



EMORY

ROLLINS
SCHOOL OF
PUBLIC
HEALTH

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, BSHE 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

Date	Topic	Instructor
Module I Week 1		
Class session	Introduction to the course <ul style="list-style-type: none"> ○ First hour: Intersecting epidemics of drug-related harms and related intersectional inequities ○ Second hour: Integrating advanced data analysis into the social and behavioral lifecycle to end drug-related harms ○ Third hour: Cook & Campbell's Validity framework 	Hannah Cooper/ Lance Waller
Lab	<ul style="list-style-type: none"> ● 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 <ul style="list-style-type: none"> ○ First hour: Theory and the Public Health Law Research Model ○ Second hour: Introduction to Administrative Data ○ Third hour: Integrating administrative data from multiple sources to explore the impacts of laws on intersectional inequities in drug-related harms 	Hannah Cooper
Lab	<ul style="list-style-type: none"> ● 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 <ul style="list-style-type: none"> ○ First hour: The Census, gender, and race: from the 3/5th compromise to the carceral state ○ Second hour: NSDUH & coverage of people who use drugs ○ Third hour: Ethics and administrative data 	Hannah Cooper
Lab	<ul style="list-style-type: none"> ● 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 <ul style="list-style-type: none"> ○ First hour: Construct validity, units of analysis, & time ○ Second Hour: Construct validity, instrumentation, and algorithms to assess opioid overdose mortality ○ Third hour: Construct validity, instrumentation, and measuring "race" and "ethnicity" over time 	Hannah Cooper
Lab	<ul style="list-style-type: none"> ● 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

Burnett, J. C., Broz, D., Spiller, M. W., Wejnert, C., & Paz-Bailey, G. (2018). HIV infection and HIV-associated behaviors among persons who inject drugs—20 cities, United States, 2015. *Morbidity and Mortality Weekly Report*, 67(1), 23.

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

[Van Handel MM, Rose CE, Hallisey EJ, et. al. County-Level Vulnerability Assessment for Rapid Dissemination of HIV or HCV Infections Among Persons Who Inject Drugs, United States. *J Acquir Immune Defic Syndr*. 2016 Nov 1;73\(3\): 323-331.](#)

Wilson N, Kariisa M, Seth P, Smith H IV, Davis NL. Drug and Opioid-Involved Overdose Deaths — United States, 2017–2018. *MMWR Morb Mortal Wkly Rep* 2020;69:290–297. DOI: <http://dx.doi.org/10.15585/mmwr.mm6911a4>

Big Data and the Social and Behavioral Sciences

Kaplan, R. M., Riley, W. T., & Mabry, P. L. (2014). News from the NIH: leveraging big data in the behavioral sciences. *Translational behavioral medicine*, 4(3), 229-231.

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

Introduction to Validity

Matthay, E. C., & Glymour, M. M. (2020). A Graphical Catalog of Threats to Validity: Linking Social Science with Epidemiology. *Epidemiology (Cambridge, Mass.)*, 31(3), 376.

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

Read if interested:

Forum, N. C. P., Nass, S. J., Patlak, M., Zevon, E., Balogh, E., & National Academies of Sciences, Engineering, and Medicine. (2020, July). Proceedings of a Workshop. In *Applying Big Data to Address the Social Determinants of Health in Oncology: Proceedings of a Workshop*. National Academies Press (US). <https://www.ncbi.nlm.nih.gov/books/NBK561300/>

Week 2:

The Importance of Theory:

Krieger, N. (2011). *Epidemiology and the people's health: theory and context*. Oxford University Press. Chapter 1 (pgs 3-42).

Theoretical Model: Public health law research

Wagenaar, A. C., & Burris, S. C. (Eds.). (2013). *Public health law research: theory and methods*. John Wiley & Sons. Chapter 3.

Administrative data

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social science research*, 59, 1-12.
<https://www.sciencedirect.com/science/article/pii/S0049089X1630206X>

Integrating Administrative Databases

Hagger-Johnson, G., Harron, K., Gonzalez-Izquierdo, A., Cortina-Borja, M., Dattani, N., Muller-Pebody, B., ... & Goldstein, H. (2015). Identifying possible false matches in anonymized hospital administrative data without patient identifiers. *Health services research*, 50(4), 1162-1178. https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-6773.12272?casa_token=jFHbS_steT4AAAAA:751LXojaeM2NarEeV64JoC15SfHpLBocgxJQTiVxAkr1omC-zDITkCNSwv2ptI3XMDPcRrddLnn4hgQ

Week 3

Who Counts in Administrative Data?: Coverage

Runes, Charmaine. *Urban Wire: Race and Ethnicity*, 2019, www.urban.org/urban-wire/following-long-history-2020-census-risks-undercounting-black-population .

