

Contextual data in Add Health

Lauren Gaydos and Iliya Gutin
University of Texas at Austin

Add Health Users Conference
July 11, 2022

Outline

- Brief history of contextual data efforts
- Existing resources
- Longitudinal considerations in contextual data analysis
- Future opportunities

A brief history

Existing Resources

Core contextual data

Ancillary studies – neighborhoods and schools

Core contextual data

- Merged at all Waves (I – V) and planned for Wave VI
- At the “local” level – county, neighborhood, Census tract and block group
- ~7000 variables
- Three sources:
 1. US Census (Decennial, CPS, ACS)
 2. Uniform Crime Report
 3. Climate atlas

US Census



- From Decennial Census, Current Population Survey, and American Community Survey
- Race/ethnicity, age, household and family type
- Linguistic isolation, foreign born
- Educational attainment, income, labor force participation, poverty status
- Industries, occupations, transportation to work
- Housing characteristics

Uniform Crime Report



- Total adult and juvenile arrests
- Property crime arrests
- Violent crime arrests
- Arrests and crimes by age and type

Climate Atlas



- Days $<32^{\circ}\text{F}$ and $>90^{\circ}\text{F}$
- Average min and max daily temperatures
- Average precipitation
- Average snowfall
- Rural-Urban Commuting Area Codes

Ancillary studies

Neighborhood

- Vital statistics
- ONE
- Political context
- Food environment
- Air pollutants
- Sexual Minority Policy
- Health context
- Income mobility

School

- Transcript
- Wave III EduContext
- Wave IV IPEDS
- Mobility report cards

Neighborhoods

Vital Statistics



Waves I - III

- Marriage rates
- Fertility rates
- Migration
- Nativity
- Mortality
- Disability

Obesity and Neighborhood Environment (ONE)



Waves I and III physical, social, and environmental measures

- ACCRA cost of living index
- Land cover, parks, length of day
- Road type, road connectivity
- Population density
- Physical activity and diet resources

Political context database



Waves I - III

- Election results – presidential, gubernatorial, senatorial
- Voter registration laws
- Turnout rate

Modified Retail Food Environment Index



Wave IV

- From CDC Division of Nutrition, Physical Activity, and Obesity
- The mRFEI represents the percentage of healthy food retailers relative to less healthy food retailers within each census tract

Ambient Air Pollutants



Wave IV

- Community Multiscale Air Quality model runs
- Ambient air pollutants, individual pollutants, daily particulate matter, toxic gas estimates

Sexual minority policy

Waves III and IV

- Human Rights Campaign and American Civil Liberties Union
- State prohibits employment discrimination based on sexual orientation;
State has hate crime statutory provisions based on sexual orientation;
State allows same-sex marriage/domestic partnership/civil union/reciprocal beneficiary relationships;
State allows same-sex joint adoption and/or second-partner adoption

Health context



Waves I, IV and V

- IHME:
Life expectancy and age-specific mortality;
Diabetes prevalence, awareness, control;
Alcohol use, binge drinking;
Smoking; Obesity; Physical activity
- State-level cigarette tax per pack
- Health insurance coverage
- RWJF county health rankings:
Health outcome and factor ranking quartiles

Income mobility



Waves I, IV, and V

From Opportunity Insights

- Relative and absolute mobility – total and by race, gender, and maternal nativity status
- Causal effect of county of childhood residence on adult household income at p25
- Income Gini coefficient
- Probability of reaching top 1% of income distribution – by race and gender
- Probability of living in census tract with poverty rate <10% – by race and gender

Schools

Secondary school educational context



From Adolescent Health and Academic Achievement Study

- Common Core of Data: free lunch, location, student/teacher ratio
- Private School Survey
- NCES: district size, expenditures
- Attached to school/district of the participating Add Health schools

College Mobility Report Cards



From Opportunity Insights, attached to college enrolled (Wave III) or graduated (Wave IV)

- Selectivity, SAT scores
- Parental income
- Adult income
- Mobility rate
- Cost
- Endowment, expenditures
- Graduation rate
- Student demographics
- Major shares

IPEDES College characteristics



From IPEDES, attached to college enrolled (Wave III) or graduated (Wave IV)

