

Accessing Restricted Data at the University of Maryland Federal Statistical Research Data Center

Michel Boudreaux - HLSA

mhb@umd.edu

Andrew Fenelon - HLSA

afenelon@umd.edu



SCHOOL OF
PUBLIC HEALTH

DEPARTMENT OF HEALTH SERVICES
ADMINISTRATION



MPRC
Maryland Population Research Center

What is the RDC?

- A consolidated statistical computing platform linking researchers to protected datasets through federal agencies
- Run by the US Census Bureau
- Participating organizations include NCHS, AHRQ, BLS, and Census

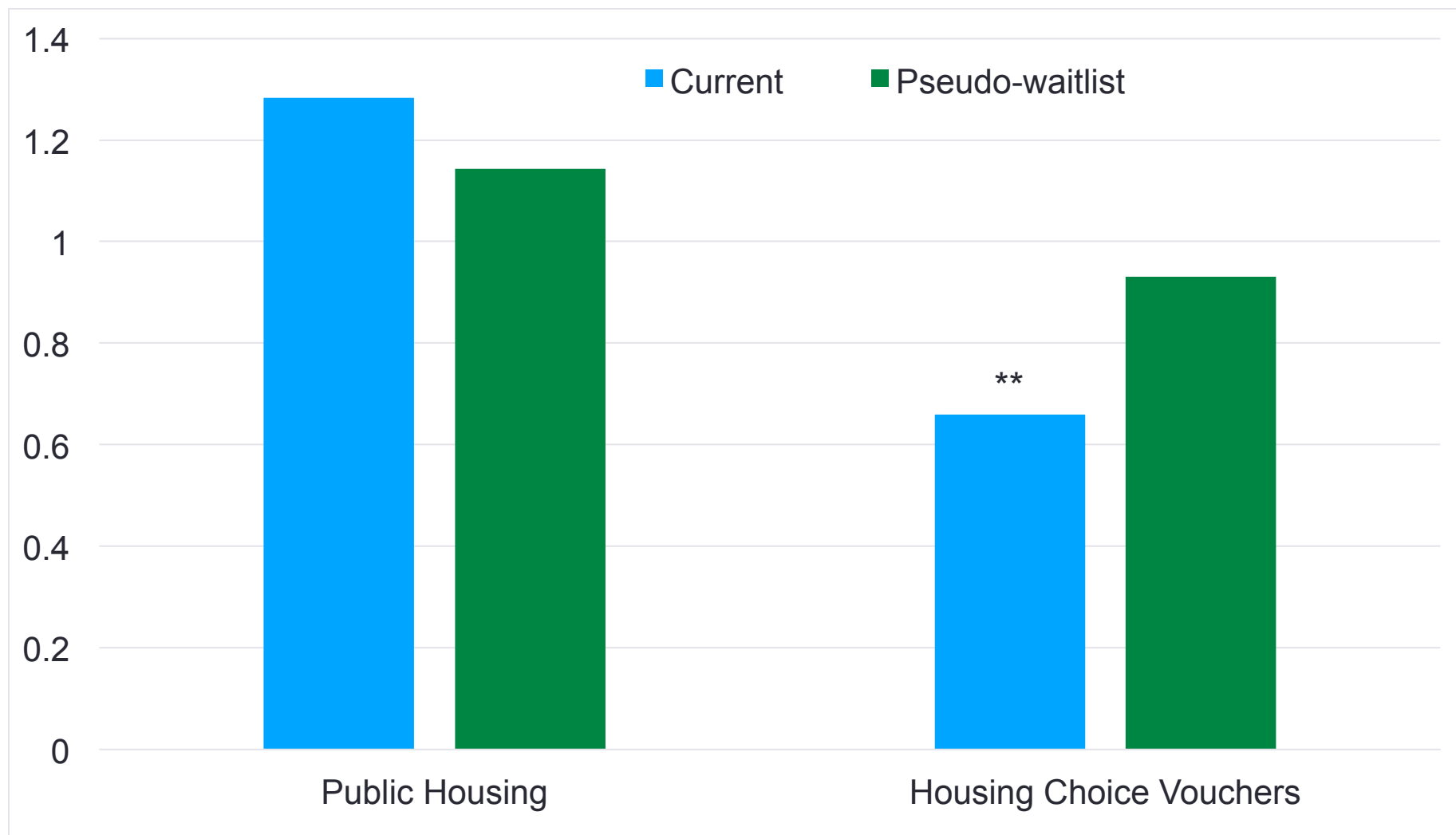
What kind of variables can you get?

