The Health and Retirement Study: Contextual Data Augmentation

November, 2021

Christopher Dick, Demographic Analytics Advisors

Introduction

The Health and Retirement Study is an amazing resource for those studying aging in the United States, and a fantastic model for other countries who have created similar longitudinal studies. The raw amount of information, from data on income, wealth, and use of health services to employment, retirement, and family connections on to the collection of clinical biomarkers can be both empowering and overwhelming to a researcher. Luckily through the process of engagement with the research community and constant improvement, these reams of data are not only consistently growing in a thoughtful and focused direction, they are also explained and summarized to increase the ease of use for all.

One of the very useful areas of the HRS is the Contextual Data File (CDF), which is the focus of this review. The CDF provides access to easy to use helpful community-level data in a secure environment that has allowed researchers to answer questions that would have otherwise been difficult or impossible to tackle. The current CDF includes data in 6 categories:

1. Socio-economic Status and Demographic Structure
2. Psychosocial Stressors
3. Health Care
4. Physical Hazards
5. Amenities
6. Land Use and the Built Environment[1]

Each of these areas have allowed researchers to answer interesting questions such as what is the impact of air pollution on cognition in older adults[2], the impact of neighborhood

---

[1] University of Michigan Institute for Social Research (2017) *HRS Data Book: The Health and Retirement Study: Aging in the 21st Century, Challenges and Opportunities for Americans.* Available online: https://hrs.isr.umich.edu/about/data-book

[2] Ailshire, Jennifer and Walsemann, Katrina M. (2021). "Education Differences in the Adverse Impact of PM 2.5 on Incident Cognitive Impairment Among U.S. Older Adults". *Journal of Alzheimer's Disease*, 79(2):615-625. DOI: 10.3233/JAD-200765

characteristics on obesity in older adults[3], or even what do we gain from introducing contextual data to a survey analysis[4]?

My review focuses on the potential to expand contextual data in a few of these areas. From new data sets developed and released by the U.S. Census Bureau, to improved measurements of climate and environmental risk, there are numerous new data sources that would be a boon to the research community if they were joined together with the HRS. The following section begins by breaking down the opportunity provided by community or place based data before moving on to specific recommendations for new data that could be included in the HRS contextual data file.

Community-level Data: The Opportunity

The last couple of years have laid bare the importance of community resiliency and the impacts of both natural and human-induced disasters on our society. While these factors are important for any social or policy research, their bearing on the areas covered by the HRS are critical. Because of these issues, I have focused my review on data sources that could help those using the HRS to better interrogate questions such as: What are the impacts of climate induced disasters on economic well-being in retirement? Or, how does community resiliency impact the health outcomes of an aging population?

Each of the data sources below fits either the profile of a general use or a more specialized use data source that helps us to better understand these areas of inquiry. For each resource I discuss the following:

1. A brief summary of how the dataset was built and by whom.
2. What can be gained by including the data in the HRS Contextual data file.
3. How it can be obtained, and are there multiple versions that can be obtained.
4. Level or levels of analysis and update cadence.
5. Is it a model or observed data, and if it is a model, the level of transparency we have into the inner workings.

Community Resilience Estimates (US Census Bureau)

Why and How the Data Set Was Constructed

The new Community Resilience Estimates (CRE) are a great example of a general use data resource that many academic and policy researchers could benefit from having in easy to use format with the HRS. The COVID-19 pandemic has changed the way that many of us, whether

---

[3] Grafova, Irina B., Vicki A. Freedman, Rizie Kumar, and Jeannette Rogowski. (2008). "Neighborhoods and Obesity in Later Life", *American Journal of Public Health*, 98:2065-2071, DOI: 10.2105/AJPH.2007.127712

[4] Wilkinson, Lindsay R., Kenneth F. Ferraro, and Blakelee R. Kemp. (2017). "Contextualization of Survey Data: What Do We Gain and Does It Matter?", *Research in Human Development*, 14:3, 234-252, DOI: 10.1080/15427609.2017.1340049

in academia, the private sector, or government, think about the timeliness and level of geographic specificity in our data measurement. This has led to many innovations at places like the US Census Bureau, including the development of an ongoing pulse survey for both the population as well as business. One of these innovations could provide a new depth of information for those using the Health and Retirement Survey, the CRE. These estimates were in an experimental phase in 2018[5], and were turned into an ongoing official product in 2019[6].

