Yi Lian

yi.lian@pennmedicine.upenn.edu

EDUCATION

- McGill University — Montréal, Québec, Canada
  - B.Sc. Pharmacology, Statistics — 2014
  - M.Sc. Epidemiology — 2016
    Supervisor: Samy Suissa, Laurent Azoulay
    Thesis: Phosphodiesterase-5 inhibitors and the risk of melanoma skin cancer
  - Ph.D. Biostatistics — 2023
    Supervisor: Robert W. Platt, Archer Y. Yang
    Thesis: Some new computational methods in high-dimensional statistical learning in biostatistics

RESEARCH INTERESTS

High-dimensional statistics, federated learning, missing data methods, clinical trials, cancer epidemiology, cardio-oncology, pharmacoepidemiology.

RESEARCH APPOINTMENTS

- Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Philadelphia, PA
  - Postdoctoral research fellow in Biostatistics — 2023 – Present
  - Supervisor: Qi Long
- Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital — Montréal, Canada
  - Research Assistant — 2014 – 2022
- The Research Institute of the McGill University Health Centre — Montréal, Canada
  - Research Assistant — 2020 – 2022
- Department of Epidemiology, Biostatistics and Occupational Health — McGill University
  - Intern in Biostatistics — Summer 2013

AWARDS

- First Place Innovation Award
  - American Medical Informatics Association (AMIA) Knowledge Discovery and Data Mining (KDDM) Working Group Meeting — 2024
- Natural Sciences and Engineering Research Council of Canada (NSERC; Canadian equivalent of NSF)
  - NSERC Postgraduate Scholarship – Doctoral (PGS-D) — 2018 - 2020
    Discipline: Statistics and Probability
    Competitive merit-based federal scholarship
    $21,000 per year for two years
    Comparable to the NSF Graduate Research Fellowship
- McGill Initiative in Computational Medicine

  - MICM Graduate Award ($1,350) 2019 – 2020

- Lady Davis Institute, Jewish General Hospital
  - Lady Davis Institute/TD Bank Studentship Award ($10,000 for one year) 2017 – 2018

- Department of Epidemiology and Biostatistics and Occupational Health, McGill University
  - CRM Statistics Lab Award ($500) April 2021
  - Graduate Excellence Award ($15,000) 2016 – 2017
  - Graduate Research Enhancement and Travel Award ($500) April 2016
  - Biostatistics Education Award ($3,000) June 2013

- International Society for Pharmacoepidemiology
  - 2018 ICPE Annual Conference, Montréal August 2017
    Scholarship ($390)
  - 2016 Mid-Year Meeting, Baltimore April 2016
    Awarded student oral presentation (US $465)

- McGill University
  - James McGill Scholarship ($3,000 per year) 2009 – 2011

PUBLICATIONS

- Biotatistics and Machine Learning Methodology

S1 **Lian, Y.**, Long, Q., & Chang, C. (2025). Flexible and adaptive multitask regression for fairness-aware integration of high-dimensional data. Submitted to *Biometrics*.

S2 **Lian, Y.**, Jiang, X., & Long, Q. (2025). Federated multiple imputation for multicenter studies using random forests: robust to model misspecification and easy to apply. In progress.

S3 **Lian, Y.**, Long, Q., Hu, C., (2025). An efficient two-stage basket trial design with interim data-driven pooling to improve power and precision. In progress.

S4 **Lian, Y.**, Jiang, X., & Long, Q. (2025). `D3MI`: an efficient federated imputation method for enhanced analysis of distributed heterogeneous incomplete data. Under review at *Journal of Biomedical Informatics*.

S5 **Lian, Y.**, Long, Q., & Chang, C. (2025) `fairee`: a flexible and adaptive semi-parametric method for fairness-aware integration of heterogeneous high-dimensional data. In progress.

S6 Fan, J., Gu, Y., **Lian, Y.**, Zhang, R. (2025). Adaptive group lasso penalized quantile regression. In progress.

S7 Zhang, R., Fan, J., **Lian, Y.**, & Yan, A. (2025). An efficient algorithm for the weighted elastic net penalized quantile regression. *Communications in Statistics-Simulation and Computation*: 1-27.