- Restricted geography
 - State, ZIP, tract, county, etc etc
- Full variable distributions
 - No top-coding
- Full count files and unedited variables
 - No PUMS for us!
- Linking Keys
 - Personal Identification Keys (PIKs) for merging across individual data
- Administrative data

University of Maryland RDC

- <https://marylandrdc.umd.edu/>
- Our RDC is a fantastic resource that allows affiliates of founding schools to use the RDC with no seat charge
- (Researchers who use the RDC at NCHS in Hyattsville are charged \$300 per day of use)
- Director: Professor Liu Yang – lyang@rhsmith.umd.edu
- Administrator: Currently Vacant (Veronika Penciakova starting in October!)

Neighborhood Disadvantage Index



† p<0.10 * p<0.05 ** p<0.01

Steps in the RDC Process

1. Proposal Application
2. Certification (Special Sworn Status)
3. Data Merging and Upload (NCHS Only)
4. Analysis
5. Disclosure

Example Health Datasets in the RDC

- [National Health and Nutrition Examination Survey \(NHANES\)](#)
- National Health Care Surveys
 - [National Ambulatory Medical Care Survey \(NAMCS\) and National Hospital Ambulatory Medical Care Survey \(NHAMCS\)](#)
 - [National Hospital Discharge Survey \(NHDS\)](#)
 - [National Nursing Home Survey \(NNHS\) and National Nursing Assistant Survey \(NNAS\)](#)
 - [National Home and Hospice Care Survey \(NHHCS\) and National Home Health Aide Survey \(NHHAS\)](#)
 - [National Survey of Residential Care Facilities \(NSRCF\)](#)
 - [National Study of Long-Term Care Providers \(NSLTCP\)](#)
 - [National Hospital Care Survey \(NHCS\)](#)
- [National Health Interview Survey \(NHIS\)](#)
- [National Survey of Family Growth \(NSFG\)](#)
- [State and Local Area Integrated Telephone Survey \(SLAITS\)](#)
 - National Survey of Children's Health (NSCH)
 - National Survey of Children with Special Health Care Needs (CSHCN)
- NCHS Data Linkage Activities
 - [Linked Mortality Data Products](#)
 - [Linked Medicare Enrollment and Claims Files Data](#)
 - [Linked Medicaid Enrollment and Claims Data](#)
 - [Linked Social Security Benefit History Data](#)
- [National Vital Statistics System \(NVSS\) Data Release and Access Policy](#)
- [National Maternal and Infant Health Survey](#)

Examples of Census Data

- Demographic
 - Decennial, ACS, CPS, AHS, etc
- Economic
 - Economic Census, LEHD, Annual Survey of Manufacturers, etc
- Administrative Data
 - IRS, Public Programs, SSA
 - Hard to get public information on all that they have, but they have a lot. Assume they have what you want.
 - IRS data is usually very hard to get, best to have inside collaborator
- Proprietary Data
 - In theory, Census will PIK any data set you can send them which allows it to be merged with any of these files. No extra charge for this service.

