Philippe Boileau

Curriculum Vitae

Updated March 2025

philippe.boileau@mcgill.ca
pboileau.ca

Education

2020–2023 **PhD in Biostatistics**, *University of California, Berkeley*, Berkeley, CA
Thesis: Nonparametric Methods for High-Dimensional Data Analysis
Minor: Computational and Genomic Biology
Committee: Sandrine Dudoit, Mark van der Laan, Alan Hubbard, Andres Cardenas

2018–2020 **MA in Biostatistics**, *University of California, Berkeley*, Berkeley, CA
Thesis: Exploring High-Dimensional Biological Data with Sparse Contrastive Principal Component Analysis
Committee: Sandrine Dudoit, Elizabeth Purdom, Haiyan Huang

2016–2018 **BSc in Honours Statistics**, *Concordia University*, Montréal, QC
Thesis: Modelling the Obesity Epidemic with Networks
Advisors: Lisa Kakinami, Lea Popovic

Appointments

2025–Pres. **Assistant Professor of Biostatistics**, *McGill University*, Montréal, QC
Departments of Epidemiology, Biostatistics and Occupational Health and of Medicine

2025-Pres. **Junior Scientist**, *Research Institute of the McGill University Health Centre*, Montréal, QC
Centre for Outcomes Research and Evaluation

Selected Experience

2023–2025 **Associate**, *Analysis Group*, Montréal, QC
Healthcare Economics and Outcomes Research Practice

2021–2023 **Data Science Intern**, *Genentech and Roche Canada*, South San Francisco, CA
Product Development Data Science

2020–2022 **Graduate Student Researcher**, *University of California, Berkeley*, Berkeley, CA
UC Berkeley Superfund Research Program

2020–2021 **Statistician**, *University of California, San Francisco*, San Francisco, CA
Ha Lab, Department of Otolaryngology – Head and Neck Surgery

2015-2018 **Undergraduate Research Assistant**, *Research Institute of the McGill University Health Centre*, Montréal, QC
Canadian Longitudinal Study on Aging

Awards and Honours

Amounts in Canadian dollars unless otherwise specified

2025 Postdoctoral Research Funding, IVADO ($200,000 [declined])

2022 Student & Early Career Travel Award, Symposium on Data Science & Statistics ($750 [USD])

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, Natural Sciences and Engineering Research Council of Canada ($105,000 [declined])

Postgraduate Scholarship – Doctoral, Natural Sciences and Engineering Research Council of Canada ($63,000)

Bourse de doctorat en recherche, Fonds de recherche du Québec, Nature et technologies ($90,000)

Extraordinary Teaching in Extraordinary Times Award: DATA 8, Summer 2020, University of California, Berkeley ($2,000 [USD])

2020 Biostatistics Block Grant and Non-Resident Student Tuition Award, University of California, Berkeley ($25,000 [USD])

2019 Bourse de maîtrise en recherche, Fonds de recherche du Québec, Nature et technologies ($38,000)

Biostatistics Block Grant and Non-Resident Student Tuition Award, University of California, Berkeley ($13,000 [USD])

2018 Biostatistics Block Grant and Non-Resident Student Tuition Award, University of California, Berkeley ($18,500 [USD])

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 ($2,000)

Faculty of Arts and Science Scholar, Concordia University

Publications

** indicates shared first-authorship; trainees' names underlined*

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*, 2025+. 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**, 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

[6] **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

[7] 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

[8] 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

[9] 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

[10] 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

[11] 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

[12] **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

[13] **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

[14] **P. Boileau**. *unihtee: Univariate heterogeneous treatment effect estimation*, 2023. URL https://github.com/insightsengineering/unihtee. R package version 0.1.0

[15] **P. Boileau**. *uniCATE: Univariate Conditional Average Treatment Effect Estimation*, 2022. URL https://github.com/insightsengineering/uniCATE. R package version 0.4.0

[16] **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

[17] 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

[18] **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

Theses

[19] **P. Boileau**. *Nonparametric Methods for High-Dimensional Data Analysis*. PhD thesis, University of California, Berkeley, Berkeley, CA, 2023

[20] **P. Boileau**. Exploring high-dimensional biological data with sparse contrastive principal component analysis. Master's thesis, University of California, Berkeley, Berkeley, CA, May 2020

[21] **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. 2025+

[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. 2025+

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, P. Gagnon-Sanschagrin, J. Maitland, **P. Boileau**, K. Yokoji, A. Guerin, V. Guan, and G. Joseph. Time to real-world progression among patients with relapsed/refractory testicular germ cell tumors undergoing palliative chemotherapy in the united states. 2025 ASCO Genitourinary Cancers Symposium, San Francisco, CA, February 2025

[2] D. R. Feldman, K. Habucky, P. Gagnon-Sanschagrin, 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

[3] **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

[4] 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

[5] 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

[6] 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`

[7] **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

[8] **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

[9] 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

[10] **P. Boileau**, N. S. Hejazi, and S. Dudoit. Sparse contrastive principal component analysis. Berkeley Statistical Annual Research Symposisum (cancelled), March 2020

[11] D. Scheel, A. Perkova, P. Danieles, **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

[12] **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

[13] **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 mmrm 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

Peer Review: Ad Hoc Referee

2021–Pres. 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, npj Digital Medicine, PLOS One, Proceedings of the National Academy of Sciences, Scientific Reports

Committee Work

2025–Pres. Steering Committee, Accelerating Clinical Trials – Clinical Trial Unit of the Research Institute of the McGill University Health Centre

2020-2022 Biostatistics Diversity, Equity, Inclusion, and Belonging Committee, University of California, Berkeley

Teaching Experience

University of California, Berkeley

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

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

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

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

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

Formal Mentoring

University of California, Berkeley

2022-2023 Jonathan Lin (undergraduate), Statistics Research Apprenticeship Program

2020-2021 Brian Collica (masters), STAT 197: Field Study in Statistics

2020 Jamarcus Liu (undergraduate), Statistics Research Apprenticeship Program

2020 Star Li (undergraduate), Statistics Research Apprenticeship Program

2020 David Liu (undergraduate), Statistics Research Apprenticeship Program

Computing

Languages R, Bash, LaTeX, SAS, SQL

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

Languages

English Native speaker

French Native speaker