S8 Fan, J., Kong, L., **Lian, Y.**, & Zou, H. (2025) Analysis of best subset regression: global algorithm and asymptotic normalities. To be submitted to the *Journal of the American Statistical Association.*

S9 **Lian, Y.***, Wang, G.*, Yang, A. Y., Platt, R. W., Wang, R., Perreault, S., ... & Schnitzer, M. E. (2024). Structured learning in time-dependent Cox models. *Statistics in Medicine*, 43(17), 3164-3183. (*Co-first author)

S10 **Lian, Y.***, Wang, G.*, McGrath, S.*, & Dahabreh, I. (2024). `CausalMetaR`: An R package for performing causally interpretable meta-analyses. *Research Synthesis Methods.* (Accepted for publication). (*Co-first author)

S11 **Lian, Y.**, Jiang, X., & Long, Q. (2024). Federated multiple imputation for variables that are missing not at random in distributed electronic health records. Proceedings at *the American Medical Informatics Association (AMIA) 2024 Annual Symposium.* (First Place Innovation Award by the AMIA Knowledge Discovery and Data Mining (KDDM) Working Group.)

S12 Yang, A. Y., Zhao, Y., **Lian, Y.**, Gu, Y., & Fan, J. (2024). Regularized estimating equations: some new perspectives. In progress

S13 **Lian, Y.**, Yang, A. Y., Gu, Y., Fan, J., Zhao, Y., & Platt, R. W. (2024). Fast computation of high-dimensional estimating equations using Anderson acceleration. (Completed, submission follows completion of S5.)

S14 **Lian, Y.**, Yang, A. Y., Gu, Y., Fan, J., Zhao, Y., & Platt, R. W. (2024). `free`: The R package for flexible regularized estimating equations with high-dimensional covariates. (Completed, submission follows completion of S5.)

S15 **Lian, Y.**, Yang, A. Y., Wang, B., Shi, P., & Platt, R. W. (2023). A Tweedie compound Poisson model in reproducing kernel Hilbert space. *Technometrics*, 1-15.

S16 Moodie, E. E. M., Bian, Z., Coulombe, J., **Lian, Y.**, Yang, A. Y., & Shortreed, S. M. (2022). Variable selection in high dimensions for discrete-outcome individualized treatment rules: A case study in reducing severity of depression symptoms. *Biostatistics*, 2023;, kxad022

S17 Lu, X. H., Liu, A., Fuh, S. C., **Lian, Y.**, Guo, L., Yang, Y., Marelli A., & Li, Y. (2021). Recurrent disease progression networks for modelling risk trajectory of heart failure. *PLoS One*, 16(1), e0245177

- Epidemiology

E1 **Lian, Y.***, Voruganti, T.*, Lu, J., Long, Q., & Mamtani, R. (2025) Survival trends in urothelial cancer before and after ICIs and antibody drug conjugates. *JAMA Network Open*, 8(7):e2519524. (*Co-first author)

E2 Nimgaonkar, V., Kurian, M., **Lian, Y.**, & Mamtani, R. (2025). Real-world prevalence of toxicities associated with enfortumab-vedotin-based treatment in advanced urothelial carcinoma. In progress.

E3 Du, F., Guo, Y., Khosrow-Khavar, F., ..., **Lian, Y.**, ..., and Jiao, T. (2025). The association between RAAS inhibitors and new-onset pancreatic cancer: A target trial emulation. Submitted to the *Annals of Internal Medicine*.

E4 Zhao, Z., Ruzieh, M., **Lian, Y.**, ..., & Jiao, T. (2024). High dimensional propensity score to reduce selection bias and confounding: the association between use of Watchman device and 1-year mortality in a real-world setting. *Circulation: Cardiovascular Quality and Outcomes*, e011188.

E5 Ruzieh, M., Du, F., Foy A. J., Blotner, M., Naccarelli, G. V., **Lian, Y.**, ... & Jiao, T. (2024). Trial emulation of percutaneous left atrial appendage occlusion versus direct oral anticoagulants. Submitted to *Europace*.