RDC Application

- NCHS
 - <https://www.cdc.gov/rdc/b3prosal/PP300.htm>
- Census
 - <https://www.census.gov/ces/rdcresearch/howtoapply.html>
- AHRQ
 - https://meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp
- BLS
 - <https://www.bls.gov/rda/>

NCHS Application

Proposal Format (Revised 2/04/2014)

Project Title, Researcher

RDC Research Proposal

General Information	
Date:	
Title of Project:	
NCHS Data System and Years:	
Non-NCHS Data Files:	
Mode of Access:	<input type="checkbox"/> NCHS RDC, Hyattsville, MD (Washington, DC-metro) <input type="checkbox"/> NCHS RDC, Washington, DC (Government Only) <input type="checkbox"/> NCHS RDC, Atlanta, GA <input type="checkbox"/> Remote Access (ANDRE) <input type="checkbox"/> Census RDC, specify: _____
Statistical Software: (Check all that apply)	<input type="checkbox"/> SAS/Sudaan <input type="checkbox"/> Stata <input type="checkbox"/> Other, specify: _____ * Remote access users can only use SAS/Sudaan
Proposed Start Date:	
Funding Source:	
Billing Address: (include contact person)	

NCHS Application

- A. Abstract
- B. Research Question
- C. Background
- D. Public Health Benefit
- E. Data Requirements (data you need, years of surveys, restricted variables, additional data merged)
- F. Methodology (what you'll be doing)
- G. Output (Table Shells and presentation of results)
- H. Data Dictionary (list of all public and restricted variables you will use)

Census Application

- Generally starts with a 'pre-proposal' that local RDC admin will give you feed back on
- Provide benefit to Census
 - Hardest part, but RDC Admin will help you figure this out.
- Demonstrate Scientific Merit
- Requires non-public data
- Be feasible
- Pose no disclosure risk

Special Sworn Status

- To use the RDC, you become an “employee” of the census bureau, and you have to go through the background check process to obtain that status
- The whole process can take several months, so start early (but it lasts for life!!!!)
- In addition to filling out a ton of forms, you also have to
 - Submit fingerprint cards
 - A passport photo
 - Be interviewed by the OPM Federal Investigator