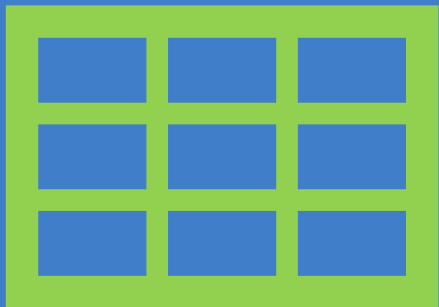
- Type of institution
- Cost – tuition, room and board
- SAT score percentile
- Size
- Student demographics – race/ethnicity, sex, financial aid
- Academic and professional services

Data documentation and access

- Explore data using online [Navigator](#)
- Read [codebook](#) documentation
- [Public use](#): ICPSR, ARDA, Odum Institute
- [Restricted use](#): Add Health Contract

Longitudinal Considerations in Contextual Data Analysis

Challenges with Contextual Data : Preparing



Identify

- What is **already available** in Add Health
- **Public vs. proprietary/restricted** sources
- Organizations/institutions requiring **data use agreements and/or proposals**

Define "Space"

- Manipulate raw data to operationalize **proximity, density, dissimilarity, etc.**
- **GIS** software and methods (and/or expertise)

Define "Time"

- **Point/Place-in-time** associations
- Moving **averages** or "**lagged**" influences
- **Longer trends**, independent of respondents

Challenges with Contextual Data : Processing



Format

- Organize data to **clearly distinguish** different contextual factors over waves/time (e.g., *Non-white population*_{CountyWave V/2018/2016-2020})
- Account for **longitudinal grouping id**

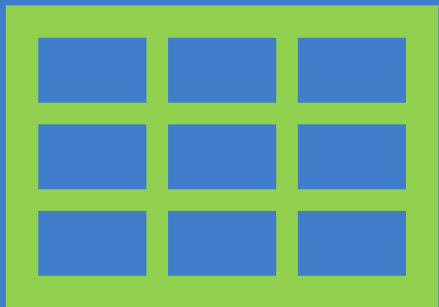
Merge

- **Merge datasets** based on context
- Calculate harmonized measures **across multiple levels of context**

Match

- Assign tracts and counties to all respondents with geocoded data **in secure facility**
- Match respondents to contextual measures **across multiple Waves**

Challenges with Contextual Data : Dissemination



Deductive Disclosure

- Analyze data for **small cell counts**
- **Merge categories** when necessary

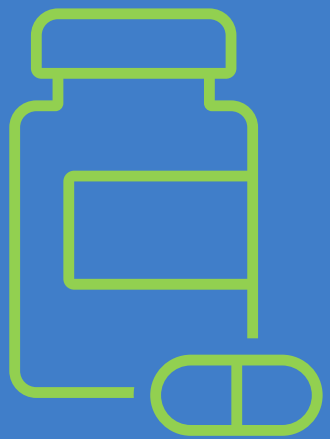
Release

- What is **releasable?** (E.g., DUAs)
- Create **documentation**
- **Disseminate** to users

Update

- As **new data become available** from both the source(s) and the study
- **Harmonize** across Waves
- **Changing** identifiers and boundaries

Ancillary Study Example

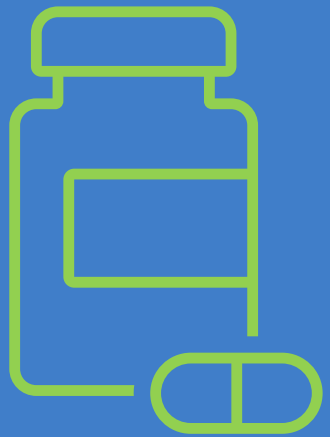


Contextual Despair and Risk Behaviors in Midlife: Extending Innovative Measures to Add Health

Lauren Gaydosch, Iliya Gutin, Tse-Chuan Yang

- Motivation and research objectives
- Novel contextual measures
- Thinking *longitudinally* about context

Ancillary Study Example



Motivation

Declining life expectancy, potentially explained by rising midlife mortality from drug overdose, alcohol-related diseases, and suicide (a.k.a., “**deaths of despair**”)



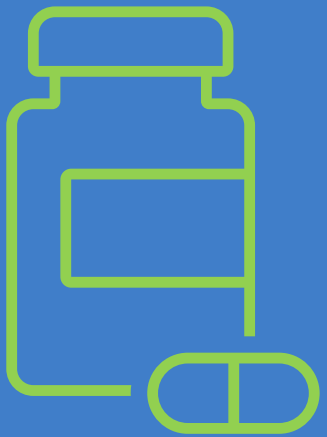
Middle-Aged White Americans Are Dying of Despair

Even as longevity increases across the rich world, uneducated white Americans are living sicker and dying earlier. Two economists speculate on the reasons why.



