Project Name: Development of innovative Bayesian methods to address statistical challenges in longitudinal electronic health record data

Supervisor:
- Name: Zihang Lu
- Affiliation: **Assistant Professor (tenure track)** Biostatistics Program, Department of Public Health Sciences, Queen's University
- Email: zihang.lu@queensu.ca
- Website (if available): https://phs.queensu.ca/faculty-research/zihang-lu

Co-supervisor:
- Name: Kuan Liu
- Affiliation: Assistant Professor in Biostatistics (cross-appointment), University of Toronto
- Email: kuan.liu@utoronto.ca
- Website (if available): https://www.dlsph.utoronto.ca/faculty-profile/liu-kuan/

Location/University:
- Queen's University
- University of Toronto

Abstract:
The use of Electronic Health Records (EHR) has become increasingly common in medical research and it is viewed as an important data source for scientific discovery. Several Bayesian methods for analyzing EHR data have been proposed recently. Despite these advances, there remain several methodological research gaps especially for longitudinal EHR data and poor uptake of Bayesian methods for analyzing EHR data in practice, due to hurdles such as the generalizability, computational challenges, interpretability, and the lack of user-friendly statistical software.

To this end, the postdoctoral fellow will work on developing innovative, flexible, generalizable and interpretable Bayesian methods to analyze large and complex real-world clinical and public health data. The methodological contribution will focus on patient phenotyping and causal inference under longitudinal settings. This research also includes the development of accompanying software packages, open-access code, and tutorials to facilitate a wide application of the proposed methods.

Interdisciplinary/applied experience:
This proposal facilitates new collaboration among faculty members and PDF from multiple institutes with appointments across several departments including statistics, biostatistics, and medicine. Apart from working with the supervisory team (Dr.s Lu, Liu and Stephens), the PDF will work collaboratively with Dr. Eddy Fan (University of Toronto) and Dr.Laveena Munshi (University of Toronto). Drs. Fan and Munshi are leading researchers in critical care medicine who will facilitate access to intensive care EHR data and provide clinical expertise on the application of the proposed research.

CANSSI Distinguished Postdoctoral Fellows Projects List

The proposed research will be carried out at both Queen's University and University of Toronto. Drs. Lu and Liu will be supervisor and co-supervisor who will lead the training and work closely with the PDF on a daily basis. Dr. Stephens will serve as research mentor and project collaborator. He will provide methodological guidance and professional development support including career development on a monthly basis. During the analysis stage, monthly virtual team meetings with clinical collaborators (Drs. Fan and Munshi) will also be organised.

This interdisciplinary training partnership will foster an interdisciplinary mentorship environment and promote training experience to equip the fellow with both technical and professional skills to succeed in academia and industry as an independent scientist. Training will encompass further development of technical and scientific communications skills(including non-technical presentation, grant and proposal writing), teaching and mentorship skills, leadership and collaborative research skills. There will be ample opportunities for the PDF to be involved in other collaborative biostatistical research projects at the host universities.

Teaching/training/education:
The planned training will take place at Queen's University (Kingston, ON) between September and April in Year 1 and Year 2and at University of Toronto (UofT , Toronto, ON) between May and August in Year 1 and Year 2. The PDF will be engaged in a variety of teaching and mentorship activities. Specifically, the PDF will teach one three-credit course in Biostatistics each year at Queen's University. In addition, together with Drs. Lu and Liu, the PDF will participate in organizing and teaching a one-day pre–conference short course on applied Bayesian statistics in Year 2.

The PDF will be provided with opportunities to engage in primary supervision of summer undergraduate research students in statistics at UofT, and joint supervision of graduate students in biostatistics from supervisors' research groups at Queen's University and UofT throughout the two-year training program.

The PDF will be funded to attend and present at national and international conferences every year, such as the SSC Annual Meeting, the Joint Statistical meeting, and the Eastern North American Region Biometric Society Meeting. Additionally, PDF will be encouraged and supported to present and participate in various research seminars and workshops within and outside the host institutes, such as Biostatistics research seminars, clinical epidemiology rounds, and the CANSSI research day.

Mentoring of the postdoctoral fellow:
The PDF will be supported to direct their own research, engage in problem solving and critical scientific appraisal of the work. All technical training uses a structured open-door policy, with team meetings between PDF and supervisors scheduled once a week and additional support offered by email or in-person as needed. Monthly meetings between PDF, supervisors and mentor will support the ongoing research and professional training progress.

Drs. Lu and Liu will function as supervisor and co-supervisor who will lead the PDF training and work closely with the PDF on a daily basis. Dr. Stephens will function as research mentor and collaborator who will provide methodological guidance and professional development support including career development on a monthly basis. During the analysis stage, monthly virtual team meetings with clinical collaborators (Drs. Fan and Munshi) will also be organised.

Queen's University and UofT are Canada's leading research-intensive universities with a global reputation for cutting-edge collaborative research intertwining statistics, data sciences, medicine and public health. The PDF will have access to seminars, training, scholarships, networking opportunities, collaborations, and peer support through various initiatives across the two universities including the Data Sciences Institute and Vector Institute at Toronto and the Canadian Cancer Trials Group at Queen's University.

Proposed schedule:

Year 1:
   (a) Research: The PDF will work on developing a scalable hidden Markov model for electronic health record data with informative missingness (Objective 1). Under the supervision of Drs Lu, Liu and Stephens, the PDF will work on the model specification, developing efficient algorithms and software code in Year 1 (Sep-Apr). This newly developed method will be applied to real data applications (Year 2, May-Aug).
   (b) Teaching: The PDF will teach a full course (three-credit) at Queen's University and engage in supervising graduate students on their research projects.
   (c) Other activities: Prepare the manuscript and submit it for publication. Attend and present research findings in seminars, workshops and conferences. Help organize seminars and workshops.

Year 2:
   (a) Research: The PDF will work on developing Bayesian causal methods with time-invariant and time-varying clustering (Objective 2). Under the supervision of Drs Lu, Liu and Stephens, the PDF will work on the model specification, developing efficient algorithms and software code in Year 2 (Sep-Apr). This newly developed method will be applied to real data applications (Year 2, May-Aug).
   (b) Teaching: The PDF will teach a full course (three-credit) at Queen's University and engage in supervising graduate students on their research projects.
   (c) Other activities: Prepare the manuscript and submit it for publication. Attend and present research findings in seminars, workshops and conferences. Help organize seminars, workshops, and pre-conference short courses.

Qualifications:

1. A doctoral degree in Biostatistics or Statistics
2. Strong programming skill; Proficient in R, C++ and SAS3.
3. Knowledge of Bayesian statistical methods and modelling
4. Strong communication, written and analytical skills
5. Experience in collaborative research
6. Ability to independently organize workload, set goals and work effectively towards deadlines
7. Experience with analyzing clinical and public health data is an asset