

Add Health Data Dissemination: Challenges, Procedures and Successes

Kathleen Mullan Harris

University of North Carolina at Chapel Hill

Meeting of C-StARR, Duke University

August 25, 2014



Topics to cover

- Design of Add Health Study
- Data released and user base
- Security risks in data dissemination and user access
- Data sharing plan
- Successes and ongoing challenges
- Ancillary studies

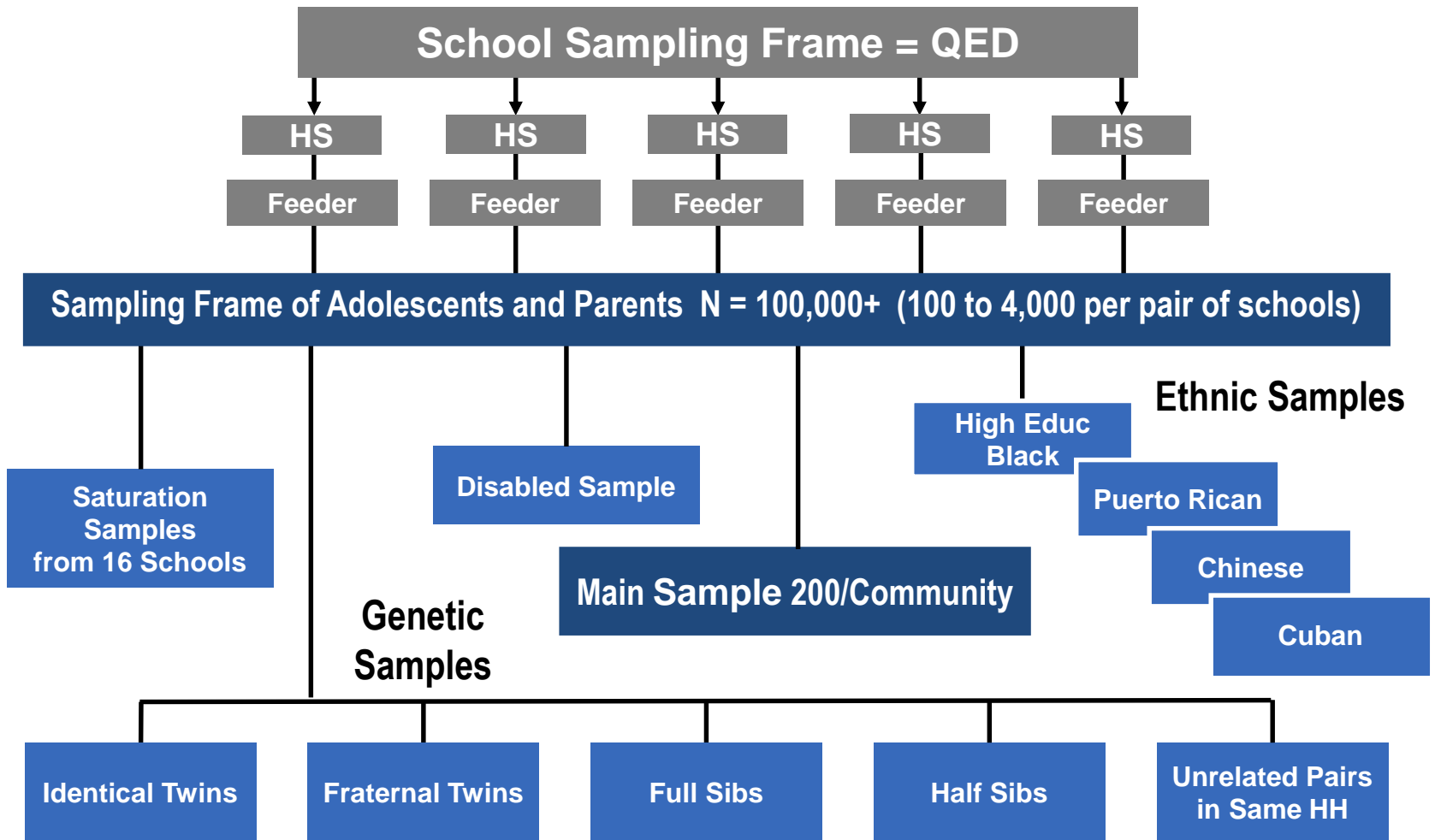
Add Health

- National Longitudinal Study of Adolescent to Adult Health
- On-going program project that began in 1994 funded by NIH.
- Developed in response to a congressional mandate to fund a study of adolescent health.
- Mandate to understand the role of social environments in which adolescents live.