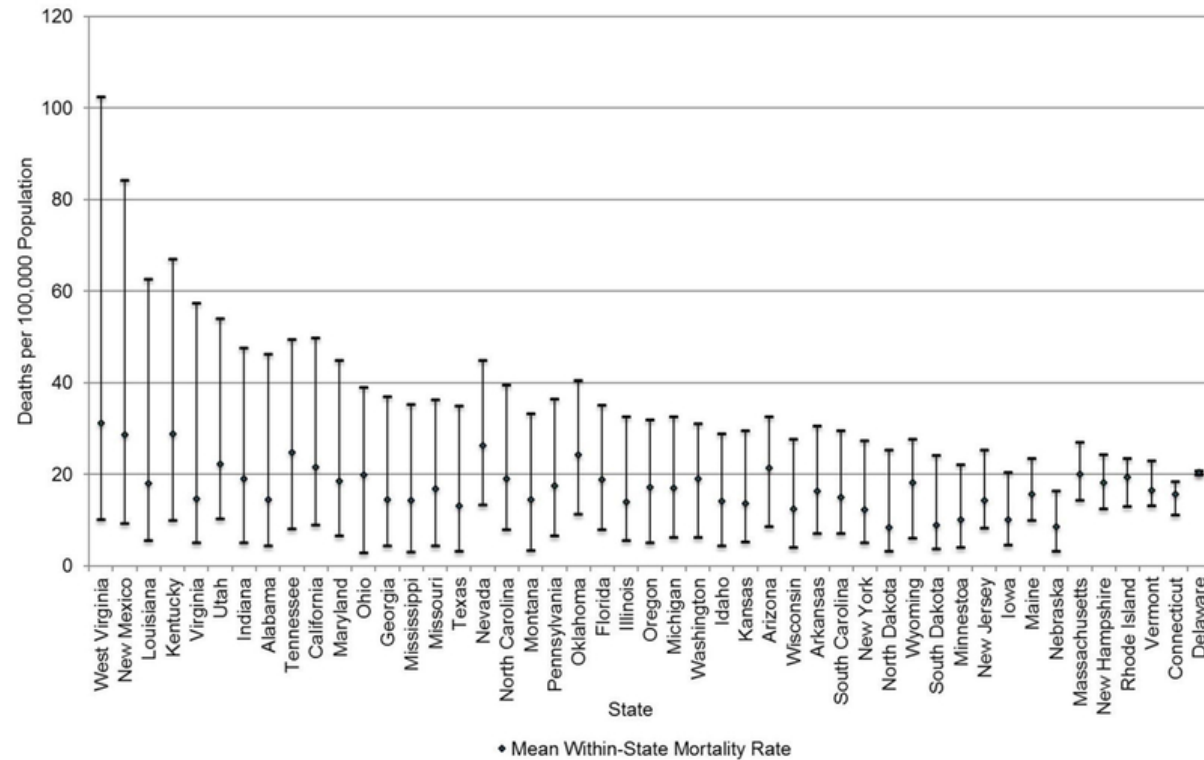
Case and Deaton 2015

Ancillary Study Example



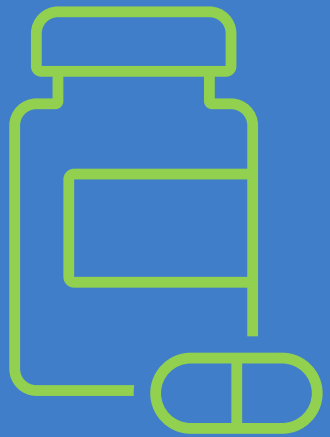
Motivation

Despair mortality exhibits wide variation based on “place”, suggesting **geographic context matters** in understanding social etiology of declines in life expectancy



