hejazi, and S. Dudoit. Sparse contrastive principal component analysis. Joint Statistical Meetings, Washington, DC (declined), August 2022
- [5] **P. Boileau**, N. T. Qi, M. J. van der Laan, S. Dudoit, and N. Leng. unicate: Flexible predictive biomarker discovery. Symposium on Data Science and Statistics, Pittsburgh, PA, June 2022. URL https://github.com/PhilBoileau/SDSS-2022_unicate-talk
- [6] **P. Boileau**, N. S. Hejazi, M. J. van der Laan, and S. Dudoit. Cross-validated covariance matrix estimator selection in high dimensions. Statistics 2021 Canada, online, July 2021. URL <https://github.com/PhilBoileau/stats-2021-canada-presentation>
- [7] **P. Boileau**, N. S. Hejazi, and S. Dudoit. Sparse contrastive principal component analysis for exploring high-dimensional biological data. Bioconductor 2020, online, July 2020. URL <https://github.com/PhilBoileau/scPCA-presentation-bioc2020>

Contributed Posters

- [1] D. R. Feldman, K. Habucky, P. Gagnon-Sanschagrín, J. Maitland, **P. Boileau**, K. Yokoji, A. Guerin, and G. Joseph. Real-world evidence of overall survival and treatment patterns of patients with testicular germ cell tumors receiving palliative chemotherapy in the united states. 2024 American Society of Clinical Oncology Annual Meeting, Chicago, IL, June 2024
- [2] **P. Boileau**, N. Q. Ting, M. J. van der Laan, S. Dudoit, and N. Leng. A flexible approach for predictive biomarker discovery. Bay Area Biotech-Pharma and Statistics Workshop: Resilience and Reinvention, November 2022
- [3] A. K. Bozack, **P. Boileau**, A. E. Hubbard, F. C. M. Sillé, C. Ferreccio, C. M. Steinmaus, M. T. Smith, and A. Cardenas. Evidence of epigenetic age acceleration among adults with prenatal and early life arsenic exposure in Nothern Chile. Annual Conference of the International Society for Environmental Epidemiology, September 2022

- [4] A. Cardenas, **P. Boileau**, N. Ghildayal, M. Doyon, P. Perron, L. Bouchard, and M.-F. Hivert. Cell-type specific DNA methylation signature of prenatal smoking on human placenta. DOHaD World Congress, Vancouver, Canada, August 2022
- [5] J. P. Duncan, T. Tang, C. F. Elliot, **P. Boileau**, and B. Yu. simChef: An intuitive framework for reliable simulation studies in R. useR! 2022: The R User Conference, online, June 2022. URL <https://github.com/jpdunc23/simChef-poster-useR-2022>
- [6] **P. Boileau***, N. S. Hejazi*, I. Malenica*, P. G. Gilber, S. Dudoit, and M. J. van der Laan. Identifying direct causal effects under unmeasured confounding. American Causal Inference Conference, Berkeley, CA, May 2022. URL https://github.com/PhilBoileau/ACIC-2022_ndeconf-poster
- [7] **P. Boileau**, N. Q. Ting, M. J. van der Laan, S. Dudoit, and N. Leng. A flexible approach for predictive biomarker discovery. American Causal Inference Conference, Berkeley, CA, May 2022. URL https://github.com/PhilBoileau/ACIC-2022_uniCATE-poster
- [8] J. Cevallos, **P. Boileau**, J. Humstoe, L. Jones, and P. Ha. Exploration of the transcriptome of the adenoid cystic carcinoma patientderived xenograf. Virtual American Head and Neck Society 10th International Conference on Head and Neck Cancer, online, July 2021
- [9] **P. Boileau**, N. S. Hejazi, and S. Dudoit. Sparse contrastive principal component analysis. Berkeley Statistical Annual Research Symposium (cancelled), March 2020
- [10] D. Scheel, A. Perkova, P. Daniele, **P. Boileau**, and L. Kakinami. Does survey design information matter? Assessing the impact on population estimates of hypertension in canada. 2018 Annual Meeting, Statistical Society of Canada, Montreal, QC, June 2018
- [11] **P. Boileau**, L. Popovic, T. A. Barnett, M. Henderson, and L. Kakinami. Ego network exploration with heatmaps: A case study on pediatric obesity. Canadian Statistics Student Conference, Montreal, QC, June 2018
- [12] **P. Boileau**, L. Popovic, T. A. Barnett, M. Henderson, and L. Kakinami. L'analyse de réseaux sociaux avec cartes thermiques, une étude de cas avec l'obésité pédiatrique. Québec society for lipid, nutrition and metabolism scientific meeting, Magog, QC, February 2018

Other Talks

- [1] **P. Boileau**. Simulation studies as integration tests for statistical software: A case study with the mrm R package. Genentech Statistics Community Forum, April 2023
- [2] **P. Boileau**, N. Q. Ting, M. J. van der Laan, S. Dudoit, and N. Leng. UniCATE: A flexible approach for predictive biomarker discovery. Genentech Statistics Community Forum, May 2022
- [3] **P. Boileau**, J. Zhou, and J. Fridlyand. Model-based, model-guided or no model at all: Pragmatic evaluation of dose-finding methods. Genentech Statistics Community Forum, May 2022

Workshops

- [1] D. Kelkhoff, **P. Boileau**, and D. S. Bove. Good software engineering practices for R packages. McGill Initiative in Computational Medicine Workshop Series, 2023. URL <https://openpharma.github.io/workshop-r-swe-mtl/>

Service

Ad Hoc Referee Annals of Applied Statistics, Bioinformatics, GigaScience, IEEE Transactions on Signal Processing, Journal of Causal Inference, Journal of Computational and Graphical Statistics, Journal of Machine Learning Research, PLOS One, Proceedings of the National Academy of Sciences

Committees Biostatistics Diversity, Equity, Inclusion, and Belonging Committee, University of California, Berkeley

Teaching Experience

University of California, Berkeley

Instructor: DATA 8: The Foundations of Data Science *Summer 2020*

Teaching Assistant: PB HLTH 241: Statistical Analysis of Categorical Data *Spring 2020*

Teaching Assistant: DATA 100: Principles and Techniques of Data Science *Fall 2019*

Teaching Assistant: DATA 100: Principles and Techniques of Data Science *Spring 2019*

Teaching Assistant: PB HLTH 142: Introduction to Statistics in Public Health *Fall 2018*

Mentoring

University of California, Berkeley

Jonathan Lin: Undergraduate, Statistics Research Apprenticeship Program *2022-2023*

Brian Collica: MA student, STAT 197: Field Study in Statistics *2020-2021*

Jamarcus Liu: Undergraduate, Statistics Research Apprenticeship Program *Fall 2020*

Star Li: Undergraduate, Statistics Research Apprenticeship Program *Spring 2020*

David Liu: Undergraduate, Statistics Research Apprenticeship Program *Spring 2020*

Computing

Languages: R, Bash, \LaTeX , SAS, SQL

Tools: Git, GitHub/GitLab, Emacs, Vim, RStudio

Languages

English: Native speaker

French: Native speaker