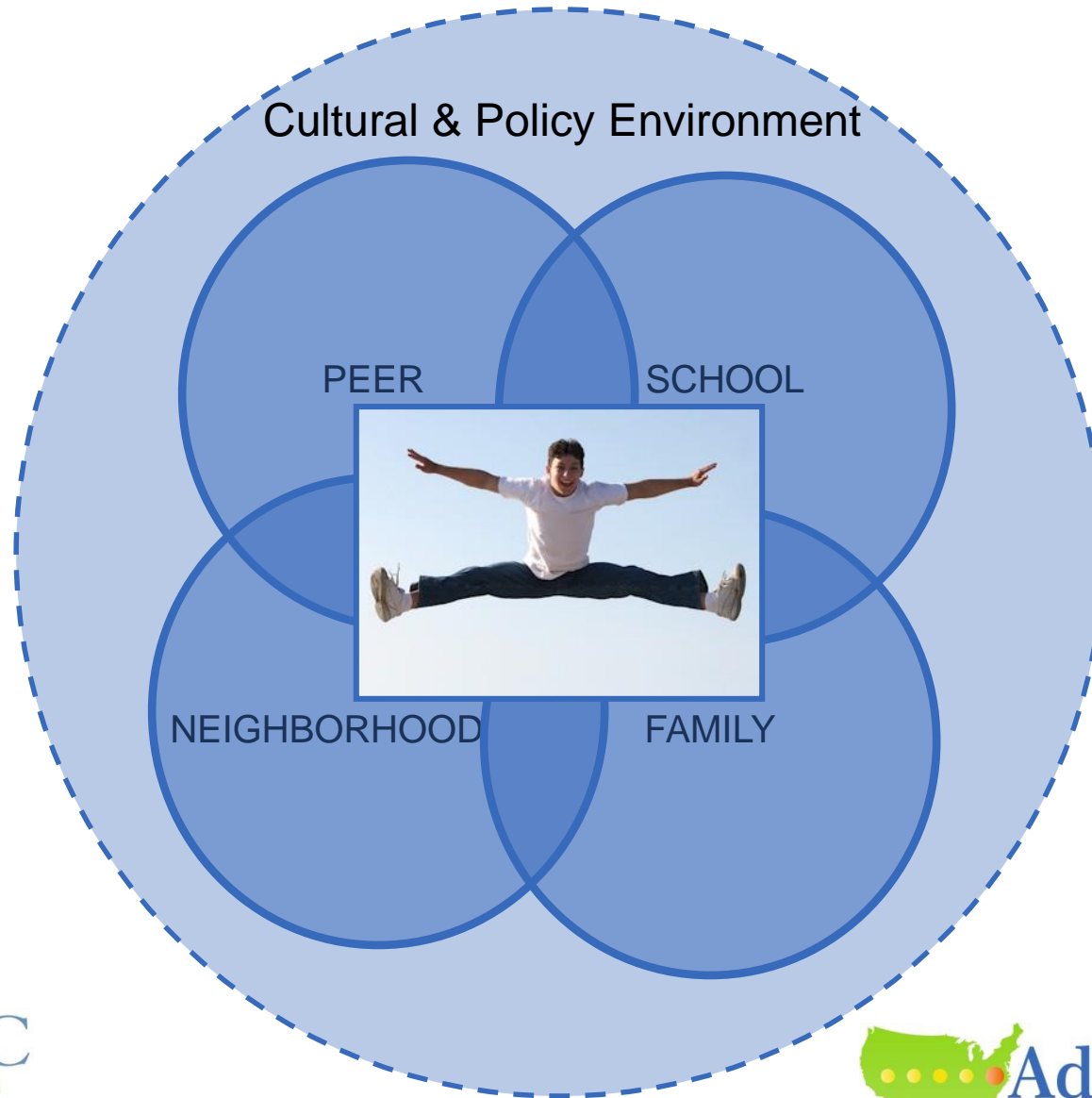
Key Features of Add Health Design

- Nationally representative study that explores the cause of health and health-related behaviors of adolescents and their outcomes in adulthood.
- Multi-survey, multi-wave inter-disciplinary design.
- Direct measurement of the social contexts of adolescent life and their effects on health and health behavior.
- Unprecedented racial and ethnic diversity and genetically informed sibling samples.
- Extensive biomarker collection across all waves.

Sampling Structure



Add Health Contextual Model



Race and Ethnic Diversity in Add Health

Race/Ethnicity	N	%
Mexico	1,767	8.5
Cuba	508	2.5
Central-South America	647	3.1
Puerto Rico	570	2.8
China	341	1.7
Philippines	643	3.1
Other Asia	601	2.9
Black (Africa/Afro-Caribbean)	4,601	22.2
Non-Hispanic White (Eur/Canada)	10,760	52.0
Native American (non-Hispanic)	248	1.2
Total N	20,686	100.0

Family Structure of Adolescents Add Health 1995

	N	%
2 biological parents	10,339	53.3
2 adoptive parents	403	0.7
Bio Mom/ Step Dad	2,756	13.6
Bio Dad/ Step Mom	591	2.6
Single Mom	4,520	20.4
Single Dad	637	3.1
Surrogate parent(s)	1,499	6.3
Total	20,745	100.0

Strong Motivation for Data Sharing

- Add Health very expensive;
- National representation;
- Design provided unprecedented and unique opportunities for research that no other study allowed;
- Multi-disciplinary design;
- Omnibus study with comprehensive coverage of health and health behavior in adolescence;
- Longitudinal study to help sort out cause and effect.
- Add Health policy: No proprietary period for investigators.

Special Data Security Concerns to Address in Add Health Data Sharing Plan

- Contextual nature of design;
- Third-party nominations of friends;
- Third-party nominations of romantic and sex partners:
 - No consent from nominated partners
- Geocodes of adolescent's home address;
- Sensitive nature of data, especially during adolescence;
- With each additional wave of longitudinal data and new data sources added, data security risks increase.

Pledge of Confidentiality to Add Health Participants

Add Health Wave I Consent form for Adolescent,
1994-95.

- “I understand that my statements and answers will be completely protected so that no one will be able to connect my answers to me. My answers will not be given to my parents/guardians or to any unauthorized person by project staff.”

Two Major Security Risks to Privacy and Confidentiality of Add Health Respondents

- Direct disclosure
 - link between name and questionnaire information.
- Deductive disclosure
 - Discerning a respondent's identity and their responses through use of known characteristics of that person;
 - Add Health especially high risk of deductive disclosure because of the clustered study design;
 - Almost 400,000 people knew of the participation of at least one, if not more, the adolescents in Add Health at Wave I.

For Example

- In more than 90,000 cases in in-school administration, takes only 5 variables to distinguish as few as 7 records:
- Number of respondents = 90,118
 - Gender = female (44,482)
 - Age = 16 (8,250)
 - Race = Asian (570)
 - Does not live with father figure (104)
 - Participates in chorus (7)

Data Security Procedures

- Add Health Data Management Security Plan:
- Protects against breach of confidentiality from both direct and deductive disclosure;
- Protects from subpoena (or hostile sharing).
- Also have a certificate of confidentiality

Protections from Direct Disclosure

- Based on the separation of identities and data.
- Names and addresses of respondents reside with a security manager (i.e., honest broker) outside U.S.
- Identifying information only in U.S. during data collection.
- Multiple identification numbers used for each wave of data collection.
- Respondent names never connected to data.

Protections from Deductive Disclosure

- Tiered data dissemination plan:
 - Differs by level of risk of deductive disclosure;
 - Requirements for access and use of data differ by the level of risk;
 - Add Health users must meet these requirements and assurances for safeguarding Add Health data.
- Users also required to abide by procedures for the dissemination of research results.
- Access to all data via restricted data use contract.