Ethics

Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (Eds.). (2016). *Big data and social science: A practical guide to methods and tools*. crc Press. Chapter 12: Privacy and Confidentiality. <https://textbook.coleridgeinitiative.org/chap-db.html>

Goroff, D., Polonetsky, J., & Tene, O. (2018). Privacy protective research: Facilitating ethically responsible access to administrative data. *The ANNALS of the American Academy of Political and Social Science*, 675(1), 46-66. https://fpf.org/wp-content/uploads/2017/07/privacy_protective_research_-_facilitating_ethically_responsible_access_to_administrative_data.pdf

Week 4:

More on Validity

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company. Chapters 2 and 3 (Statistical Conclusion Validity and Internal Validity; Construct Validity and External Validity)

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

Mezey, N. (2002). Erasure and Recognition: The Census Race and the National Imagination. *Nw. UL Rev.*, 97, 1701.
<https://heinonline.org/HOL/LandingPage?handle=hein.journals/illlr97&div=47&id=&page=>

Prewitt, K. (2005). Racial classification in America: where do we go from here?. *Daedalus*, 134(1), 5-17.

https://www.jstor.org/stable/pdf/20027956.pdf?casa_token=ySx4rRQx5Y0AAAAA:7zUrYGi2S565sJqBG56zdj2NNGb54bS986jeNjCfS8L32wwT4IT7pANUt7IawY3nmMe-UrBnKF4EkwyzXSuKgBxuALE4fDLMNAHeqW6E22-1mDcOFt0n

Williams, D. R. (1999). The monitoring of racial/ethnic status in the USA: data quality issues. *Ethnicity & health*, 4(3), 121-137.

https://www.tandfonline.com/doi/pdf/10.1080/13557859998092?casa_token=dNyb-J3lNY4AAAAA:jGjWgn4A6lljNNAG8dwxfrhW-4K-5tDuPRE4aPdcKfcxVKfySi-3AKGXvQeu7Z7OYQrKtzH-MHkNsA

Week 5

Integrating data to describe places

Cooper, H. L., Linton, S., Kelley, M. E., Ross, Z., Wolfe, M. E., Chen, Y. T., ... & Semaan, S. (2016). Racialized risk environments in a large sample of people who inject drugs in the United States. *International Journal of Drug Policy*, 27, 43-55.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4715941/pdf/nihms720813.pdf>

Probabilistic Matching

Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (Eds.). (2016). *Big data and social science: A practical guide to methods and tools*. crc Press. Chapter 3: Record Linkage

<https://textbook.coleridgeinitiative.org/chap-db.html>

Capture/Recapture Methods

Jouanjus, E., Pourcel, L., Saivin, S., Molinier, L., & Lapeyre-Mestre, M. (2012). Use of multiple sources and capture–recapture method to estimate the frequency of hospitalizations related to drug abuse. *Pharmacoepidemiology and drug safety*, 21(7), 733-741.

https://onlinelibrary.wiley.com/doi/pdf/10.1002/pds.3280?casa_token=cCLEMbaAfcIAAAAAA:GRfifA-fNdutND_dM1Mx3Vm6WqDzMicVqP_FaNINK2hCUPwYPk114GZbJV-ML8eBcfjF8PvbAk28I-E

Hickman, M., Cox, S., Harvey, J., Howes, S., Farrell, M., Frischer, M., ... & Tilling, K. (1999). Estimating the prevalence of problem drug use in inner London: a discussion of three capture–recapture studies. *Addiction*, 94(11), 1653-1662.

https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1360-0443.1999.941116534.x?casa_token=evyyLEoGNDYAAAAA:i7KCB5O2_1YLIU5zGxcM0jGX52CXOvdx54PyugVb8BwPDaxd_bVzAauk60mFYHUjNJ5NeB4wejQGAuU

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

Date	Topic	Instructor
Module II Week 1		
Class session	Theory of local influences on drug-related health outcomes <ul style="list-style-type: none"> • The Risk Environment Model • Maps and health • Neighborhoods and health 	Lance Waller/ Hannah Cooper
Lab	<ul style="list-style-type: none"> • Reading data into ArcGIS • Making a map in ArcGIS 	Lance Waller
Module II Week 2		
Class session	Cartography and communication <ul style="list-style-type: none"> • Effective display of geographic information • Communicating with maps • Ethical issues in analyzing geospatial data for drug-related research 	Lance Waller
Lab	<ul style="list-style-type: none"> • Layering and joining multiple data sets 	Lance Waller
Module II Week 3		
Class session	Maps and theories of substance use <ul style="list-style-type: none"> • Maps of administrative data • Maps of substance use • Scale, aggregation, and the modifiable areal unit problem 	Lance Waller
Lab	<ul style="list-style-type: none"> • Access to services: Buffering locations 	Lance Waller
Module II Week 4		
Class session	Local risk and resiliency <ul style="list-style-type: none"> • Risk and resilience factors • Vulnerability indices • Spatial regression 	Lance Waller
Lab	<ul style="list-style-type: none"> • Case study work time 	Lance Waller
Module II Week 5		
Class Session	Geographic analysis <ul style="list-style-type: none"> • Building a geographic analysis • GIS for processing data • R for analyzing data 	Lance Waller
Lab	<ul style="list-style-type: none"> • Case study work time 	Lance Waller