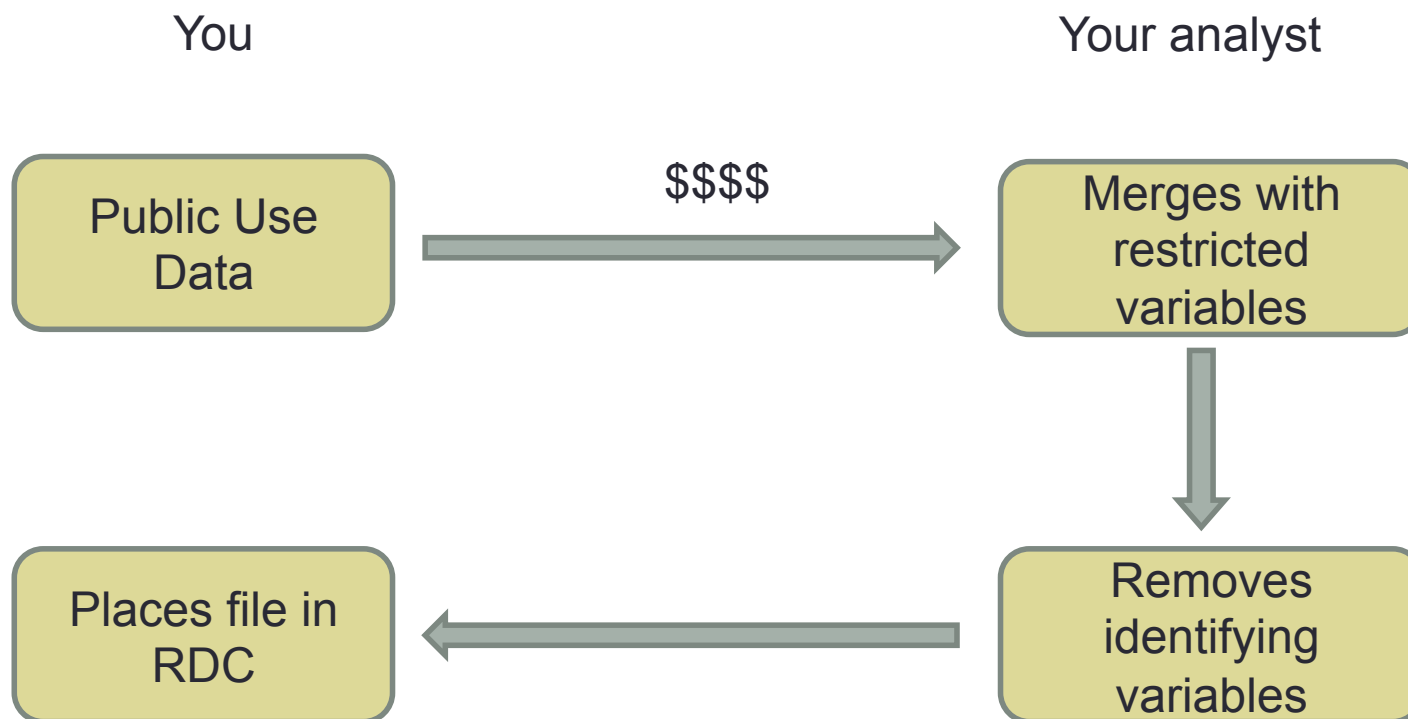
List of SSS documents

Checklist:

- PIV Training Certificate
- I-9 Form + supporting documentation
- BC-1759
- CES Researcher information sheet
- Initial Information Sheet
- CD-591
- Information Technology Rules of Behavior
- New Fair Credit Reporting
- OF-306
- For Non-Citizens only:*
 - Copies of Visa/green card, picture page of passport
 - Foreign National Info form
 - Foreign National Residence History form
 - DAO Attachment 3
- Fingerprint cards (x2)
- Passport quality photo (*email JPEG format to Sandy Dietrich*)
- Data Stewardship Training Certificate
- Title 26 Training Certificate

Data Merging for NCHS Data

- You provide your public use data, then the analyst merges it with the restricted variables



Data Merging for Census

- You send any external data (not Census data) and then you are on your own, which is good and bad.
- You will be given access to internal files that you are granted permission for. It is then up to you to merge
- Greater control, but requires you have a deeper understanding of the data

Analysis

- Data analysis can be done with a variety of statistical packages including SAS, Stata, and R via Unix
 - Stata at least is interactive. Not true for SAS.
- You can have your analyst upload program code to the RDC
- Unfortunately you have to leave the RDC to consult any online statistical programming help