Four tier data dissemination according to disclosure risks

- Public-use data
- Restricted-use data
- High-security restricted data
- Secure data facility for analyzing high school transcript data and for using geocodes to link contextual data.
- [Secure server for access to genome-wide data.]

Add Health Dissemination Plan

	Public Use Data^a	Restricted Data	High Security Restricted Data	Secure Data Facility Use Data
Data Types				
Direct Identifiers	-	-	-	-
Geographic Coordinates	-	-	-	Yes
Neighborhood Characteristics	Yes ^b	Yes ^b	Yes ^b	Yes ^b
Individual Characteristics	Yes	Yes	Yes	Yes
Sensitive Behavior	Yes ^b	Yes	Yes	Yes
Procedural Requirements				
Read and agree to confidentiality guidelines	Yes	-	-	-
Institutional Contractual Agreement	-	Yes	Yes	Yes
Brief research proposal	-	Yes	Yes	Yes
Data security plan	-	Yes	Yes	Yes
Signed security pledges	-	Yes	Yes	Yes
Institutional Review Board approval	-	Yes	Yes	Yes
Site Visit by Add Health Staff	-	Yes	Yes ^c	-
Fee	Yes	Yes	Yes	Yes

^a Public Use Data distributed and managed by Sociometrics Inc.

^b Some of the characteristics treated for disclosure protection or are not included.

^c An annual site visit is required for high security restricted data contracts.

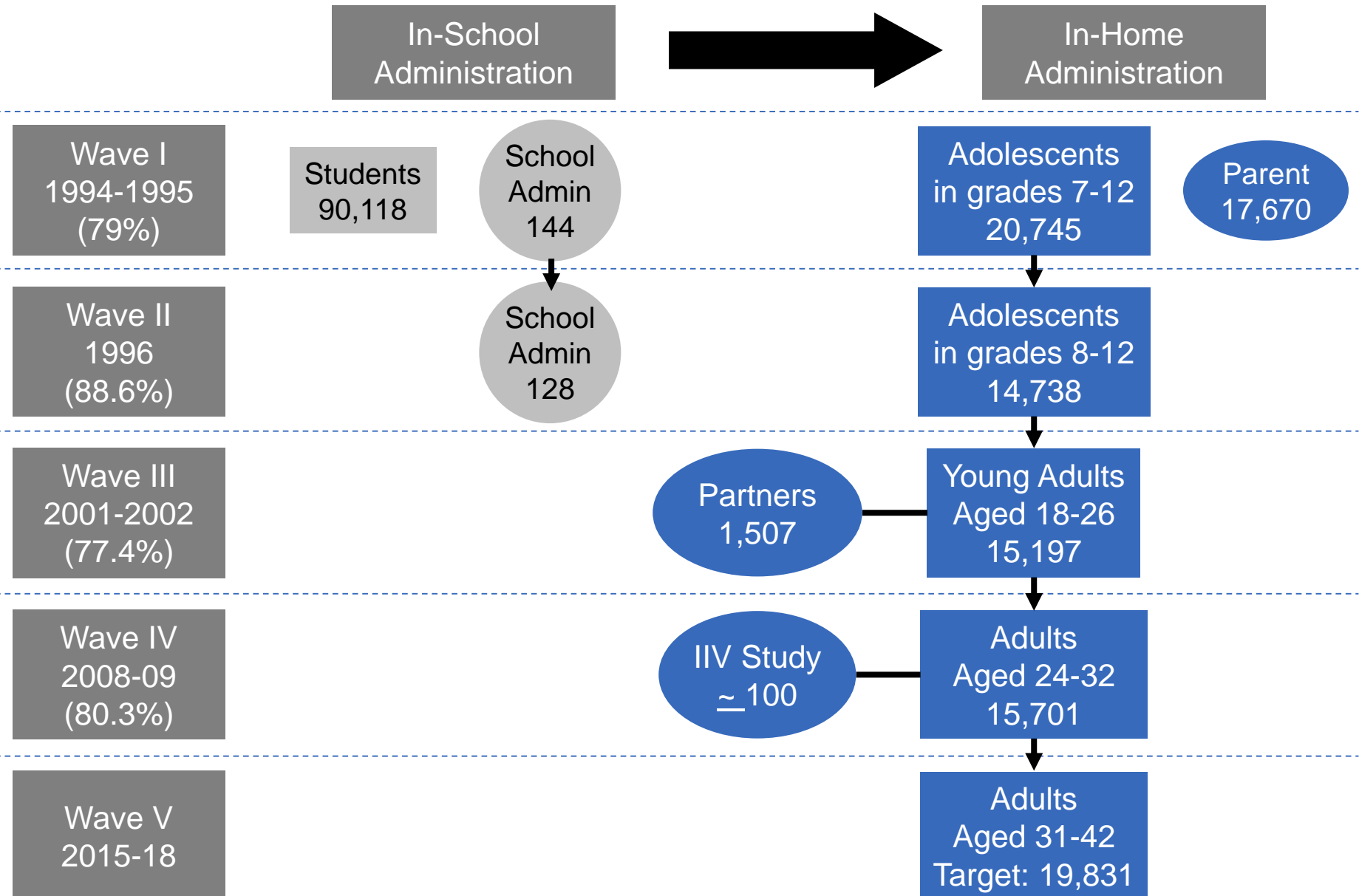
Data Users

- User requirements to protect from deductive disclosure:
 - Pledge of confidentiality
 - Monitoring of data use
 - Store data securely
 - Deletion of temporary data analysis files every six months.
 - Security of printed information
 - Password protected screen saver, set to activate after 3 minutes of idle time.
 - Access data only from approved locations.


