DEPARTMENT: Behavioral, Social & Health Education Sciences

COURSE NUMBER: BSHES 760R/GRAD 700R (+lab)

CREDIT HOURS: 4 credits SEMESTER: Spring ’21

COURSE TITLE: Reducing Drug-Related Harms Using Big Data: Administrative, geospatial, and network sources

DATES & LOCATION: Fridays 10 AM to 1 PM (Consists of a 2-hour lecture and 1-hour lab). This will be offered as an on-line course.

INSTRUCTORS: Hannah Cooper, ScD
Professor
Department of Behavioral, Social and Health Education Sciences
Rollins School of Public Health
Email: hcoope3@emory.edu

Lance A. Waller, Ph.D.
Professor
Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Email: lwaller@emory.edu

Weihua An, PhD
Associate Professor
Departments of Sociology & Quantitative Theory and Methods
College of Arts and Sciences
Email: weihua.an@emory.edu

OVERALL COURSE DESCRIPTION:

This course will prepare students to conduct ethical, rigorous, and theoretically-informed analyses of three types of “big data” (administrative, geospatial, and social network data) in the context of research and interventions into intersecting crises of substance use disorders (SUDs) and drug-related harms. It will apply the strengths of social and behavioral sciences – including a focus on theory and validity – to the emerging field of advanced data analytics.

This course is one of two courses on analyzing “big data” to study and intervene in drug-related harms (the other course is entitled “Reducing Drug-Related Harms Using Big Data: Gene*Environment Interplays and Machine Learning”). We recommend, but do not require, taking this course first.
COURSE COMPETENCIES:

Trainees will learn to:

- Design and conduct theoretically-informed analyses of distributions and ecologies of SUD-related harms by applying advanced data science methods to administrative data, geospatial data, and sociometric network data.
- Design and conduct theoretically-informed analyses assessing policies and programmatic interventions that may affect SUD-related harms and services by applying advanced data science methods to administrative data, geospatial data, and sociometric network data.
- Communicate findings to select stakeholder communities to strengthen efforts to end SUD-related harms.
- Critically apply principles of the ethical and responsible conduct of research.

OVERALL COURSE LEARNING OBJECTIVES:

Upon completion of this course, the student will be able to:

1. Design theoretically guided analyses describing distributions and ecologies of SUD-related harms and services using appropriate data science methods for administrative data, geospatial data, and sociometric network data.
2. Conduct theoretically guided analyses describing distributions and ecologies of SUD-related harms and services using appropriate data science methods for administrative data, geospatial data, and sociometric network data.
3. Design theoretically guided analyses of policies and programmatic interventions that may affect SUD-related harms and services using appropriate data science methods for administrative data, geospatial data, and sociometric network data.
4. Conduct theoretically guided analyses of policies and programmatic interventions that may affect SUD-related harms and services using appropriate data science methods for administrative data, geospatial data, and sociometric network data.
5. Compare the rigor (e.g., validity) of various data science methods as tools to study and intervene in SUD-related harms and services.
6. Communicate the rationale, methods, findings, and conclusions of theoretically guided analyses of administrative data, geospatial data, and sociometric network data describing distributions and ecologies of SUD-related harms and services to diverse audiences.
7. Communicate the rationale, methods, findings, and conclusions of theoretically guided analyses of policies and programmatic interventions that may affect SUD-related harms and services to diverse audiences.
8. Assess ethical issues posed by each data science method and consider the responsible conduct of related analyses, particularly as applied to SUD-related research.

REQUIREMENTS

This course will require computing in different programs and different environments. Familiarity and comfort with the following is needed to successfully complete the course:

- Regression (examples: BIOS 501, BSHES 700)
- SAS (examples: BIOS 501)
- R (examples: BIOS 544)
EVALUATION

Class participation 30%
Case study final project 40%
Other Written Assignment + Posts 15%
Laboratory assignment 15%

RSPH POLICIES

Accessibility and Accommodations

As the instructors of this course we endeavor to provide an inclusive learning environment. However, if you experience barriers to learning in this course, do not hesitate to discuss them with us and the Office of Accessibility Services (OAS). Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact the OAS. It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed.

Students who registered with OAS and have a letter outlining their academic accommodations are strongly encouraged to coordinate a meeting time with me to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible.