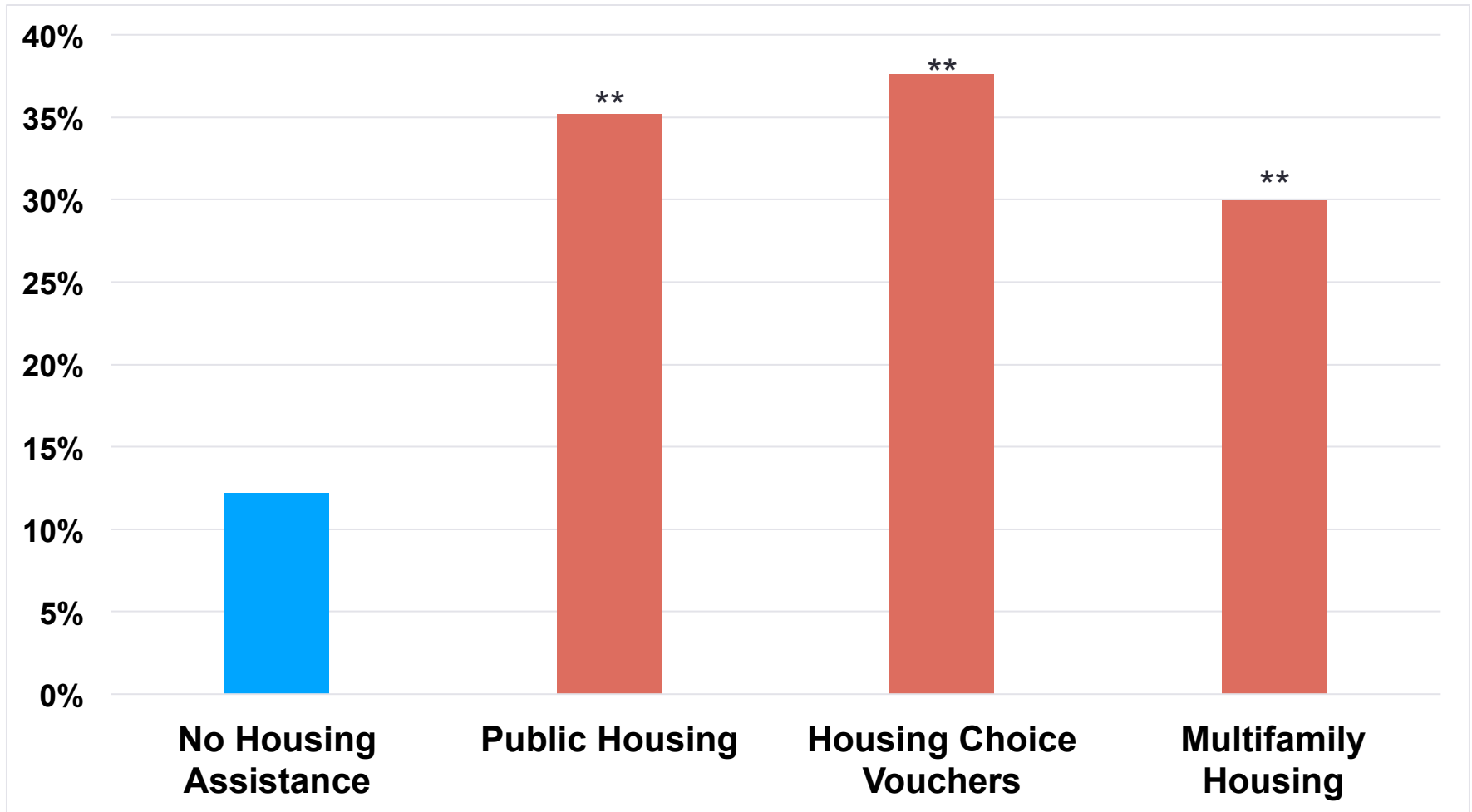
Disclosure

- Unfortunately, you can't just take your results with you when you leave. Disclosing results involves several steps and differs by agency and analyst.
 1. Run analysis and place .log file in 'disclosure' folder
 2. Send analyst a disclosure request along with path of folder
 3. Wait...
 4. Receive results in email as long as analyst deems them releasable (this can be annoying and frustrating)