Promote Use of Add Health data

- Presentations about Add Health design and unique opportunities for research
 - Conferences: Sociological, population, family, public health, behavioral genetics, medical, adolescent health;
 - Presentations, posters, workshops, didactic sessions
 - Distributed our public use data on CDs.
 - Seminar series in multiple departments, institutes and institutions;
 - Congressional briefings;
 - I gave 16 such presentations between 1995-96.
- Presentations at NIH whenever they called
- Add Health Users Conference on NIH campus





Biological Data Across Waves

Adolescence	Transition to Adulthood	Young Adulthood	Adulthood
Wave I-II (Ages 12-20)	Wave III (Ages 18-26)	Wave IV (Ages 24-32)	Wave V (Ages 31-42)
Embedded genetic sample of ~3,000 pairs			
Physical development 			
Height, weight	Height, weight	Height, weight, waist	Height, weight, waist
	STI tests (urine)	Metabolic	Metabolic
	HIV test (saliva)	Immune function	Immune function
	Genetic (buccal cell DNA)	Inflammation	Inflammation
		Cardiovascular	Cardiovascular
		Genetic (buccal cell DNA)	Genetic (whole blood)
		Medications	Medications
			Renal

Social and Biological Longitudinal Data in Add Health

Adolescence



Adulthood

Wave I-II
(12-20)

Wave III
(18-26)

Wave IV
(24-32)

Wave V
(31-42)

Social environmental data:

school

college

college

work

family

family

family

family

romantic rel

romantic rel

romantic rel

romantic rel

neighborhood

neighborhood

neighborhood

neighborhood

community

community

community

community

peer

peer

Biological data:

Biological resemblance to siblings in household on 3,000 pairs

height

height

ht, wt, waist, BMI

ht, wt, waist, BMI

weight

weight, BMI

BP, pulse

BP, pulse

BMI

STI test results

immune

immune

HIV test results

inflammation

inflammation

DNA

diabetes

diabetes

DNA

kidney disease

GWAS

Original Data Sharing Plan able to accommodate collection of new sensitive data

- STD and HIV results (on both members of couple sample);
- Links to military and other administrative data;
- Biomarkers;
- Genetic data.

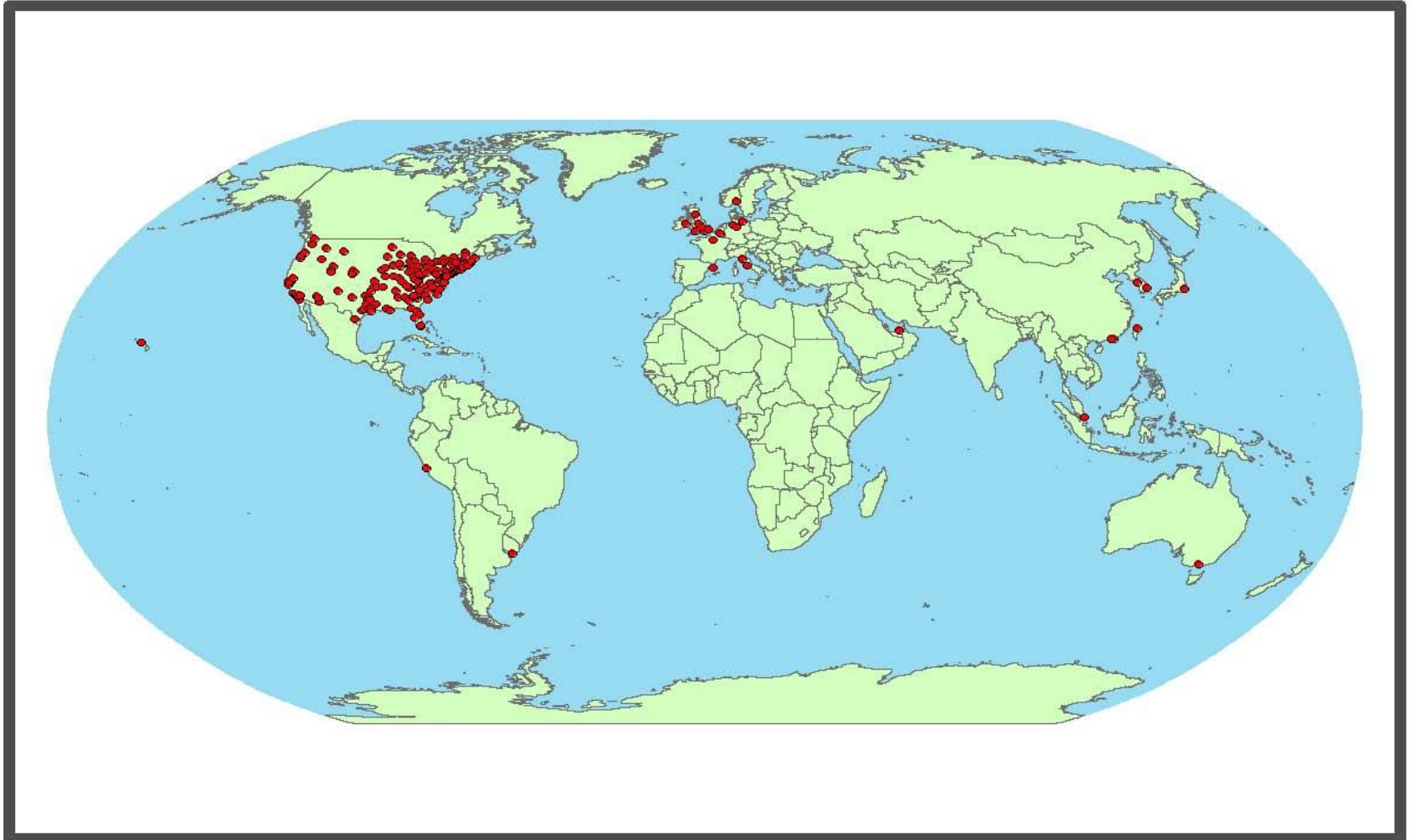
Other Data Security

- Keep up with technological advances in data access, hacking, and intruding.
- SANS Institute
 - Organization that provides computer security training and certification;
 - Continuing training and updating security systems on 20 top threats to data security;
 - Add Health staff data security specialist.
- Procedures for PI, study administrators, survey field staff and interviewers, data manager, and data users.

