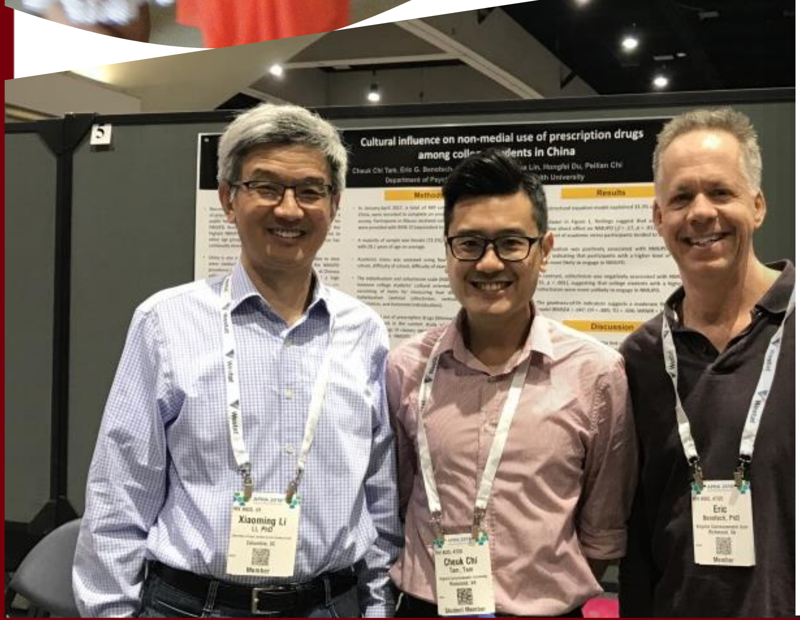
MACHINE-LEARNED IDENTIFICATION OF PRESCRIPTION OPIOID MISUSE AND ITS PSYCHOSOCIAL AND BIOMEDICAL DETERMINANTS IN PEOPLE LIVING WITH HIV

Cheuk Chi (George) Tam, Ph.D., M.S., M.A.
Research Assistant Professor
Department of Health Promotion, Education, and Behavior
Arnold School of Public Health

Unlocking the Power of BIG DATA IN HEALTH
Empowering Scientific and Healthcare Communities with Data Analytics

5TH NATIONAL BIG DATA HEALTH SCIENCE CONFERENCE
February 2-3 2024 | Columbia, SC



Background

- Clinical training in psychotherapy and HIV prevention intervention

- Doctoral training in Health Psychology focusing on psychometric and theory-driven investigation of **prescription opioid misuse (POM)** among vulnerable population

- Research faculty member in Health Promotion, Education, and Behavior at Arnold School of Public Health, USC
  - Apply advanced methodologies to examine or address psychosocial aspects of **POM** or relevant treatment
  - Cascade of care for opioid use disorder
  - Technology-based intervention for prescription opioid misuse

Research Plan

The successful risk assessment and POM prevention in PLWH relies on valid measures of POM and the understanding of multifaced features of POM,

- the use of multiple sources of high-dimensional data to examine the complex mechanism underlying the **interactions among biomedical, psychiatric, and social determinants**

Training Aims

- AI modeling on aggregated data from various sources of clinical and addiction-related datasets (EHRs, lab reports, prescription records, self-reported surveillance, and state-based data)
- Frontier insights and models of complex medical and psychiatric needs of people with POM/treatment
- Valid use of HIV surveillance EHR data

Gaps

- My previous training relied solely on self-collected psychometric data from a selected group of the target population
- My data analytic skill sets are built upon traditional statistics, which has various restrictions on data (e.g., normality, linearity, outliers, and non-collinearity),

R25 Fellowship

Mentored Training

Big-Data-based analytic training and data aggregation

Bankole Olatosi, Ph.D.

Assistant Professor and MHA Program Director, Health Service, Policy and Management, Arnold School of Public Health

Alain Litwin, M.D.