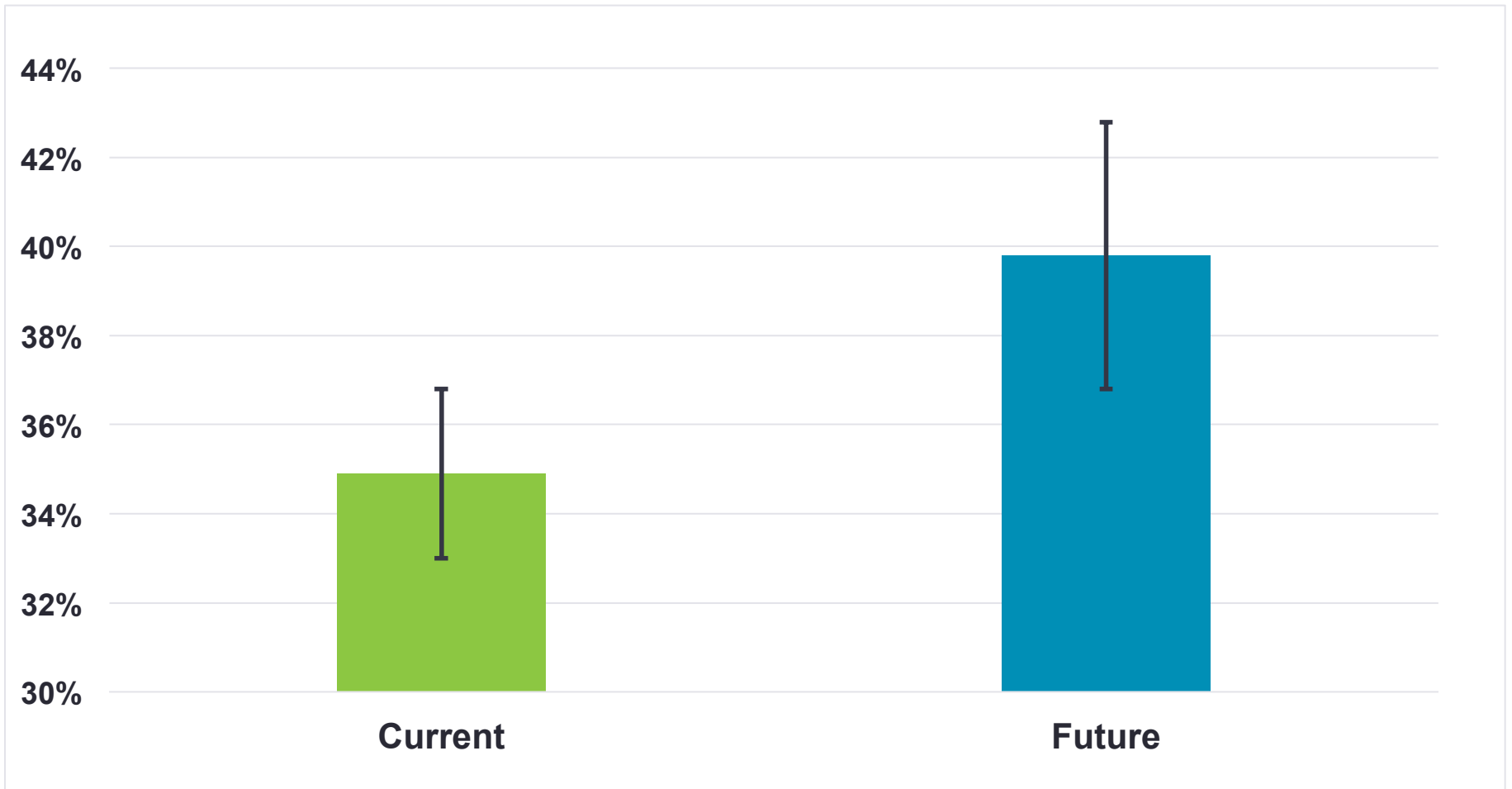
Examples – Our Current Projects

- HUD housing assistance and child health and well-being (Andy, Michel, and Natalie)
 - NHIS/NHANES data linked with HUD records and geographic info
- NHIS and NHANES merged with census tract and zip code data (Quynh)
- American Housing Survey and census tract data (Andy)
- School-based health centers, school screening mandates, and Medicaid eligibility: Effects to health and health care access in the NHIS (Michel)