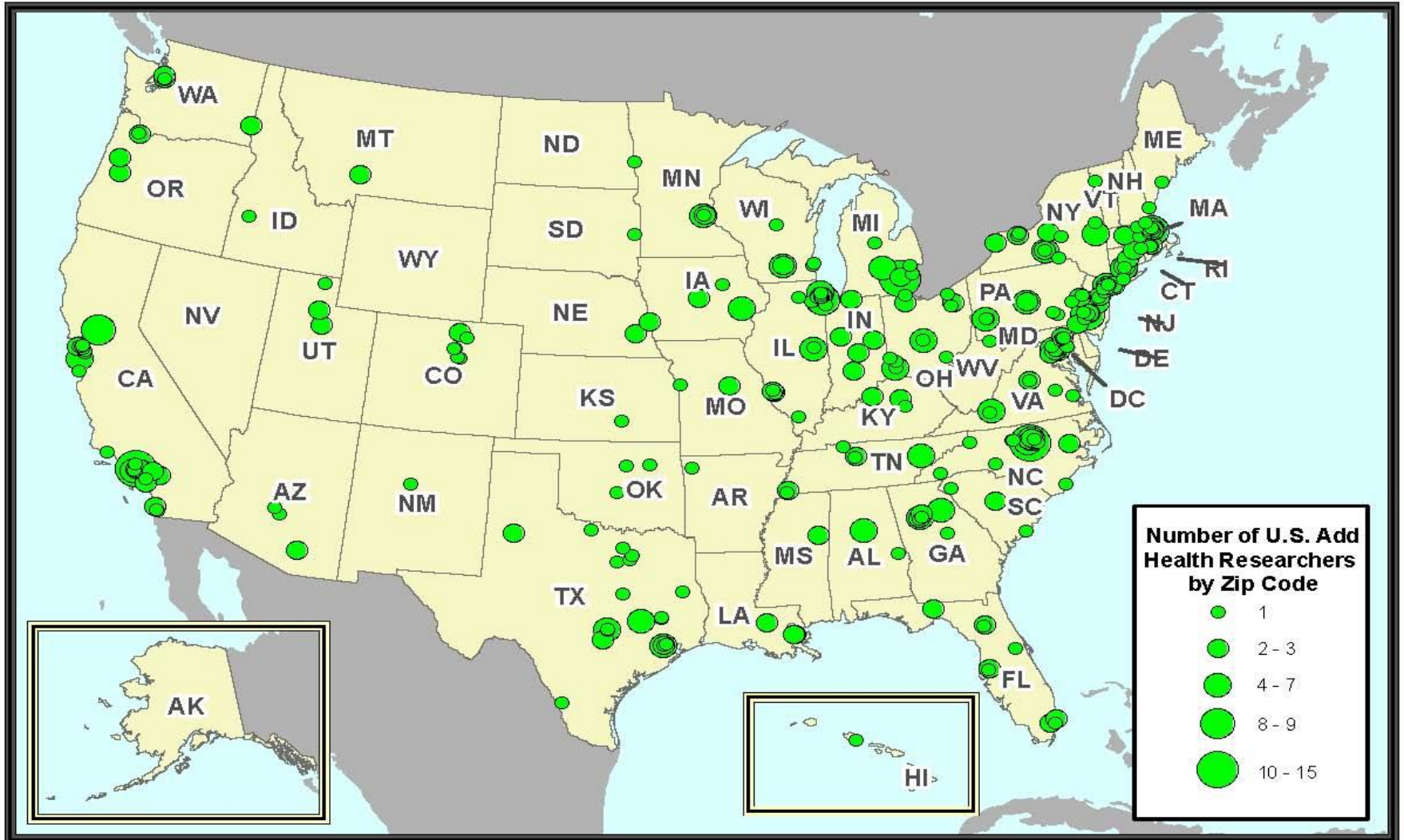
Dissemination Successes

- Data made available to more than 10,000 researchers for analyses;
- 650+ grants awarded to analyze data
- 2,000+ peer-reviewed publications that have appeared in > 350 different disciplinary journals
- 20 books, 100 reports, and 75 book chapters based on Add Health data
- 450+ doctoral dissertations and masters theses

Over 10,000 Add Health Users Around the World



Add Health Users in the US



Ongoing Dissemination Challenges

- Time-consuming
- Multiple staff demands and skills
- Resort to custom-approach with users to set up security plans at IT-poor institutions
- Educate
 - IRBs
 - Users
 - Investigators
 - Never-ending with rapidly changing technology and methods of data intrusion.

Ancillary Studies

- Researchers may propose ancillary studies to add data to Add Health;
 - No access to respondents;
 - Geographic, political, environmental (via geocodes);
 - Biological and genetic data (via biological specimen archive);
 - Review ancillary study proposals for scientific merit, value and burden to Add Health, scarcity and amount requested of specimen archive.
 - Cover all costs, including Add Health costs to check and merge data via Data Security Manager outside the U.S. and disclosure analysis;

Ancillary Studies

- Add Health policies:
 - Ancillary investigators must check and clean data unlinked to Add Health longitudinal data;
 - Contextual data has no proprietary period
 - The day we provide the data to the Ancillary Study PI is the day we distribute to community of Add Health users.
 - Biological and genetic data have a 1-year proprietary period.

Research Findings Unique to Add Health Design

- Network data
 - Bearman, Peter S.; & Bruckner, Hannah. (1999). *Power in Numbers: Peer Effects on Adolescent Girls' Sexual Debut and Pregnancy*. Washington, DC: National Campaign to Prevent Teen Pregnancy.
- Biomarker data
 - Brückner, Hannah; & Bearman, Peter S. (2005). After the promise: The STD consequences of adolescent virginity pledges. *Journal of Adolescent Health, 36(4)*, 271-278.
- Longitudinal health data
 - Harris, K. M.; Gordon-Larsen, P.; Chantala, K.; & Udry, J. R. (2006). Longitudinal trends in race/ethnic disparities in leading health indicators from adolescence to young adulthood. *Archives of Pediatrics and Adolescent Medicine, 160(1)*, 74-81.
- Contextual data to study gene – environment interplay
 - Daw, J.; Shanahan, M.; Harris, K. M.; Smolen, A.; Haberstick, B.; & Boardman, J. D. (2013). Genetic sensitivity to peer behaviors: 5HTTLPR, smoking, and alcohol consumption. *Journal of Health and Social Behavior, 54(1)*, 92-108.
 - Boardman, J. D.; Domingue, B. W.; & Fletcher, J. M. (2012). How social and genetic factors predict friendship networks. *Proceedings of the National Academy of Sciences of the United States of America, 109(43)*, 17377-17381.