Contact Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at http://equityandinclusion.emory.edu/access/students/index.html

Honor Code

You are bound by Emory University’s Student Honor and Conduct Code. RSPH requires that all material submitted by a student fulfilling his or her academic course of study must be the original work of the student. Violations of academic honor include any action by a student indicating dishonesty or a lack of integrity in academic ethics. Academic dishonesty refers to cheating, plagiarizing, assisting other students without authorization, lying, tampering, or stealing in performing any academic work, and will not be tolerated under any circumstances.

The RSPH Honor Code states: “Plagiarism is the act of presenting as one’s own work the expression, words, or ideas of another person whether published or unpublished (including the work of another student). A writer’s work should be regarded as his/her own property.” (http://www.sph.emory.edu/cms/current_students/enrollment_services/honor_code.html)
**MODULE I: ADMINISTRATIVE DATABASES**

In this module, students will be introduced to the course and will learn to conduct ethical, rigorous, and theoretically-informed analyses of administrative data in the context of research and interventions into intersecting crises of SUDs and drug-related harms.

**SCHEDULE**

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Module I Week 1</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Class session | Introduction to the course  
   o First hour: Intersecting epidemics of drug-related harms and related intersectional inequities  
   o Second hour: Integrating advanced data analysis into the social and behavioral lifecycle to end drug-related harms  
   o Third hour: Cook & Campbell’s Validity framework | Hannah Cooper/Lance Waller |
| Lab | • Introduction to the case study | Hannah Cooper |
| **Module I Week 2** | | |
| Class session | Administrative Data and their utility in studying the impacts of laws on drug-related health outcomes and intersectional inequities  
   o First hour: Theory and the Public Health Law Research Model  
   o Second hour: Introduction to Administrative Data  
   o Third hour: Integrating administrative data from multiple sources to explore the impacts of laws on intersectional inequities in drug-related harms | Hannah Cooper |
| Lab | • Case Study: Exploring the legal databases held at the Public Health Law Project and APIS | Hannah Cooper |
| **Module I Week 3** | | |
| Class session | Deep dive into Administrative Data and Coverage  
   o First hour: The Census, gender, and race: from the 3/5th compromise to the carceral state  
   o Second hour: NSDUH & coverage of people who use drugs  
   o Third hour: Ethics and administrative data | Hannah Cooper |
| Lab | • Case Study: Estimating the impact of undercoverage by race/ethnicity, gender, and citizenship in the US census on denominators | Hannah Cooper |
| **Module I Week 4** | | |
| Class session | Administrative Data and Validity and Construct Validity  
   o First hour: Construct validity, units of analysis, & time  
   o Second Hour: Construct validity, instrumentation, and algorithms to assess opioid overdose mortality  
   o Third hour: Construct validity, instrumentation, and measuring “race” and “ethnicity” over time | Hannah Cooper |
| Lab | • Case Study: Applying and comparing different algorithms to assess overdoses | Hannah Cooper |
| **Module I Week 5** | | |
Class Session | Administrative Data and Validity: Linking data across databases to describe people, institutions, and places  
First hour: probability-based matching for individual-level data  
Second hour: Linking databases to describe places  
Third hour: Linking databases to estimate the size of hidden populations | Hannah Cooper
---|---|---
Lab | Case study: Capture/Recapture exercise | Hannah Cooper

**Readings (Available on Canvas)**

**Week 1:**

**The Epidemiology of drug-related harms**

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5769798/


**Big Data and the Social and Behavioral Sciences**

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4167902/

**Introduction to Validity**

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7144753/

**Read if interested:**

https://www.ncbi.nlm.nih.gov/books/NBK561300/

**Week 2:**

**The Importance of Theory:**

Theoretical Model: Public health law research

Administrative data

Integrating Administrative Databases

Week 3
Who Counts in Administrative Data?: Coverage

Ethics


Week 4:
More on Validity

Measuring “Race” and “Ethnicity” in the US Census

Week 5
Integrating data to describe places
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4715941/pdf/nihms720813.pdf