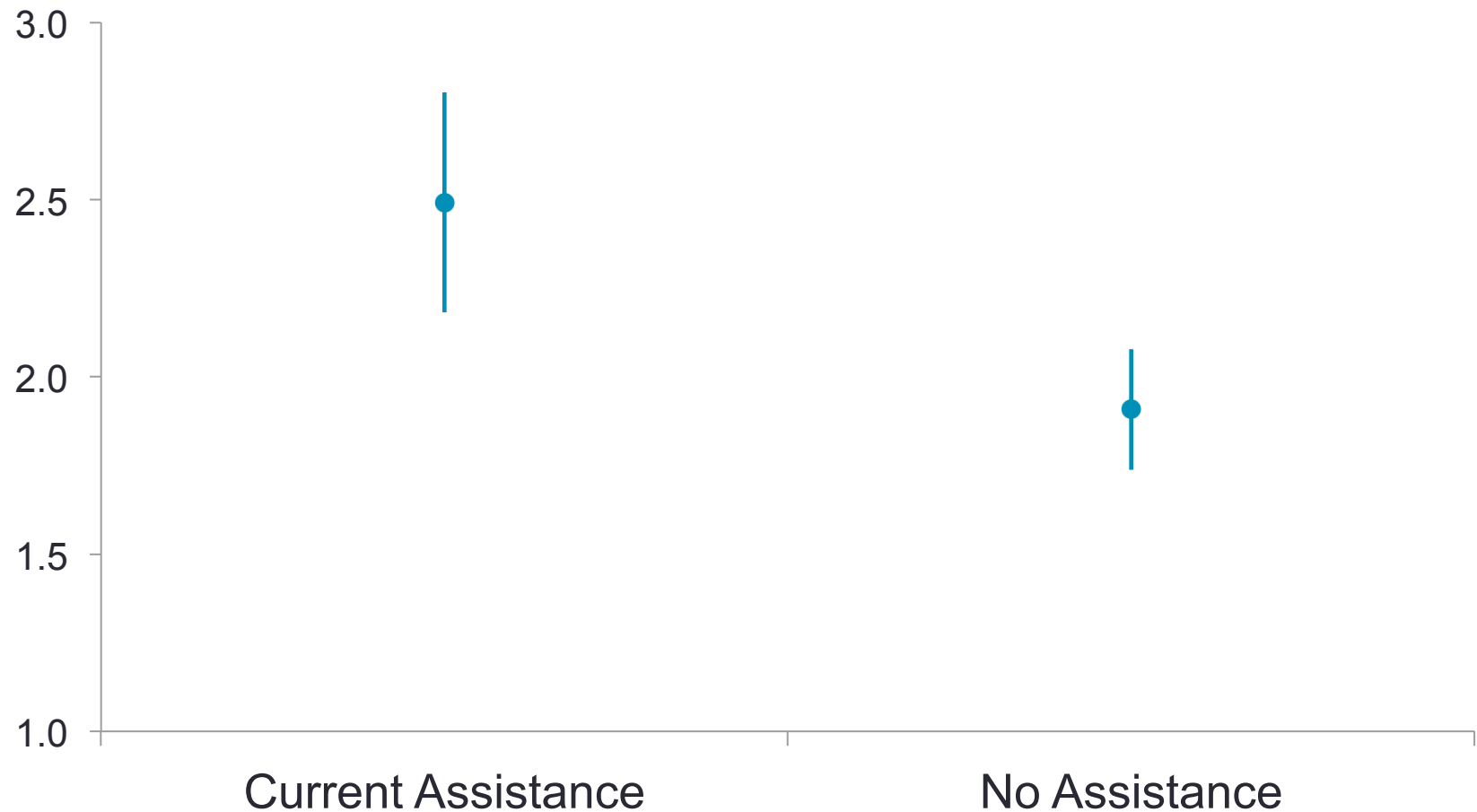
Adult Fair/poor health by Housing Assistance