Monnat 2018

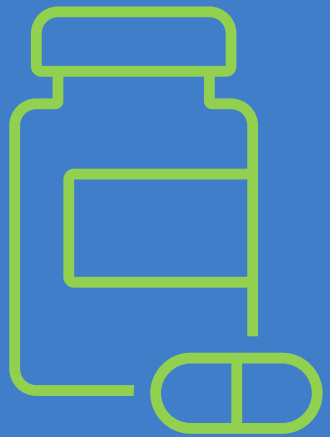
Ancillary Study Example



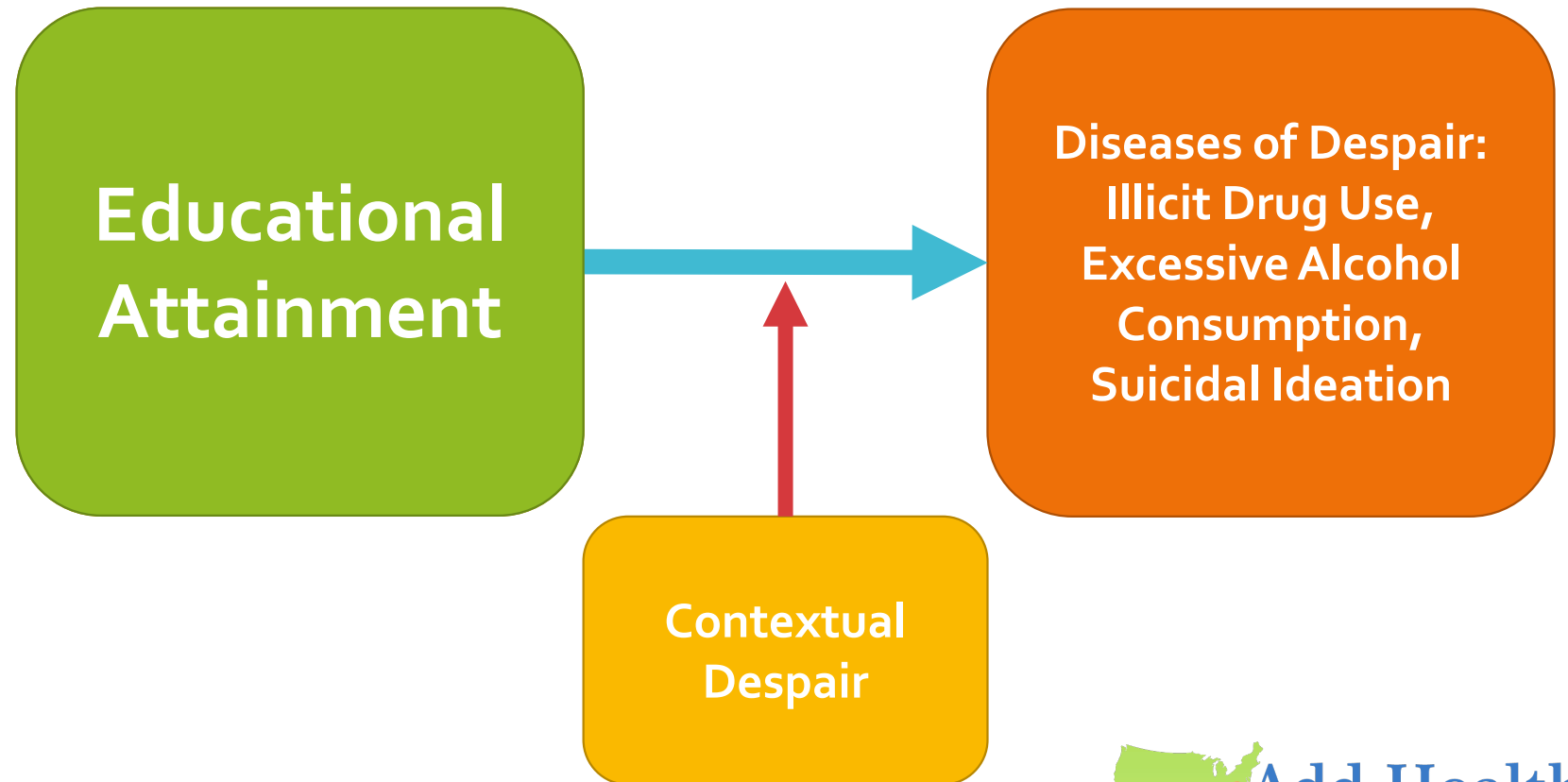
What can individual-level, longitudinal data help explain?