Professor and Vice Chair of Internal Medicine at USC School of Medicine-Greenville

Clinical and addiction-related research and data guidance

LEANING ACTIVITIES

- Journal club
- BDHSC Seminars
- CITI: Technology, Ethics, and Regulation
- Data Science Ethics: University of Michigan by Coursea (14 hours)
- BDHSC Linux and Cloud Computing for Bioinformatics workshop (20 hours)
- SAS courses

Ellis et al. BioData Mining (2019) 12:3
https://doi.org/10.1186/s13040-019-0193-0

BioData Mining

RESEARCH Open Access

Predicting opioid dependence from electronic health records with machine learning

Randall J. Ellis[1], Zichen Wang[1], Nicholas Genes[2] and Avi Ma'ayan[1*]

* Correspondence: avi.maayan@mssm.edu
[1]Department of Pharmacological Sciences, Mount Sinai Center for Bioinformatics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
Full list of author information is available at the end of the article

Abstract

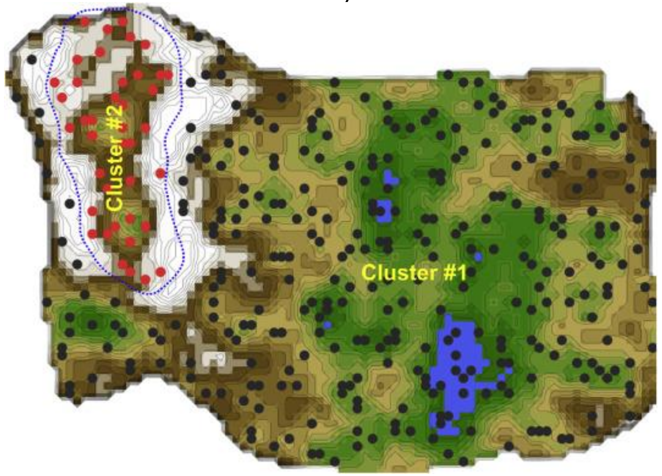
Background: The opioid epidemic in the United States is averaging over 100 deaths per day due to overdose. The effectiveness of opioids as pain treatments, and the drug-seeking behavior of opioid addicts, leads physicians in the United States to issue over 200 million opioid prescriptions every year. To better understand the biomedical profile of opioid-dependent patients, we analyzed information from electronic health records (EHR) including lab tests, vital signs, medical procedures, prescriptions, and other data from millions of patients to predict opioid substance dependence.

LEARNING ACTIVITIES

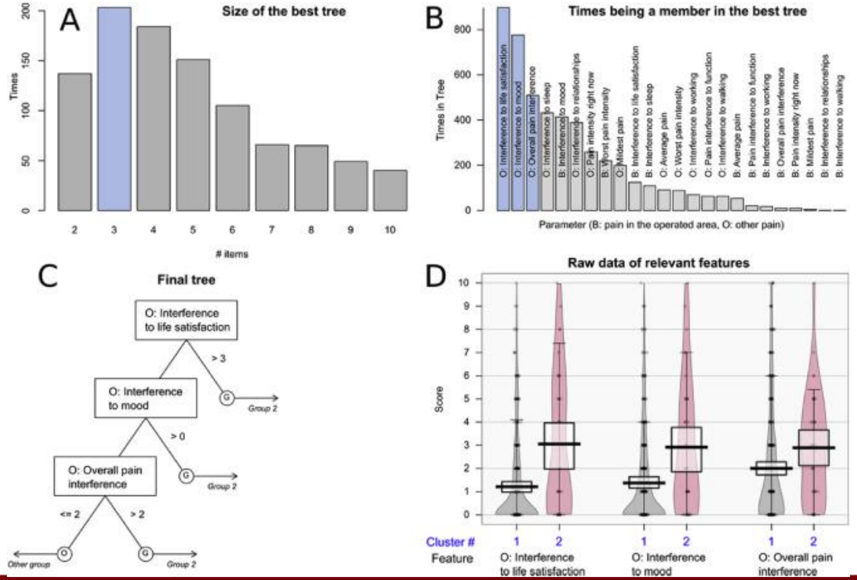
- SAS Courses:
  - Introduction to Data Science Statistical Methods (7 hours)
  - Deep Learning Using SAS® Software (14 hours)
  - SAS Viya and R integration for Machine Learning (2 half days)
  - Neural Network: Essentials (10.5 hours)

RESEARCH PLAN

- By aggregating multiple sources of data ranging from individual-level dimensions (e.g., self-reported surveys and electronic health records [EHRs]) to county-level dimensions (e.g., service data from state-based health institutes/agencies), recent research on patients with chronic pain used artificial-intelligence-based algorithms to (1) identified subgroups based on the data structure among intercorrelated psycho-behavioral parameters and (2) relate this structure to diagnostic and clinical data for interpreting the psycho-behavioral characteristics in terms of pain symptoms, treatments, and live impacts
- Such an approach appears to be valuable for POM in PLWH research in that it enables to handle massive data from high dimensions and robustly classify at-risk groups based on an array of behavioral and clinical proxies of POM (e.g., diagnosed/self-reported behaviors or symptoms of prescription opioid misuse) extracted from multiple data sources with the interpretation on the multifaced risk factors from various levels (e.g., time to initiate POM, pain interference/severity, linkage/access to treatments, psychiatric symptoms, and availability of treatment facility, opioid treatment access act, and socioeconomic characteristics)