The goal of the CRE is to measure the ability of a community to respond to disaster, including weather based disasters, pandemics, or other disasters. The model uses small area estimation techniques on data from the yearly population estimates as well as the restricted use American Community Survey micro-data. Specifically, the CRE aggregates the existence of resiliency risk factors at the household level at the census tract, county, state, and national levels, measuring the number of households who have 0 risk factors, 1 to 2 risk factors, and 3 or more risk factors out of a possible 10. The risk factors measured by the Census Bureau here include (note that for #4 a house will not be double counted if both communication barriers exist):

1. Income-to-Poverty Ratio (IPR) < 130 percent (Household)
2. Single or zero caregiver household - only one or no individuals living in the household who are 18-64 (Household)
3. Unit-level crowding with >= 0.75 persons per room (Household)
4. Communication Barrier defined as:
    1. No one in the household has received a high school diploma
    2. No one in the household speaks English "very well"
5. Aged 65 years or older
6. No one in the household is employed full-time, year-round. The flag is not applied if all residents of the household are aged 65 years or older (Household)
7. Disability, at least one serious constraint to significant life activity
8. No health insurance coverage
9. No vehicle access (Household)
10. Households without broadband internet access (Household)[7]

The CRE shares some common concepts and measurement with the Social Vulnerability Index[8], in that it is trying to determine the areas that are likely to be most vulnerable during a disaster. However, it differs in its use of micro-level data and small area estimation to create a

[5] United States Census Bureau. (2020). *Community Resilience Estimates: Quick Guide.* Last Accessed: 11/1/2021. https://www2.census.gov/data/experimental-data-products/community-resilience-estimates/2020/technical-document.pdf

[6] United States Census Bureau. (2021). *2019 Community Resilience Estimates: Quick Guide.* Last Accessed: 11/1/2021. https://www2.census.gov/programs-surveys/demo/technical-documentation/community-resilience/2019/cre_quickguide_2019.pdf

[7] United States Census Bureau. (2021). *2019 Community Resilience Estimates: Quick Guide.* Last Accessed: 11/1/2021. https://www2.census.gov/programs-surveys/demo/technical-documentation/community-resilience/2019/cre_quickguide_2019.pdf

[8] Agency for Toxic Substances and Disease Registry. (2021). *CDC/ATSDR Social Vulnerability Index.* Last Accessed: 11/1/2021. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

more timely and granular accounting of resilience factors. These data could be helpful in many contexts to users of the HRS.

What Will Be Gained By Including the CRE in the HRS Contextual Data?

Incorporating the Census Bureau's CRE data into the contextual data file for the HRS would help to ground research in several areas. The most obvious area for inquiry would be any research with a focus on disaster preparedness or response. Incorporating these community-level data into work that looks at individual level disaster preparedness[9][10] or reducing health and economic risks after a disaster[11] could help to explain some of the community level factors that may be at play. Further, some of these resiliency factors could be helpful to better understand short and long term outcomes of communities and individuals as we move into new stages, and hopefully at some point beyond, the COVID-19 pandemic. Finally, work to understand the efficacy of different interventions to boost resilience, for both individuals and communities, would be wise to use some of these community level data as well as individual-level data like those from the HRS as a starting point.

Logistical and Methodological Considerations

Is the resource public or private?: Public
How can the data be accessed?: Through public website or API
Unit(s) of analysis provided: Census tract, county, state, and national
Update cadence: Yearly
Coverage: All of the United States
Model or Observed Data?: Model
Transparency of sourcing or model: The model used to create the estimates are highly transparent through the Census Bureau's methodology documentation[12].