- Individuals embedded in places over time helps to **distinguish between effects of individual versus community factors** (i.e., avoiding the ecological fallacy)
- Associations with “**precursor**” risk behaviors (i.e., “diseases” of despair) earlier in life
- Are deaths of despair “**contagious**”?

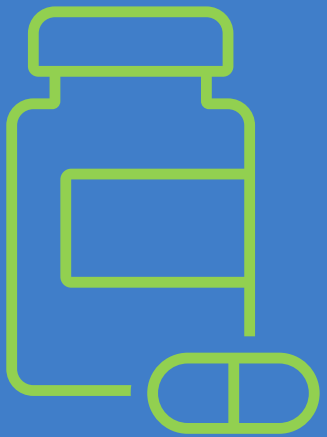
Ancillary Study Example



Research Objectives



Ancillary Study Example



Novel Contextual Measures

Socioeconomic

- Industrial composition
- Local housing market
- Social decline (e.g., social isolation, social capital index, community membership)
- Crime rates

Built Environment

- Opioid-dispensing per capita
- Medication assisted treatment programs
- Opioid treatment programs
- Distance/density of firearms dealers and businesses selling alcohol

Population Health

- Percent of adults with kidney problems and COPD
- Poor physical and mental health days
- Depression
- Self-rated health
- Preventable hospital stays
- Cause-specific mortality

Ancillary Study Example



Thinking *Longitudinally* About Context



Ancillary Study Example



Thinking *Longitudinally* About Context

$$\log\left(\frac{\varphi_{ij}}{1 - \varphi_{ij}}\right) = \gamma_{00} + \sum \gamma_{0l} \mathbf{w}_{lj} + \sum \beta_{kj} x_{ijk} + u_{0j} + r_{ij}$$

\mathbf{w}_{lj} indicates the feature l of j neighborhood (e.g., social isolation or distance to firearms dealer) and γ_{0l} refers to the estimated association of neighborhood characteristics l with the dependent variable

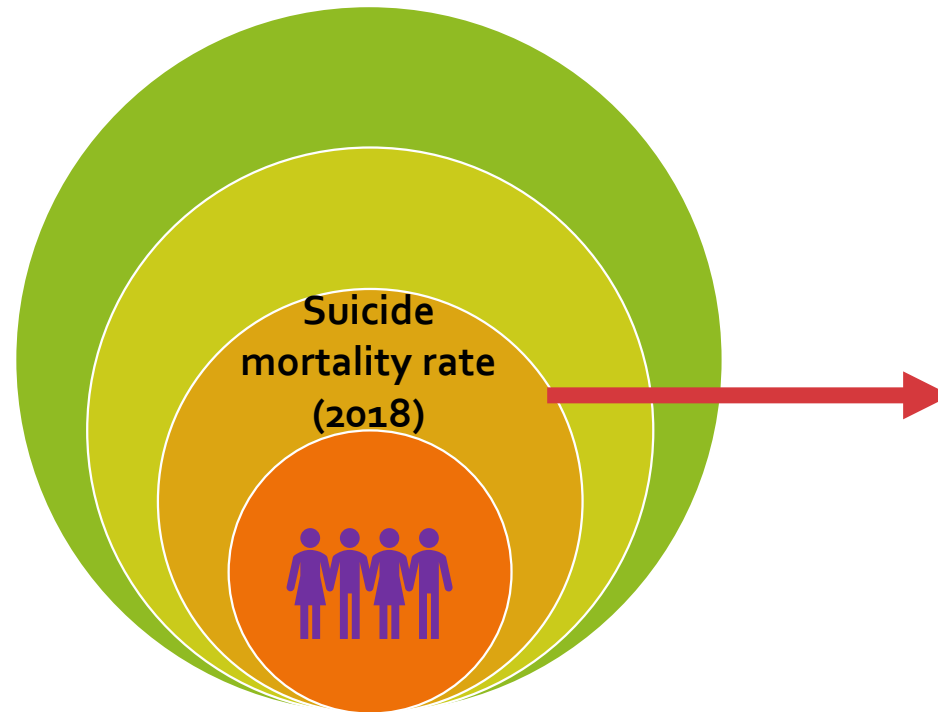
- Point-in-time measures
- Lagged measures or (moving) averages
- Trends “outside” the study/observation window