Major Findings


- **Importance of social connections for health and well-being across the life course;**
- Transition from adolescence into early adulthood is a vulnerable period for health that sets trajectories into adulthood;
- Mapping the obesity epidemic;
- Gene—environment interplay in health and behavior;
- Young adult health at risk;

Major Findings

- Importance of social connections for health and well-being across the life course;
- **Transition from adolescence into early adulthood is a vulnerable period for health that sets trajectories into adulthood;**
- Mapping the obesity epidemic;
- Gene—environment interplay in health and behavior;
- Young adult health at risk;

Health During the Transition to Adulthood

- 
- Depression
 - Suicidal thoughts
 - Self-reported poor health

- 
- Obesity
 - No physical activity
 - No breakfast
 - Eating fast food
 - No health insurance
 - No annual check-up
 - No dental check-up
 - Foregone care
 - Asthma
 - STDs
 - Smoking
 - Marijuana use
 - Hard drug use
 - Binge drinking



UNC
CAROLINA
POPULATION
CENTER

Harris et al. 2006

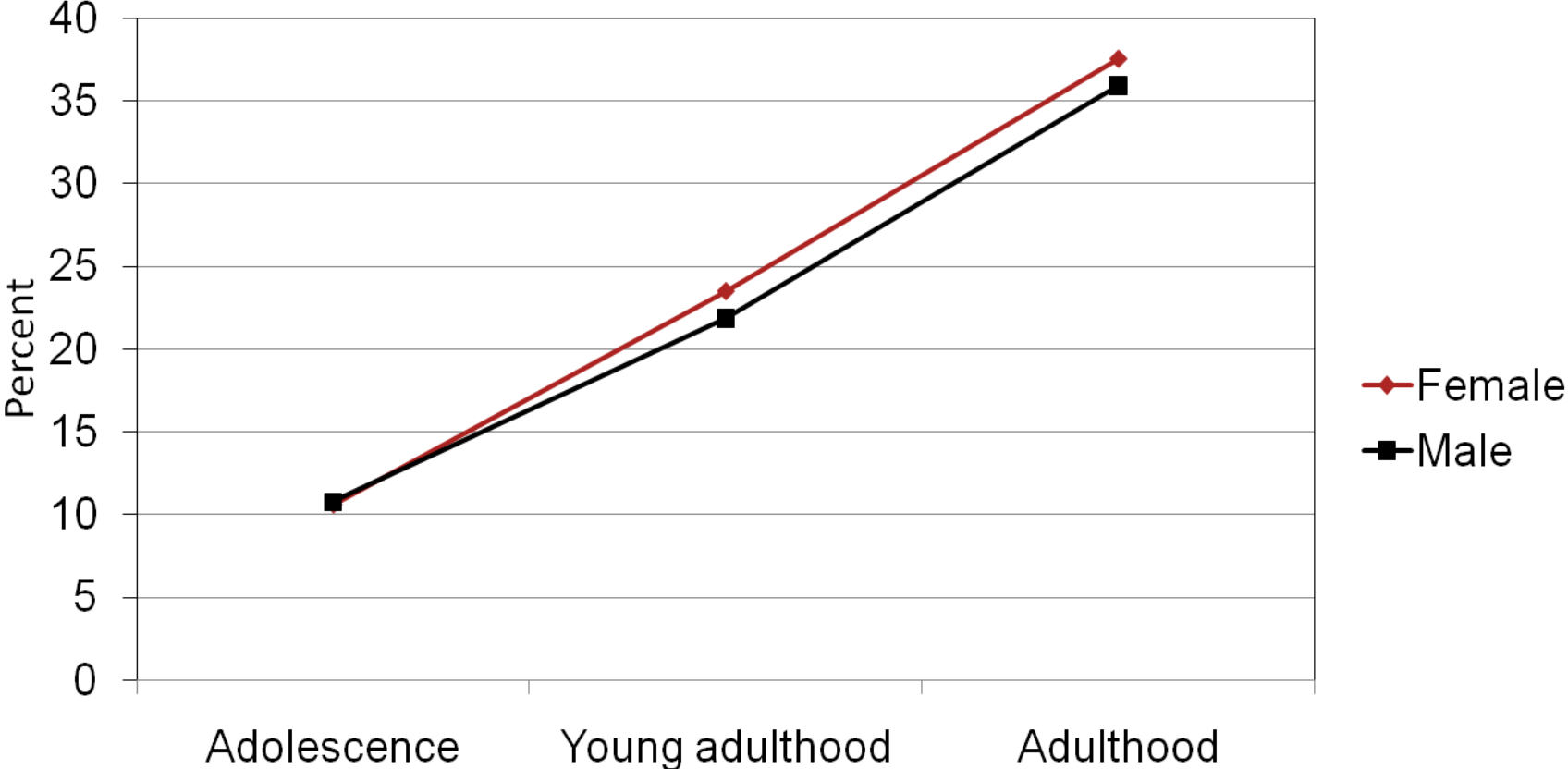


Add Health
The National Longitudinal Study of Adolescent to Adult Health

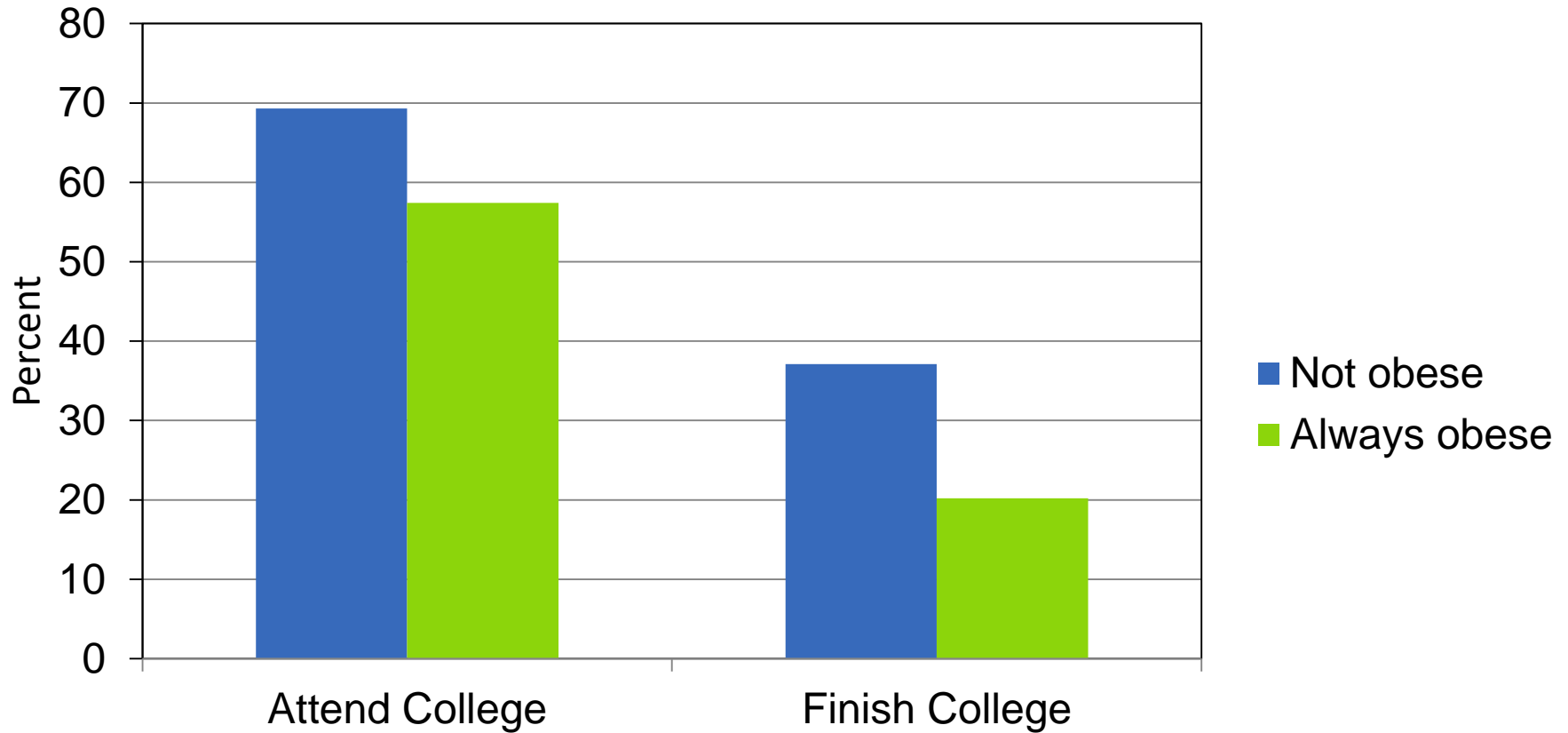
Major Findings

- Importance of social connections for health and well-being across the life course;
- Transition from adolescence into early adulthood is a vulnerable period for health that sets trajectories into adulthood;
- **Mapping the obesity epidemic;**
- Gene—environment interplay in health and behavior;
- Young adult health at risk;

Obesity from adolescence into adulthood



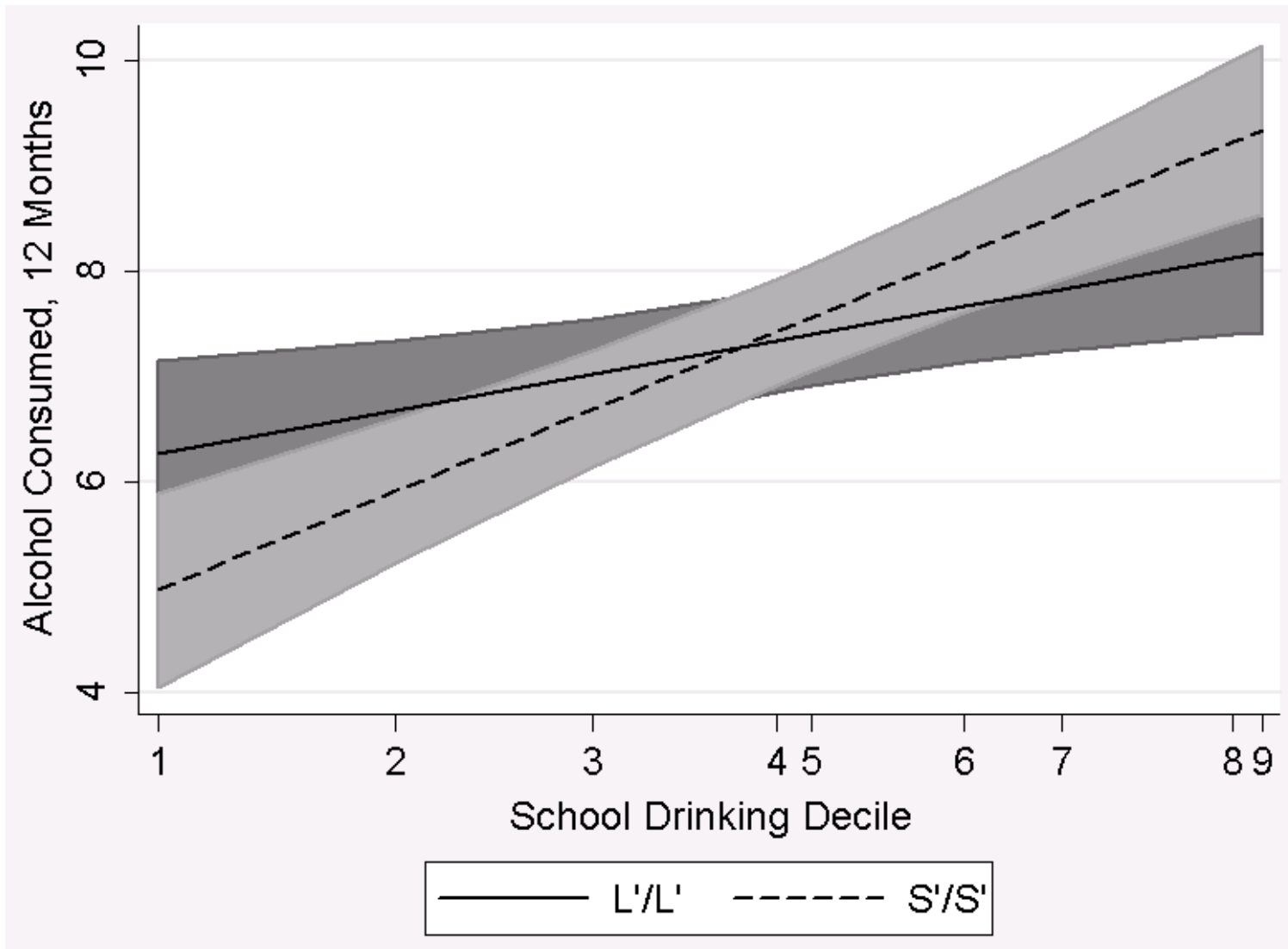
Obesity trajectories from adolescence to young adulthood associated with Educational Outcomes