Flood Factor (First Street Foundation)

Why and How the Data Set Was Constructed

From understanding community level resiliency, we turn to a quantifiable measure of natural disasters. The impacts of climate change on our daily lives is being felt to a greater and greater extent each year by individuals across the globe. Flooding, together with fires, are perhaps the

---

[9] Cox, Katherine & BoRin Kim. (2018). Race and income disparities in disaster preparedness in old age. *Journal of Gerontological Social Work*, 61:7, 719-734, DOI: 10.1080/01634372.2018.1489929
[10] Al-rousan, Tala M., Linda M. Rubenstein, Robert B. Wallace. (2014). "Preparedness for Natural Disasters Among Older US Adults: A Nationwide Survey", *American Journal of Public Health* 104, no. 3: 506-511. DOI: 10.2105/AJPH.2013.301559
[11] Shin, Su Hyun and Hyunjung Ji. (2021). "Health risks of natural hazards and resilience resources: Evidence from a U.S. nationwide longitudinal study", *Social Science & Medicine*, 281: 114110, DOI: 10.1016/j.socscimed.2021.114110.
[12] United States Census Bureau. (2021). *2019 Community Resilience Estimates: Quick Guide.* Last Accessed: 11/1/2021.
https://www2.census.gov/programs-surveys/demo/technical-documentation/community-resilience/2019/cre_quickguide_2019.pdf

most visible and visceral of these impacts. Because of this, understanding the likelihood of flooding at different levels of aggregation can be incredibly useful, but also incredibly difficult. For instance, in the aftermath of Hurricane Harvey, an independent analysis found that over 50% of the households within the city boundaries of Houston, TX that were flooded were outside the FEMA designated flood plain[13]. However, independent analysis after a weather event is hardly useful in understanding the changing character of flood risk, and the social and economic consequences that come from them. Further, our understanding of flood risk is often either taken from the official FEMA maps, which even the National Flood Insurance Program has found lacking[14], or based upon models that have been built locality by locality, introducing a patchwork of methodologies and assumptions. This is where new data sources like Flood Factor come in. Flood Factor, developed by the First Street Foundation, is a collaboration between over 80 experts to build a new model of flood risk at the building and parcel level (all the way down to a 3-meter level of resolution). They use historical data to build a model of flood risk, based on the same methodologies, across the United States, and then use changing environmental and climate factors to forecast risk out over 30 years[15].

An interesting piece of this data source is that the data are available all the way down to the building and parcel level, allowing for individual or household level linkages as well as aggregated merges. Available for non-commercial public use, the First Street Foundation has provided data at the congressional district, county, and zip code levels, while they have a paid product that provides data at a property level as well as hazard layers for geospatial analysis[16].

Also of note, the First Street Foundation is working on the development of a fire risk model, which would be of immense help to understand this growing area of risk for large swaths of the United States[17].

What Will Be Gained By Including Flood Factor in the HRS Contextual Data?

Data on the likelihood of future flooding as well as impacts of past flooding will allow for detailed research into the impacts of natural disasters on many forms of well-being for the population included in the HRS. Research has shown that the population over the age of 65 is often least prepared and most impacted by natural disasters[18], and the ability to provide data on actual

---

[13] City of Houston Housing and Community Development Department. (2019). *The Harvey Data Project: City of Houston Housing and Community Development Department.* Last Accessed: 11/1/2021.
https://www.civisanalytics.com/wp-content/uploads/2019/03/CityOfHoston_Report_Website-1.pdf
[14] National Flood Insurance Program. (2021). *Risk Rating 2.0 Methodology and Data Sources.* Last Accessed: 11/1/2021.
https://www.fema.gov/sites/default/files/documents/fema_risk-rating-2.0-methodology-data-sources_4-21.pdf
[15] First Street Foundation. (2020). *First Street Foundation Flood Model: Technical Methodology Document.* Last Accessed: 11/1/2021. https://assets.firststreet.org/uploads/2020/06/FSF_Flood_Model_Technical_Documentation.pdf
[16] First Street Foundation data access is laid out here: https://firststreet.org/data-access/. Many different terms are available, including partnerships that are granted through proposal submissions.
[17] Press Release. (2021). "First Street partners with Pyregence Consortium to build first property specific, climate adjusted fire risk model." Last Accessed: 11/1/2021.
https://firststreet.org/press/press-release-2021-pyregence-partnership/
[18] Shih, Regina A., Joie D. Acosta, Emily K. Chen, Eric G. Carbone, Lea Xenakis, David M. Adamson, and Anita Chandra. (2018). Improving Disaster Resilience Among Older Adults: Insights from Public Health Departments and

flood risk, both historically and projected, would provide further detail of such studies. However, the possibilities for the use of flood risk data do not end there. Policy decisions are likely to be made about how we fund things like the National Flood Insurance Program (NFIP): What impact will this have on the economics of retirement and the financial well being and health of an aging population? Other important and interesting questions could also be explored: Are there impacts for older community members who are not directly impacted by flooding, but live in communities with high risk or occurrence? Does increasing knowledge of natural disaster risk change behavior? Do local-level policy changes impact outcomes for the older population, despite high risk for climate induced flooding?