Ancillary Study Example



Point-in-time Measures

Attributes of contexts that have “immediate” (or proximate) associations with outcomes



Wave V County of Residence



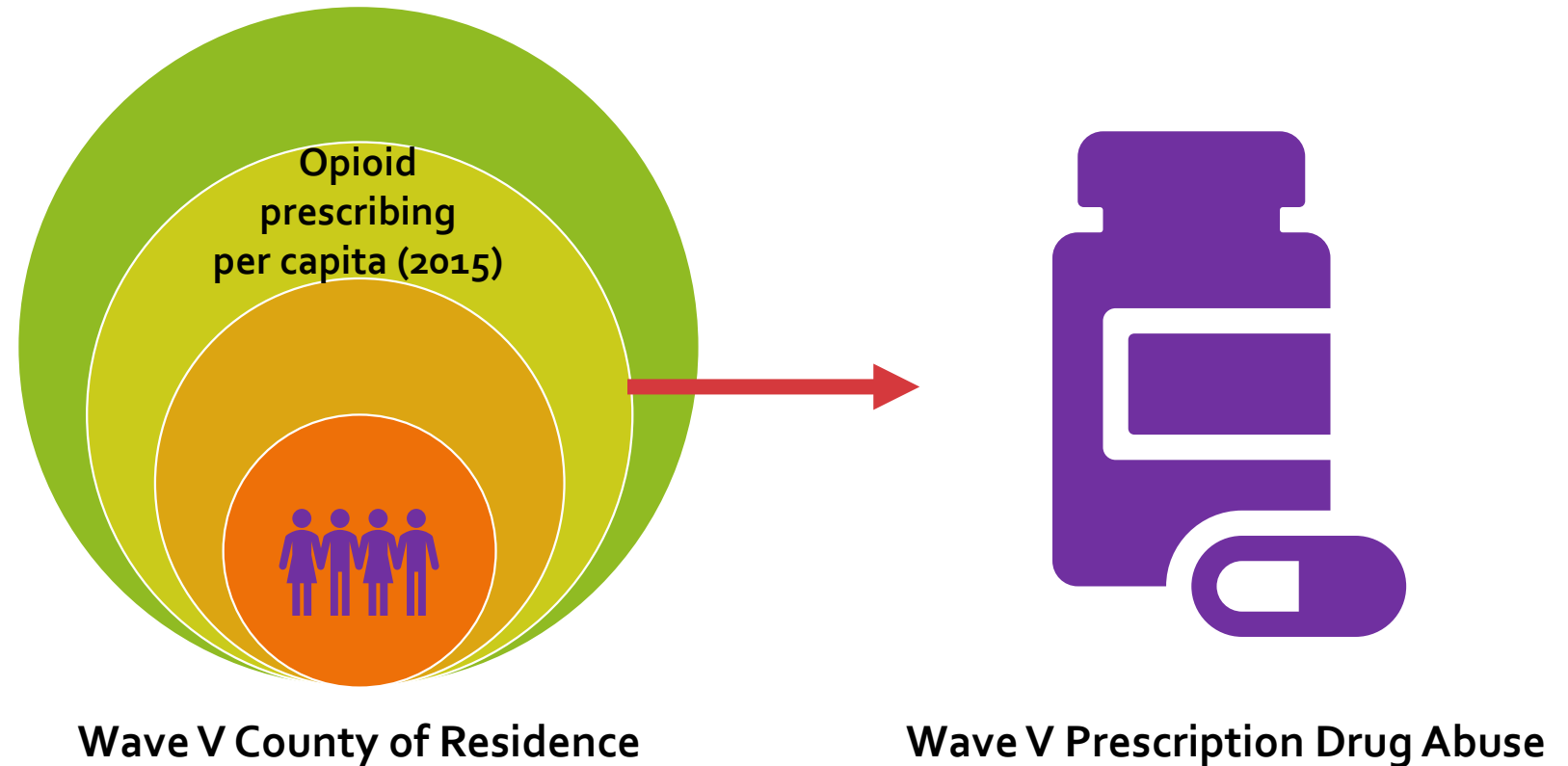
Wave V Suicidal Ideation

Ancillary Study Example



Lagged Measures or Averages

Attributes of contexts that have “**delayed**” (or staggered) associations with outcomes