E6 **Lian, Y.**, Yin, H., Pollak, M. N., Carrier, S., Platt, R. W., Suissa, S., & Azoulay, L. (2016). Phosphodiesterase type 5 inhibitors and the risk of melanoma skin cancer. *European Urology*, 70(5), 808-815 (Platinum Priority – Sexual Medicine)

- Statistical Consulting

C1 Chen, S. S., Wang, T. Q., Song, W. C., Tang, Z. J., Cao, Z. M., Chen, H. J., **Lian, Y.**, ... (2022) A novel PM sampling and cell exposure strategy based on agar membrane for cytotoxicity study. Available at SSRN 3977371.

C2 Tang, Z., Cao, Z., Guo, X., Chen, H., **Lian, Y.**, Zheng, W., ... (2020). Cytotoxicity and toxicoproteomic analyses of human lung epithelial cells exposed to extracts of atmospheric particulate matters on PTFE filters using acetone and water. *Ecotoxicology and Environmental Safety*, 191, 110223

SOFTWARE PROJECTS

- R package `HDfair`: High-Dimension Fairness-Aware Integration toolbox in R
  - Available upon request – paper under review.
- R package `CausalMetaR`: Causally interpretable meta-analysis in R

  - https://cran.r-project.org/package=CausalMetaR
- R package `free`: Regularized estimating equations
  - Regularized estimating equation solver
  - Efficient C++ Armadillo program with R interface
  - https://cran.r-project.org/package=free
- R package `sox`: Structured learning in time-dependent Cox models
  - Efficient procedures for fitting and cross-validating the structurally-regularized time-dependent Cox models.
  - Efficient C++ program with R interface
  - https://cran.r-project.org/package=sox
- R package `ktweedie`: Tweedie compound Poisson gamma model in reproducing kernel Hilbert space
  - Flexible nonparametric Tweedie model for zero-inflated continuous outcome data
  - Efficient Fortran program with R interface
  - https://cran.r-project.org/package=ktweedie
- R package `intsel`: Automated machine learning tool for screening and identifying interactions in regression models
  - https://cran.r-project.org/package=intsel
- R package `mtool`: An efficient R toolbox for multitask regression with structured penalization
  - Efficient C++ Armadillo program with R interface
  - Comprehensive support for any combination of four types of regression models and over twenty different penalty functions
  - Private repository, available upon request
- R package `AutoAff`: Automatic author order and affiliation list generator
  - Made for colleagues in systematic review and meta analysis to automatically generate author and affiliation lists with customized formats for their manuscripts with many coauthors
  - https://github.com/ly129/AutoAff

PRESENTATIONS

- D3MI: an efficient and powerful federated imputation method for bias reduction in the analysis of distributed incomplete data by accounting for within-site correlation and between-site heterogeneity
  - Oral presentation at the 2025 National Cancer Institute (NCI) Informatics Technology for Cancer Research (ITCR) Trainee Symposium.
- Federated multiple imputation for multicenter studies using random forests: robust to model misspecification and easy to apply.
  - Oral presentation at the 2025 National Cancer Institute (NCI) Informatics Technology for Cancer Research (ITCR) Annual Meeting.
- Federated multiple imputation for variables that are missing not at random in distributed electronic health records.
  - Oral presentation at the American Medical Informatics Association (AMIA) 2024 Annual Symposium.
  - Oral presentation at the AMIA Knowledge Discovery and Data Mining (KDDM) Working Group meeting. First Place Innovation Award.

