Machine Learning for Precision Public Health Workshop –
Ethical Considerations in the Era of Big Data

Summary

8th of April, 2019, 1:15 – 3 PM
Hardwick Hall, University of British (UBC) Medical Student & Alumni Centre

Overview of the workshop

This workshop was part of the special seminar series Machine Learning for Precision Public Health, organized by the BC Centre for Disease Control (BCCDC). Following the presentation "Who's Afraid of Big (Bad) Data?"[1] by Dr. Diego Silva, an Assistant Professor of the Faculty of Health Sciences at Simon Fraser University, this workshop aimed to provide an opportunity to deeper explore potential ethical issues that arise in the application of data science approaches to public health, and to identify key principles and new considerations needed to guide decision making.

A total of 11 participants attended the workshop, run in a format of small group discussions co-facilitated by Dr. Diego Silva, Dr. Mark Gilbert, Medical Director of Clinical Prevention Services (CPS) at the BCCDC, Dr. Mike Irvine, a Postdoctoral Fellow at the BCCDC, and Hsiu-Ju Chang, Research Manager of CPS at BCCDC. Divided into three groups, participants used the BCCDC Ethics Framework and Decision Making Guide as a tool to discuss potential ethical issues in provided scenarios where data science approaches are applied in public health. The workshop closed with whole group discussions and final Q&A.

BCCDC Ethics Principles and Decision Making Guide

Effective since May 2011 and reviewed in May 2015, the current version of the BCCDC ethics framework states that the principles of the ethical practice of public health at the BCCDC:

- Shall address, principally, the fundamental causes of disease and requirements for health, aiming to prevent adverse health outcomes
- Aspires to achieve community health in a way that respects the rights of individuals in the community
- Is committed to community engagement
- Will seek the information needed to implement effective policies and programs that protect and promote health
- Acts in a timely manner on the information it has

---

[1]In the Grounds Presentation of "Who's Afraid of Big (Bad) Data?" prior to this workshop, Dr. Silva provided an overview and reconsideration of traditional bioethical and political challenge related to privacy and confidentiality, data sharing, and data security with the growing data sciences in medicine and public health.

- Promotes the empowerment of vulnerable and disenfranchised community members, aiming to ensure that the basic resources and conditions necessary for health are accessible to all
- Incorporate a variety of approaches that anticipate and respect diverse values, beliefs, and cultures in the community
- Ensures proportionality in its programs and activities
- Properly justifies the creation and implementation of its programs
- Has clearly stated program and policy goals that have proven effectiveness
- Will use the least restrictive or coercive means possible to achieve its goals
- Whenever possible, will adopt a principle of reciprocity
- Will protect the confidentiality of information that can bring harm to an individual or community if made public.
- Will ensure the professional competence of its employees
- Engages in collaborations and affiliations in ways that build the public's trust and the institution's

The ethics framework also sets out the following decision making guide:
- Identify the ethical question.
- Identify the stakeholders.
- Clarify the facts, gather information.
- Analyze the problem in light of the values and principles. Try to identify the origins of the tensions from the conflicting values and principles.
- Identify relevant legal and normative guidance: legal and legislative considerations; local policy and procedures; professional codes of ethics; research guidelines; moral intuition and ethical considerations.
- Identify possible courses of action.
- Make a decision.

Scenarios for discussions
The following scenarios were provided to workshop participants for discussions.

Scenario 1:
A provincial big data cohort that links the provincial immunization program data and other administrative and clinical datasets is used to develop a prediction algorithm for measles outbreaks. A school has been identified as having a number of pupils who have not received vaccinations and the algorithm is predicting that an outbreak is highly probable at this location. How should the results be communicated? Should the school be provided the names of pupils who aren't vaccinated?

Scenario 2:
Public health authorities have established a linked health services administrative database to identify people with opioid use disorder (OUD) and to look at their use of OAT (e.g., opioid agonist therapy, OAT). It is known that people who are not on OAT for their opioid use disorder

are at higher risk of overdose due to fentanyl in the illicit opioid drug supply, and initiatives to improve retention in OAT are ramping up in the province. Public health authorities have decided to use the linked data to identify people with OUD not on OAT so that they can be contacted by public health staff and connected to OAT programs. What do you think are the ethical considerations related to this initiative?

Scenario 3:
You have conducted a large cohort study recording the daily speech of 20,000 people with anxiety disorder, correlating speech patterns to self-rated anxiety levels, and have developed a predictive rule for severe anxiety requiring intervention. From your research you have developed a publicly available smartphone application for people with anxiety disorder that records an individual's conversations throughout the day and predicts how anxious an individual is at any given point in time. When the application predicts severe anxiety, the application can alert the user or directly call a healthcare provider. What ethical considerations should you be thinking about? What would you do to mitigate them?

Scenario 4:
A machine vision algorithm is developed to determine an individual's risk of contracting an STI by analysing a recent photograph of that individual's face. The algorithm reports a 95% accuracy in predicting whether an individual has ever had an STI within the past six months (Note: This may seem slightly far-fetched but there have been recent cases where this type of approach has been applied to criminality , or sexuality). You are a public health researcher partnering with a dating site on a number of health promotion initiatives. Should you use this algorithm to provide tailored information on STI prevention? What are some of the implications of using this kind of approach?

Potential addition of ethical considerations in the era of big data

- Evaluate the performance of a machine learning algorithm before applying it
    - How accurately does it predict the outcome of interest?
    - Does it perform better in certain subgroups than the others?
- Communicate the level of uncertainty and accountability of a prediction algorithm
    - How much uncertainty is around the prediction?
    - What is the risk communication strategy?