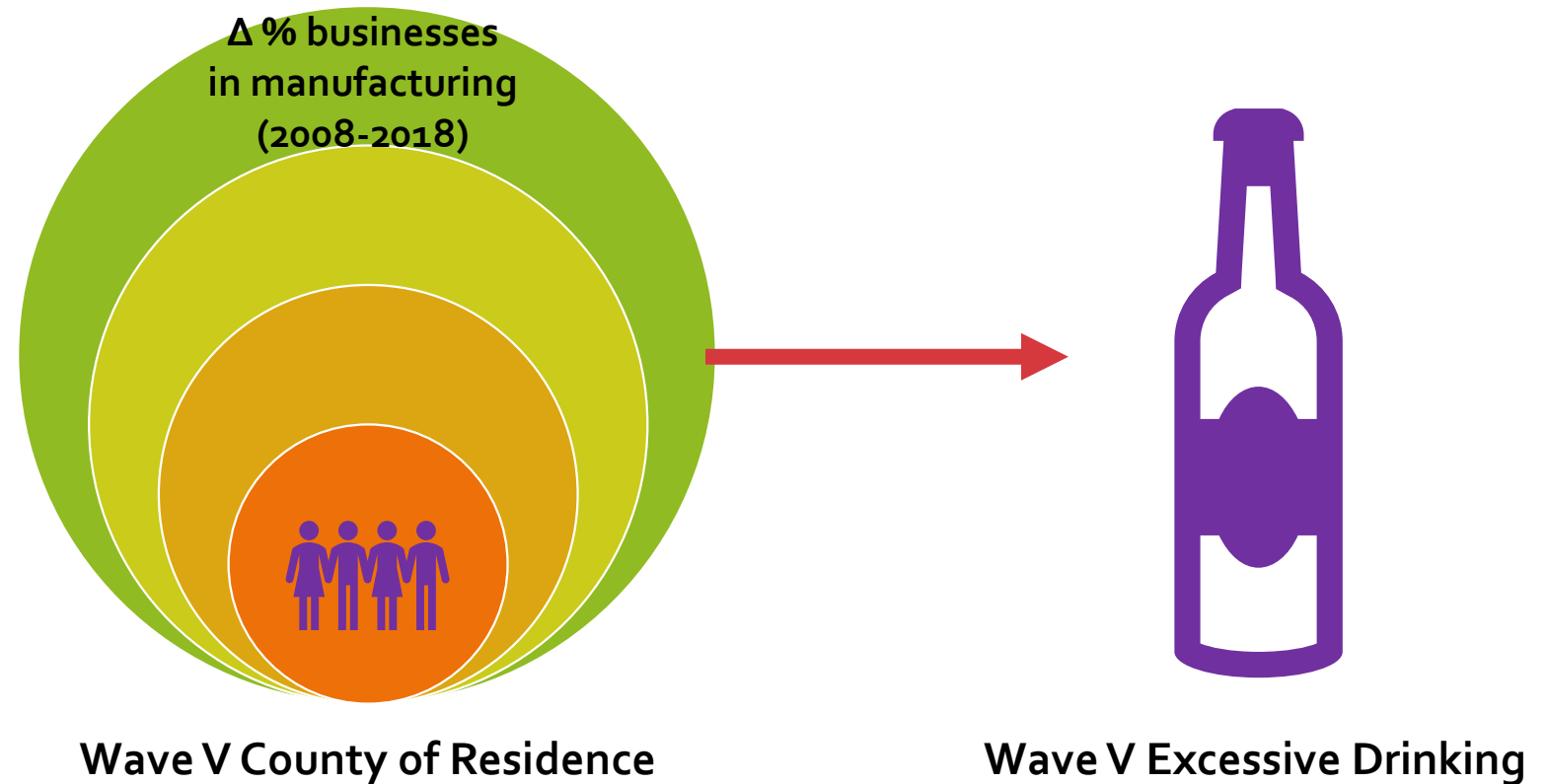


Ancillary Study Example



Long(er)-term Trends

Attributes of contexts that have “long-lasting” (or cumulative) associations with outcomes



Ancillary Study Example



Approach Depends on Motivation and Theory (and Imagination?)

