Yushu Zou

✉ yushuzou@gmail.com   📞 647-569-6874   **in** LinkedIn   🐙 Github   ⌂ Website

Qualification Highlight

- **Extensive Clinical Data Experience**: Over 3 years managing, integrating, and analyzing large-scale clinical and administrative health datasets (clinical cohort data, ICES, Statistics Canada, ON-Marg).
- **Advanced R Programming**: Expert in R for data cleaning, longitudinal and survival analysis, Bayesian inference, causal modeling, and machine learning. Developer of R packages (*Causens*, *ClickTest*).
- **Advanced Statistical Modeling & Data Analysis**: Proficient in time series analysis, regression modeling, and machine learning to assess healthcare performance and patient outcomes.
- **Collaboration & Communication**: Proven ability to work cross-functionally with clinicians, epidemiologists, and business leaders to translate analytical insights into actionable recommendations, fostering informed decision-making and accountability.

Technologies

- **Programming**: R, SAS (Base & Advanced Certified), Python
- **Modeling**: Longitudinal data analyses, Survival analyses, Mediation Analyses, Bayesian inference, Time series analyses, Machine Learning
- **Tools**: RShiny, Stan, RJags, WinBUGS, SQL, ggplot2, Matplotlib

Education

MSc., Biostatistics   *Sept. 2022 – Oct. 2024*
University of Toronto Dalla Lana School of Public Health

- GPA: 4.0/4.0
- **Thesis:** A Bayesian approach for sensitivity analysis in longitudinal data

HBSc., Statistics, Specialist in Machine Learning and Data Mining   *Sept 2018 – May 2022*
University of Toronto Scarborough

Experience

Research Coordinator   *Toronto, CA*
Public Health Ontario   *June 2024 – Present*

- Linked administrative health databases (ICES, ON-Marg) with climate (ECCC) and mortality data to study health effects of extreme temperatures.
- Implemented distributed lag nonlinear models and case-crossover designs in R to estimate heat-related mortality and identify at-risk populations.
- Designed reproducible R workflows from imputation to time-series modeling for operational heat warning analytics.
- Collaborated with epidemiologists to design and implement statistical models, supporting regulatory and operational decision-making.

Research Assistant   *Toronto, CA*
Department of Medicine, University of Toronto   *May 2022 – Present*

- Led data integration and pre-possessing of large-scale, high-dimensional longitudinal clinical datasets, revolving inconsistent and ensuring data quality through advanced programming in **R** and **Python**
- Applied advanced R programming for longitudinal analysis, causal inference, and survival analysis to support both preclinical and clinical studies.
- Prepared technical reports and translated statistical results into recommendations.

Practicum Student   *Toronto, CA*
Public Health Ontario   *Oct. 2022 – Apr.2024*

- ○ Applied a comprehensive suite of statistical models and machine learning algorithms in **Python** to forecast COVID-19 case trends in Ontario, leveraging data analytics to inform public health strategies.
- ○ Designed and tested various predictive models, including regression, time series analysis (e.g., ARIMA), and advanced machine learning techniques (e.g., Random Forest, XGBoost) to enhance forecast accuracy.
- ○ Collaborated with public health officials to interpret model outputs, provided actionable insights for pandemic response efforts and contributed to the development of targeted, data-driven intervention strategies.

Publications

1. Mitani, A. A., **Zou, Y.**, Leatherdale, T. S., Patte K. A., "Investigating the association between school substance programs and student substance use: accounting for informative cluster size" *Epidemiologic methods* vol. 14,1 20240028. 26 Aug. 2025, doi:10.1515/em-2024-0028

2. Makan, G., Wu, J.K., Chow, C.W. **Zou, Y.**, Dandurand, R.J., Czövek, D., Gingl, Z. and Hantos, Z., 2025. "Tracking of respiratory mechanics at multiple oscillation frequencies." *Computer Methods and Programs in Biomedicine*, p.108818.

3. Bosch, A., Brunsvig, J.K., Brandao, L., **Zou, Y.**, Vincelli, J., Amiri, N., Avila L., (2024). "The role of coagulation factors VIII, IX and XI in the prediction and mediation of recurrent thrombotic events in children with non-central venous catheter deep vein thrombosis", *Thrombosis Research* Volume 236, 228-235.

4. Metko, D., **Zou, Y.**, Abu-Hilal, M. (2023). "Primary care providers' awareness of comorbidities and new advanced therapies for atopic dermatitis: A Cross-Sectional Survey from Ontario." *Journal of Cutaneous Medicine and Surgery*, 276, 656-657.

5. **Zou, Y.**, Hu, L., Ricciuto, A., Deneau, M., Liu, K., "Bayesian Sensitivity Analysis for Causal Estimation with Time-varying Unmeasured Confounding" (Under Revision of *Statistics in Medicine*)

Projects

- ○ Evaluation for Disparities in Achieving Academic Promotion in a Paediatric Academic Center (**R**): Survival Analysis, Mediation Analysis, Principle Component Analysis
- ○ Predict Covid-19 case using ensemble machine learning model and statistical model (**Python**): ARIMA, XGboost, LSTM to improve prediction accuracy
- ○ Causens **(R package)**: Perform various sensitivity analysis methods to adjust for unmeasured confounding within the context of causal inference. (Link) ☑
- ○ ClickTest **(R package**): Automates the testing of assumptions for various hypothesis tests, intelligently selecting and executing the appropriate test to yield reliable hypothesis testing or correlation analysis results. (Link) ☑