Add to this the potential in the next couple of years for a rigorous wildfire risk model, and enhancing HRS data with natural disaster risk data could lead to extremely strong avenues of research and policy work.

Logistical and Methodological Considerations

Is the resource public or private?: Public aggregates and private property level data
How can the data be accessed?: Through public access (AWS or ESRI) and private website
Unit(s) of analysis provided: Building/Unit level, county, congressional district, and
Update cadence: Yearly
Model or Observed Data?: Model
Transparency of sourcing or model: The models used to build the Flood Factor model are complex and complicated. For a review like this I could never do them justice, nor would my expertise allow me to. However, the models have been peer-reviewed as well as expert-panel reviewed before inclusion in any of Flood Factors outputs. Further information on the methodologies are also available from Flood Factor[19], and the overall models are relatively transparent.
Additional Considerations: First Street Foundation has already created a program to help researchers use their data, the Flood Factor Research Lab[20], perhaps making it easier for the HRS to get up and running with their more granular data.

Climate Check (Climate Check)

Why and How the Data Set Was Constructed

Like Flood Factor, Climate Check provides risk information at the property level for flooding. However, they also have risk models for several other types of climate induced natural disasters. These include: fire risk, extreme heat risk, storm risk, and drought risk[21]. Similar to Flood Factor, Climate Check employs a group of experts to develop these models, and they

Aging-in-Place Efforts. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR2313.html.
[19] First Street Foundation. (2020). *First Street Foundation Flood Model: Technical Methodology Document.* Last Accessed: 11/1/2021. https://assets.firststreet.org/uploads/2020/06/FSF_Flood_Model_Technical_Documentation.pdf
[20] Information from the First Street Foundation website (last accessed 11/1/2021): https://firststreet.org/research-lab/
[21] Climate Check. (2021). *Methodologies.* Last Accessed: 11/1/2021. https://climatecheck.com/our-methodologies

focus their use cases strongly on the real estate sector both for individuals as well as the businesses involved. Unlike the First Street Foundation which is a non-profit, they are a for profit company. There are both benefits and drawbacks of each of these organizational models, but the HRS should consider which (or if both) would better serve the needs of the study over the long term. Questions that should be considered here involve the long-term planning for the data as well as the impacts of the assumptions that were built into the models for the specific use cases the models were designed for (again, mostly the real estate industry and individual home buyers).

Much like the First Street Foundation, Climate Check started with historical data and models while adding state of the art statistical and machine learning models to build hazard models for each of the above risks. Data are less available publically than they are for the First Street Foundation's models, but there are some publicly available data at the state or property levels[22]. Overall, the additional value of further risk models for use together with the HRS could be a big plus for researchers as long as Climate Check is able to provide visibility and transparency in their algorithms.

What Will Be Gained By Including Climate Check in the HRS Contextual Data?

The key difference between what might be gained by Climate Check in comparison to the Flood Factor models from First Street Foundation are based mainly on the additional risk categories that are included in the modeling. This would allow researchers using the HRS to expand the analyses noted above beyond the flooding (and potentially wildfire in the future) contexts. This is important because while flooding and fires are the most visible and fear inducing natural disasters that society experiences, issues like increased extreme heat and drought can have large, but less visible impacts on the well-being and economic standing of our aging population in the United States.

Logistical and Methodological Considerations

Is the resource public or private?: Mostly private, minimal public data on their website.
How can the data be accessed?: Through private website or API
Unit(s) of analysis provided: Building or parcel level, with some state and city aggregates on their website.
Update cadence: Yearly
Model or Observed Data?: Model
Transparency of sourcing or model: The models used to create the estimates are explained to some extent on the Climate Check website[23]. It is my suggestion that further discussion with Climate Check about methodology would be required to decide to use their data.