- Flexible and adaptive multitask regression for high-dimensional fairness-aware integration of data from multiple sites.
  - Oral presentation at the International Chinese Statistical Association (ICSA) – Canada Chapter 2024 Symposium
  - Oral presentation at the Inaugural NCI Informatics Technology for Cancer Research (ITCR) Trainee Symposium
  - Spotlight Talk at the 2024 NCI Junior Investigator Meeting
- Fast Computation of High-dimensional Estimating Equations Using Anderson Acceleration
  - Oral presentation at the 36th New England Statistics Symposium. Invited session: Recent Developments in High-Dimensional and Nonparametric Regression: Algorithms and Methodologies
- `mtool`: An efficient R toolbox for multitask regression with structured penalization
  - Oral presentation at the 2021 Canadian Statistical Sciences Institute Showcase
  - Oral presentation at the 2021 McGill University (Bio)Statistics Research and Career Day
- A Tweedie compound Poisson model in reproducing kernel Hilbert space
  - Oral presentation at the International Chinese Statistical Association 2021 Applied Statistics Symposium. Invited Session 15: Statistical Learning and Variable Selection
  - Oral presentation at the 2021 Joint Statistical Meetings. Topic-contributed Session 404: Recent Research in High-dimensional and Complex Data analysis
  - Invited seminar talk at Hebei University Of Technology, June 2021
  - Oral presentation at the 2019 Annual Meeting of the Statistical Society of Canada
- Utilization and patterns of potentially inappropriate use of Phosphodiesterase type 5 inhibitors in the United Kingdom
  - Poster presentation at the 2017 International Conference of Pharmacoepidemiology
- Phosphodiesterase type 5 inhibitors and the risk of melanoma skin cancer
  - Awarded student oral presentation at the International Society for Pharmacoepidemiology 2016 Mid-Year Meeting

PROFESSIONAL EXPERIENCE

- The Scleroderma Patient-centered Intervention Network
  - Statistical consultant 2020-2021
- TD Insurance Meloche Monnex Montréal, Canada
  - Intern in Advanced Analytics (data science) Summer 2019
    Development of a predictive modelling method for the the optimal auto repair shop recommendation and implementation of the corresponding software tool

TEACHING EXPERIENCE

- Department of Epidemiology, Biostatistics and Occupational Health — McGill University
  - Course Instructor
    - EPIB 613 Introduction to Statistical Software — Fall 2018, 2019, 2020
  - Teaching Assistant
    - EPIB 603 Intermediate Epidemiology — Fall 2016
    - PPHS 602 Foundations of Population Health — Winter 2017
- Department of Mathematics and Statistics — McGill University
  - Teaching Assistant
    - MATH 423/533 Regression and Analysis of Variance — Fall 2018, 2019, 2020, 2021
- McGill Initiative in Computational Medicine — McGill University
  - Workshop Instructor — Summer 2019, Winter 2020
    - R Programming beyond the Basics – Efficient Coding and Computing

MENTORING EXPERIENCE

- Andrew Cheng: research trainee, 4th year undergraduate student in Statistics — Summer 2021
  - Multitask Cox proportional hazards model with tree-guided group-lasso regularization
- Jacob Shkrob: research trainee, 3rd year undergraduate student in Statistics — Summer 2021
  - Group-lasso regularized high-dimensional additive Cox model with time-varying effects

ACADEMIC SERVICES

- Journal reviewer for
  - The Journal of American Medical Informatics Association (JAMIA)
  - Stat
- 2024 NCI Junior Investigator Meeting — Bethesda, MD
  - Planning Committee Member — August, 2024
- The Sixth ICSA – Canada Chapter Symposium — Niagara Falls, Canada
  - Session chair – Advances in Statistical Learning — June 2024
- The Inaugural NCI ITCR Trainee Symposium — Virtual
  - Session chair – General Informatics Resources and Platforms — May 2024
- The Fifth ICSA – Canada Chapter Symposium — Banff, Canada
  - Session chair – Recent Advancements in Statistical Genetics Methodology — July 2022
- McGill University — Montréal, Canada
  - Organizer of the 2019 McGill (Bio)Statistics Research and Career Day — September 2019
  - Volunteer at the McGill Mini-Med lecture series presented by the Department of Epidemiology, Biostatistics and Occupational Health for science promotion — November 2014
- International Conference on Pharmacoepidemiology
  - Abstract reviewer for the 34th ICPE annual conference — February 2018

- Lady Davis Institute, Jewish General Hospital — Montréal, Canada
  - Epidemiology Axis Representative of the Trainee Committee — 2016 – 2017
- Statistical Society of Canada
  - Member
  - Volunteer at the 2018 annual Meeting of the Statistical Society of Canada

SKILLS

Natural Languages: English, French (Canadian), Chinese (Mandarin)
Programming Languages: R, Python, C++, Fortran, MATLAB
Operating Systems: MacOS, Linux, Windows

MISCELLANEOUS

Country of citizenship: Canada

Last updated: July 14, 2025