REQUIRED READINGS (Available on Canvas)

- **Week 1:**
 - Gorman DM, Gruenewald PJ, and Waller LA (2013) Linking places to problems: Geospatial theories of neighborhoods, alcohol and crime. *GeoJournal* 78; 417-428.
- **Week 2:**

- Duncan DT, Regan SD, Chaix B (2018) Operationalizing neighborhood definitions in health research. In *Neighborhoods and Health, Second Edition*, Duncan and Kawachi (eds). New York: Oxford University Press, pp. 19-56.
- **Week 3:**
 - Gruenewald P (2007) The spatial ecology of alcohol problems: niche theory and associative drinking. *Addiction* **102**, 870-878.
 - Gruenewald P. J., Freisthler B., Remer L., LaScala E. A., Treno A. (2006) Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis. *Addiction*; 101: 666–77.
- **Week 4:**
 - Duncan DR, Goedel WC, and Chunara R (2018) Quantitative methods for measuring neighborhood characteristics in neighborhood health research. In *Neighborhoods and Health, Second Edition*. Duncan and Kawachi (eds). New York: Oxford University Press, pp. 57-90.
 - Gruenewald P. J., Freisthler B., Remer L., LaScala E. A., Treno A. (2006) Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis. *Addiction*; 101: 666–77.
- **Week 5:**
 - Gruenewald PJ, Ponicki WR, Remer LG, Waller LA, Zhu L, and Gorman DM (2013) Mapping the spread of methamphetamine abuse in California from 1995 to 2008. *AJPH* **103**: 1262-1270.
 - Gruenewald P (2007) The spatial ecology of alcohol problems: niche theory and associative drinking. *Addiction* **102**, 870-878.

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

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

Recommended Textbooks

1. Wasserman, Stanley and Katherine L. Faust. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.
2. Lusher, D., Koskinen, J. & Robins, G. 2013. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press.
3. Valente, Thomas W. 2010. *Social Networks and Health: Models, Methods, and Applications*. Oxford University Press.

REQUIRED READINGS (Available on Canvas)

o. Introduction

- Pescosolido, Bernice A. 2006. “Of Pride and Prejudice: The Role of Sociology and Social Networks in Integrating the Health Sciences.” *Journal of Health and Social Behavior* 47(3): 189-208.
- Thoits, Peggy A. 2011. “Mechanisms Linking Social Ties and Support to Physical and Mental Health.” *Journal of Health and Social Behavior* 52(2): 145-161.
- Valente, Thomas W., Peggy Gallaher, and Michele Mouttapa. 2004. “Using Social Networks to Understand and Prevent Substance Use: A Transdisciplinary Perspective.” *Substance Use & Misuse* 39: 1685-1712.

1. Network Data

Brewer, Devon and Cynthia Webste. 1999. "Forgetting of Friends and its Effects on Measuring Friendship Networks." *Social Networks* 21: 361-373.

Marsden, Peter V. 2005. "Recent Developments in Network Measurement." Pp. 8-30 in *Models and Methods in Social Network Analysis*. New York: Cambridge University Press.

2. Random Network Models

Wimmer, Andreas, and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* 116:583-642.

An, Weihua and William McConnell. 2015. "The Origins of Asymmetric Ties in Friendship Networks: From Status Differential to Self-Perceived Centrality." *Network Science* 3(2): 269-292.

3. Network Effects

An, Weihua. 2015. "Instrumental Variables Estimates of Peer Effects in Social Networks." *Social Science Research* 50: 382-394.

Cornwell, Benjamin. 2009. "Good Health and the Bridging of Structural Holes." ***Social Networks* 31:92- 103.**

Bearman, Peter S., James Moody and Katherine Stovel. 2004. "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks." *American Journal of Sociology* 110: 44-91.

4. Dynamic Network Analysis

Christakis, Nicholas A. and James H. Fowler. 2007. "The Spread of Obesity in a Large Social Network Over 32 Years." *New England Journal of Medicine* 357(4): 370-379.

Steglich, Christian, Tom A.B. Snijders, and Michael Pearson. 2010. "Dynamic Networks and Behavior: Separating Selection from Influence." *Sociological Methodology* 40(1): 329-393.

5. Network Interventions

Valente, Thomas W. 2012. "Network Interventions." *Science* 337: 49-53.

Valente, Thomas W. 2010. "Network Interventions." Chapter 11 in *Social Networks and Health: Models, Methods, and Applications*. Oxford University Press.

Paluck, E. L., H. Shepherd, and P. M. Aronow. 2016. "Changing Climates of Conflict: A Social Network Experiment in 56 Schools." *PNAS* 113(3): 566-571.

Centola, Damon. 2011. "An Experimental Study of Homophily in the Adoption of Health Behavior." *Science* 334: 1269-1272.

An, Weihua. 2015. "Multilevel Meta Network Analysis with Application to Studying Network Dynamics of Network Interventions." *Social Networks* 43: 48-56.