Probabilistic Matching
https://textbook.coleridgeinitiative.org/chap-db.html

Capture/Recapture Methods
https://onlinelibrary.wiley.com/doi/pdf/10.1002/pds.3280?casa_token=GRffAfNdu7NdM1Mx3Vm6WqDzMicVqPFaN1NKkCuPwYPkl14GZbJV-ML8eBcfjF8PvbAk281-E

https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1360-0443.1999.94111653.x?casa_token=evyyLeOGoNDYAAAA:i7KCB5O2_1YLJ5zGxcM0jGX52C XOvdx54PyugVb8BwPDaxd_bVzAauk60mFYHUjNJ5Neb4wejQGAYU
MODULE II: GEOSPATIAL DATA ANALYSIS

In this module, students will be introduced to geospatial social epidemiology, spatially referenced data, geographic information systems, cartography, and spatial analysis relating to the public health of substance use. Students will learn concepts and computational tools for linking, managing, and analyzing geographically referenced data relating to substance use and associated local drivers.

SCHEDULE

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module II Week 1</td>
<td>Theory of local influences on drug-related health outcomes</td>
<td>Lance Waller/ Hannah Cooper</td>
</tr>
<tr>
<td>Class session</td>
<td>The Risk Environment Model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maps and health</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Neighborhoods and health</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>Reading data into ArcGIS</td>
<td>Lance Waller</td>
</tr>
<tr>
<td></td>
<td>Making a map in ArcGIS</td>
<td></td>
</tr>
<tr>
<td>Module II Week 2</td>
<td>Cartography and communication</td>
<td>Lance Waller</td>
</tr>
<tr>
<td>Class session</td>
<td>Effective display of geographic information</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Communicating with maps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ethical issues in analyzing geospatial data for drug-related research</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>Layering and joining multiple data sets</td>
<td>Lance Waller</td>
</tr>
<tr>
<td>Module II Week 3</td>
<td>Maps and theories of substance use</td>
<td>Lance Waller</td>
</tr>
<tr>
<td>Class session</td>
<td>Maps of administrative data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maps of substance use</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scale, aggregation, and the modifiable areal unit problem</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>Access to services: Buffering locations</td>
<td>Lance Waller</td>
</tr>
<tr>
<td>Module II Week 4</td>
<td>Local risk and resiliency</td>
<td>Lance Waller</td>
</tr>
<tr>
<td>Class session</td>
<td>Risk and resilience factors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vulnerability indices</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spatial regression</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>Case study work time</td>
<td>Lance Waller</td>
</tr>
<tr>
<td>Module II Week 5</td>
<td>Geographic analysis</td>
<td>Lance Waller</td>
</tr>
<tr>
<td>Class Session</td>
<td>Building a geographic analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GIS for processing data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R for analyzing data</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>Case study work time</td>
<td>Lance Waller</td>
</tr>
</tbody>
</table>

REQUIRED READINGS (Available on Canvas)

- **Week 1:**

- **Week 2:**

• **Week 3:**

• **Week 4:**

• **Week 5:**
MODULE III: SOCIAL NETWORK ANALYSIS

Interest in network analysis has exploded in the past few years, due to the advancements in statistical modeling and the rapid availability of network data. This course covers the major methods to collect, represent, and analyze network data. Selected topics include basic network analysis (centrality, positions, and clustering), the exponential random graph model for modeling network formations, causal analysis of network effects, the stochastic actor-oriented model for dynamic network analysis, and social network-based interventions. Students will learn hands-on skills to conduct their own research by using network packages in R such as “statnet” and “RSiena”.

SCHEDULE

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module III Week 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class session</td>
<td>Introduction to Social Networks and Health</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Lab</td>
<td>Basic Analysis</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Module III Week 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class session</td>
<td>• Random Network Models</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Lab</td>
<td>• ERGM</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Module III Week 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class session</td>
<td>• Network Effects</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Lab</td>
<td>• Positional Analysis</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Module III Week 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class session</td>
<td>• Dynamic Network Analysis</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Lab</td>
<td>• SAOM</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Module III Week 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class session</td>
<td>• Interventions and Ethical Issues</td>
<td>Weihua An</td>
</tr>
<tr>
<td>Lab</td>
<td>• Network Interventions</td>
<td>Weihua An</td>
</tr>
</tbody>
</table>

Recommended Textbooks


REQUIRED READINGS (Available on Canvas)

o. Introduction

1. **Network Data**

2. **Random Network Models**

3. **Network Effects**

4. **Dynamic Network Analysis**

5. **Network Interventions**