- **More than one** “longitudinal” feature of individuals’ context is likely to be consequential
- **Individuals can move across contexts** characterized by different combinations of longitudinal contextual factors
- Can use additional methods – e.g., LCA, group-based trajectory modeling – to **define a “profile” of the longitudinal context**, and examine associations with outcomes
- There is a lot to consider, conceptually, **independent of modeling and analysis**

Future Opportunities

Plans for Wave VI

How to submit an ancillary study application

Plans for Wave VI

- Core contextual data merge
- Health context
- Income mobility

Goals for Wave VI

- Structural racism
- Structural xenophobia
- Structural sexism
- Structural heterosexism
- Health inequality

Structural racism

- Measures of structural racism
 - Underrepresentation in legislature – National Conference on State Legislatures
 - Disproportionate disenfranchisement – The Sentencing Project
 - Incarceration – Vera Institute of Justice
 - Historical information on redlining and lynching
 - Police brutality – police-involved killings

Structural xenophobia

- Measures of structural xenophobia
 - Immigration enforcement – ICE detentions, arrests, and removals
 - Restrictive immigration policies

Structural sexism

- Measures of structural sexism
 - Underrepresentation in legislature – National Conference on State Legislatures
 - Reproductive healthcare availability and policy

Structural heterosexism

- Measures of structural heterosexism
 - Political and legal climate
 - Sociodemographic characteristics
 - Socioeconomic disparities

Health inequality

- Measures of health inequality
 - Health policy
 - Health disparities
 - COVID-19

Your own ancillary study



An Ancillary Study is any study that derives support from independent funds outside the Add Health Study, and does one or more of the following:

- Merges secondary data sources to Add Health respondent records that requires unique identifiers (e.g., geocodes) to perform these linkages

Your own ancillary study



General Requirements For Add Health Ancillary Studies

Ancillary Study investigators must meet the following criteria:

- Have a PhD, MD, or other terminal degree.
- Hold a faculty appointment or research position at their institution.
- Work for an institution of higher education, a research organization, or a government agency.
- Have an institutional review board (IRB) that complies with applicable Federal regulations governing research involving human subjects.
- Demonstrate completion of research ethics training by all research team members who will work with the Add Health data or biospecimens.
- Have a demonstrated record of using sensitive data according to commonly accepted standards of research ethics.
- Investigators proposing to conduct an Ancillary Study must cover all costs incurred by the study, such as: sample selection; collecting or pulling samples from archive; processing and shipping biospecimens; preparing and documenting analysis files; integrating ancillary data into the Add Health Study; and archiving leftover biospecimens. Some of these activities can only be performed by the Add Health staff and/or the Add Health archive lab, which must be paid for by the Ancillary Study.

Your own ancillary study



Steps:

1. Submit a brief (1-page) Concept Proposal for Add Health review
2. Address feedback and resolve issues from preliminary review
3. Submit the [Add Health Ancillary Study Proposal Online Form](#)
4. Work with Ancillary Studies Coordinator to develop a cost estimate

Acknowledgements

This research was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R21HD095448. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health

This research is supported by the Network on Life Course Health Dynamics and Disparities in 21st Century America (NLCHDD) via grant # R24AG045061 from NIA. The NLCHDD received additional support for its annual meetings from the Michigan Center on the Demography of Aging (MiCDA) via NIA grant # P30AG012846

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>).

Questions and discussion

lauren.gaydosh@austin.utexas.edu

iliya.gutin@austin.utexas.edu

EXTRA WHERE TO PUT

- Abortion access, healthcare, STDs
- Tobacco control
- Politics – elections, government expenditures, social welfare programs