---

[22] For example, data for the state of California are available here (last accessed: 11/1/2021): https://climatecheck.com/california
[23] Climate Check. (2021). *Methodologies.* Last Accessed: 11/1/2021. https://climatecheck.com/our-methodologies

Eviction Lab

Why and How the Data Set Was Constructed

The impact of, and moratorium on, evictions has perhaps received the most attention it ever has in public policy circles and popular media during the COVID-19 pandemic. This is for obviously important reasons, but the impacts of evictions on health and economic well-being have always been an issue that has garnered some attention. The issue that has made it hard to study is the difficulty or expense in collecting data on evictions. In fact, many states and localities make it extremely difficult to get data on evictions at all. Luckily, the Eviction Lab at Princeton University has tried to change this situation. They have collected as much data as possible on evictions in 48 out of the 50 states in the US over the time period of 2000 to 2016 to create county and neighborhood level estimates of eviction rates[24]. Data are collated, purchased, or hand coded to create the fullest accounting of eviction data available. Further, they track data on an ongoing basis in 6 states and 31 cities that is published weekly since March of 2020 to track the impacts of eviction moratoriums[25].

What Will Be Gained By Including the Evictions Lab data in the HRS Contextual Data?

Many studies show the impact of individual level housing security and type, as well as community-level housing characteristics on health and well-being outcomes for older adults using HRS data. For instance, Friedman and colleagues show that living in a county with high levels of housing foreclosures has a detrimental impact on cognition in older adults[26]. Further, Mawhorter and colleagues find the older renters, especially those who experience financial strain and live with poor housing conditions, experience higher cardiometabolic risk levels[27]. Adding neighborhood and county level eviction rates into these types of analysis could further refine the individual and community impacts of housing conditions. Questions answered could include: Do we see spillover effects of high eviction rates in counties and neighborhoods for older adults in terms of cognition and other health measures? Do community-level eviction factors play a role in increased cardiometabolic risk levels as see in Manwhorter and colleagues work? What are the economic impacts of living in high eviction areas and being an older renter, versus low eviction areas and being an older renter?

More recent data in the subset of cities and states that the Eviction Lab is collecting ongoing data in would allow researchers to address the questions of the impacts of eviction changes

---

[24] Eviction Lab. (2018). *Eviction Lab: Methodology Report.* Last Accessed: 11/1/2021. https://evictionlab.org/docs/Eviction%20Lab%20Methodology%20Report.pdf
[25] Eviction Lab. (2021). *Eviction Tracking System Methodology.* Last Accessed: 11/1/2021. https://evictionlab.org/eviction-tracking/methods/
[26] Friedman, Esther M, Jason N Houle, Kathleen A Cagney, Mary E Slaughter, and Regina A Shih. (2021). "The Foreclosure Crisis, Community Change, and the Cognitive Health of Older Adults", *The Journals of Gerontology: Series B*, 76, 5: 956–967, https://doi.org/10.1093/geronb/gbaa047.
[27] Mawhorter, Sarah, Eileen M. Crimmins, and Jennifer A. Ailshire. (2021). "Housing and cardiometabolic risk among older renters and homeowners", *Housing Studies*, DOI: 10.1080/02673037.2021.1941792

during COVID-19 moratoriums on the health and economic well-being of the population covered by the HRS.

Logistical and Methodological Considerations

Is the resource public or private?: Public
How can the data be accessed?: Through public website or API
Unit(s) of analysis provided: Census tract, county, state, and national
Update cadence: Yearly
Model or Observed Data?: Model
Transparency of sourcing or model: The model used to create the estimates are highly transparent through the Census Bureau's methodology documentation ([cite])

Conclusion

There is no question that the contextual data file is a boon to research for the HRS. The inclusion of these data make possible work that would otherwise be much more time intensive or would be altogether impossible. Through this review, I have selected 5 potential new data resources that could increase the utility of the HRS contextual data resource, and by extension the overall HRS. There is little doubt that the twin crises of the pandemic and the increasing impacts of climate change have necessitated even greater research needs to understand and improve the outcomes for society. This is especially true for the aging population that is covered by the HRS. As I noted in the introduction, the goal of my review was to get the research community thinking about potential different sources of data that may be brought to bear on these issues. There are surely other topics and data sources that would be important to the research community that were not touched upon here. That being said, having a better understanding of the role resiliency, climate induced natural disasters, and community issues such as eviction play at both the individual and community levels could certainly push our understanding, and relatedly, outcomes, forward.