Major Findings

- Importance of social connections for health and well-being across the life course;
- Transition from adolescence into early adulthood is a vulnerable period for health that sets trajectories into adulthood;
- Mapping the obesity epidemic;
- **Gene—environment interplay in health and behavior;**
- Young adult health at risk;

Alcohol Consumption by *5HTTLPR*

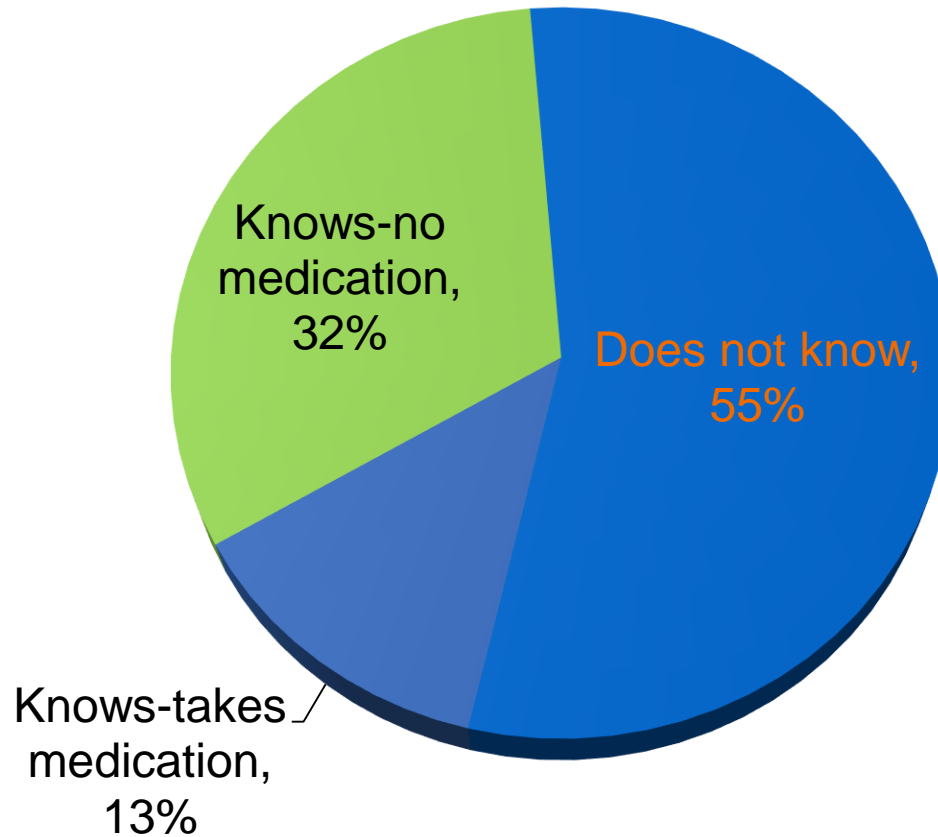


Major Findings

- Importance of social connections for health and well-being across the life course;
- Transition from adolescence into early adulthood is a vulnerable period for health that sets trajectories into adulthood;
- Mapping the obesity epidemic;
- Gene—environment interplay in health and behavior;
- **Young adult health at risk;**

Add Health uncovered a hidden epidemic of hypertension among young adults: 26% of adults aged 24-32 have hypertension (SBP \geq 140 or DBP \geq 90); and most don't know it.

Among young adults with hypertension



Add Health Info

- Website: www.cpc.unc.edu/addhealth
- Description of data access and dissemination:
<http://www.cpc.unc.edu/projects/addhealth/data>
- Restricted data use contract:
<http://www.cpc.unc.edu/projects/addhealth/data/contract>

Add Health Co-Funders

- National Institute of Child Health and Human Development*
- National Cancer Institute*
- National Center for Health Statistics, Centers for Disease Control and Prevention, DHHS
- National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, DHHS*
- National Center for Minority Health and Health Disparities*
- National Institute of Allergy and Infectious Diseases*
- National Institute of Deafness and Other Communication Disorders*
- National Institute of General Medical Sciences
- National Institute of Mental Health
- National Institute of Nursing Research*
- National Institute on Aging*
- National Institute on Alcohol Abuse and Alcoholism*
- National Institute on Drug Abuse*
- National Science Foundation*
- Office of AIDS Research, NIH*
- Office of the Assistant Secretary for Planning and Evaluation, DHHS*
- Office of Behavioral and Social Sciences Research, NIH*
- Office of the Director, NIH
- Office of Minority Health, Centers for Disease Control and Prevention, DHHS
- Office of Minority Health, Office of Public Health and Science, DHHS
- Office of Population Affairs, DHHS*
- Office of Research on Women's Health, NIH*

*Wave 4 co-funders



UNC
CAROLINA
POPULATION
CENTER