Sipilä, R., Kalso, E., & Lötsch, J. (2020). Machine-learned identification of psychological subgroups with relation to pain interference in patients after breast cancer treatments. *The Breast*, *50*, 71-80.



RESEARCH PLAN

- **Aim 1.** To evaluate POM and identify its clinical and behavioral patterns in PLWH using aggregated data of self-reported POM risk behaviors (e.g., early refills, take the medication without an approval from physician, obtaining medication, frequencies of POM), EHRs (e.g., diagnosed opioid use disorder, lab test reports of opiate component intake, drug overdoses), and prescription records (e.g., time of prescription, exposure to opioid medications, prescribe daily doses);
- **Aim 2:** To apply machine learning modeling techniques to classify subgroups of PLWH in terms of their POM risk using the clinical and behavioral indicators;
- **Aim 3:** To determine the psychosocial and biomedical predictors on the cluster structure identified in Aim 2 through machine learning modeling accounting for factors at the individual level (e.g., psychiatric symptom diagnosis, poly-drug use/co-occurring substance use disorder, self-reported stress levels, pain symptom diagnosis, and mental health/pain management service utilization, HIV progression, antiretroviral treatment engagement), county level (e.g., geographic region, availability of opioid treatment programs [OTPs]), and socioeconomic level (e.g., socioeconomic characteristics).

The emergent self-organizing feature maps (ESOM)

the simple decisions trees "Fast and Frugal Trees" (FFTs) combined with ABC analysis

RESEARCH PLAN

Individual-level data source: All of Us

In particular, the POM risk behavior and biomedical/psychosocial determinant variables in each data domains would include:

Self-reported surveys*:* Demographics and geographics (e.g., age, gender, race/ethnicity, and residency), overall health status (e.g., mental health, pain, emotional problems, qualify of life), substance use (e.g., past-3-month non-medical use of prescription opioids, poly substance use), personal medical history (e.g., mental health or substance use conditions told by healthcare providers), health care access and utilization (e.g., unaffordable prescription medicines, delayed filling of prescription, brought prescription drugs from another country, skipped doses), and social determinants of health (e.g., less respect than other people at the clinics)

EHRs domains data: Medical conditions (e.g., psychiatric symptoms, pain symptoms, substance use disorders, opioid use disorder), lab reports (e.g., opiate component intake, HIV-1 antigen), healthcare procedure data (e.g., outpatient visit, emergency department visit), Drug exposures (e.g., prescriptions for opioid medications [e.g., oxycodone, fentanyl], opioid treatments [e.g., naloxone], and antiretroviral medication [e.g., nucleoside]), lab and measurements (e.g., HIV presence in serum, opiates presence in urine/serum)

County-level data source: Substance Abuse and Mental Services Administration and Area health resources file

Substance Abuse and Mental Health Services Administration (SAMHSA) databases: National Survey of Substance Abuse and Mental Health Services (N-SUMHSS) and National Survey on Drug Use and Health (NSDUH)*:* These two databases from SAMHSA provide state-level data on drug abuse, mental health, and relevant treatment/facilities information. Data that will be extracted for the proposed study include state-level substance abuse (e.g., pain reliever misuse), mental health outcomes (e.g., serious psychological distress), substance use severity level, past-year substance use treatment (e.g., for prescription pain killers and opioids) , substance use treatment at specialty facility status, and mental health services.

Behavioral risk factor surveillance system (BRFSS): BRFSS is a CDC-funded database from state-based surveys collecting information on health-risk behaviors, prevention health practices, chronic conditions, and health care access. The proposed study will extract data regarding: substance use treatments (e.g., tobacco use and alcohol use cessation), health care access, pain severity, and mental health outcomes at the county level.

THANKS FOR YOUR ATTENTION

Cheuk Chi (George) Tam
Research Assistant Professor
ctam@mailbox.sc.edu

Big Data Health Science Center
UNIVERSITY OF SOUTH CAROLINA