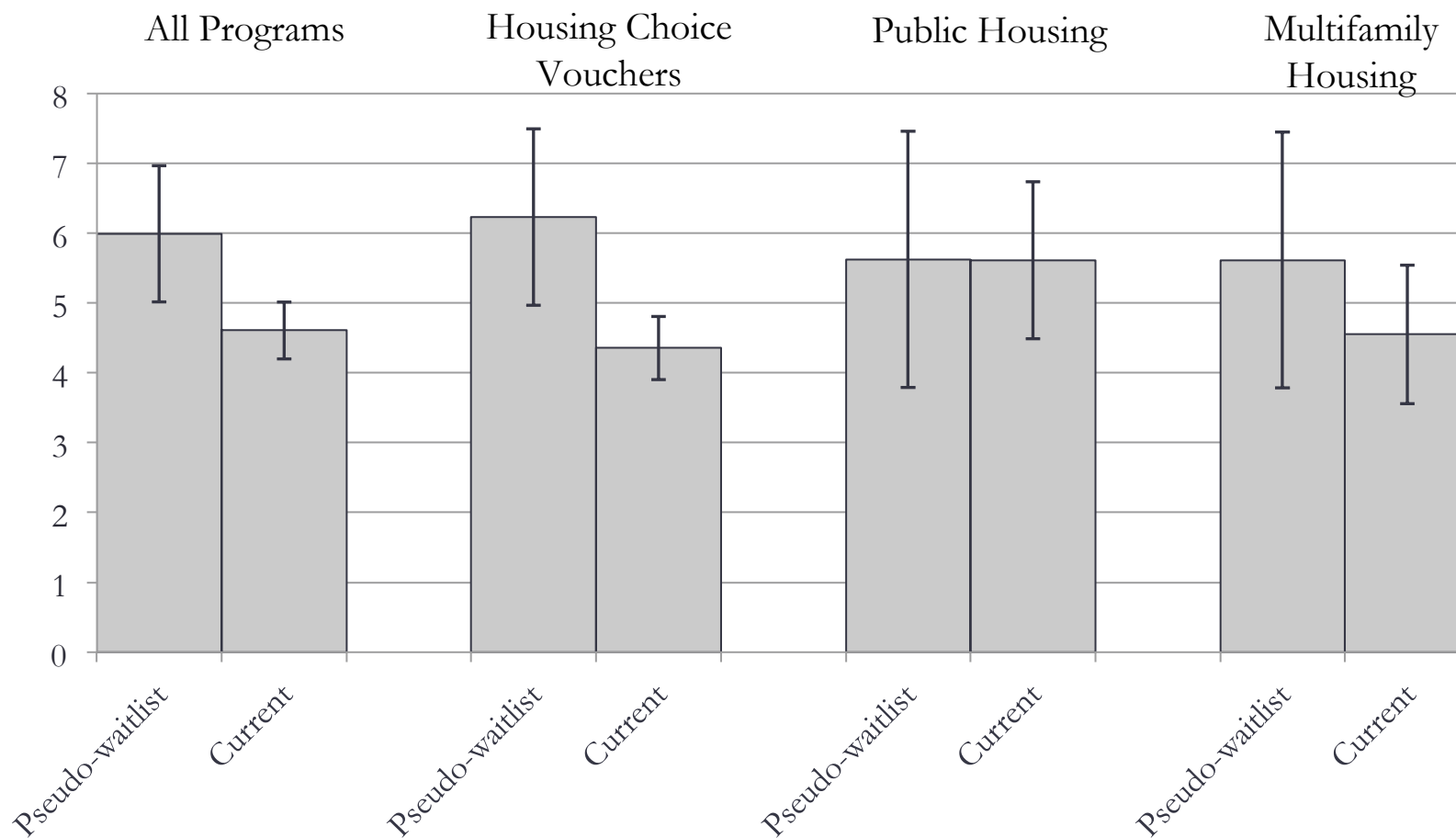
Adult Public Housing Current vs. Future



Child Mental Health by Housing Assistance



Health days of school missed



Campus Contacts

- Executive Director: Liu Yang lyang@rhsmith.umd.edu
- SPH Rep: Michel Boudreaux mhb@umd.edu
- BSOS Rep: Andres Villareal avilla4@umd.edu

- Local Health Data users:
 - Andy Fenelon afenelon@umd.edu
 - Quynh Nguyen qtnguyen@umd.edu
 - Natalie Slopen nslopen@umd.edu
- Interns:
 - Seth Murray murray@econ.umd.edu
 - Claire Hou hou@econ.umd.edu