

UCLA

Data Sets

Title

Constructing a Work History Dataset of Jobs Held During Early and Middle Adulthood Using the Health and Retirement Study

Permalink

<https://escholarship.org/uc/item/3bz58411>

Authors

Park, Sung S
Pratt, Borianna
Pebley, Anne R
[et al.](#)

Publication Date

2022-08-31

Constructing a Work History Dataset of Jobs Held During Early and Middle Adulthood Using the Health and Retirement Study

v. 8/24/22

Sung S. Park^a, Borianna Pratt^a, Anne R. Pebley^c, Noreen Goldman^b, Mara Getz Sheftel^d,
Theresa Andrasfay^e, and Keunbok Lee^f

^a Office of Population Research, Princeton University, USA

^b Office of Population Research, School of Public and International Affairs, Princeton
University, USA

^c California Center for Population Research, Fielding School of Public Health, University of
California, Los Angeles, USA

^d Population Research Institute, Penn State University, University Park, PA, USA

^e Leonard Davis School of Gerontology, University of Southern California

^f Department of Epidemiology and Biostatistics, University of Maryland School of Public Health

Corresponding Author: Sung S. Park (sung.s.park@princeton.edu)

Funding: Research reported in this publication was supported by the National Institute on Aging through grant number R01AG061094 and by the Eunice Kennedy Shriver National Institute on Child Health and Human Development through grant number P2CHD041022.

Keywords: Occupation, Employment, Health and Retirement Study, Life Course, Aging, Functional Limitations

Acknowledgments: We are grateful to Trivellore Raghunathan at the University of Michigan for his valuable advice on multiple imputation and for making his software available. The authors also thank Narayan Sastry and Michael Handel for their helpful comments.

Abstract

The Health and Retirement Survey (HRS) (<https://hrs.isr.umich.edu/about>) collects extensive data on current employment and occupation at each wave but data on occupations prior to the start of the survey are limited to an abbreviated job history in each respondent's first interview focused on recently held jobs. Therefore, using HRS data to link employment earlier in respondents' lives to socioeconomic, health, and other outcomes at older ages has been infeasible. The RAND Corporation created a dataset from the HRS Core and Exit Interviews called the RAND HRS Cross-Year Longitudinal file which is used by many researchers working with HRS. This dataset contains a variable called the "longest job held" for each respondent. However, this variable is the longest job held among the limited subset of jobs reported in HRS (most of which are recent) and not necessarily the longest job that the respondent has held to date.

In 2017, HRS conducted a Life History Mail Survey (LHMS) of HRS respondents who had participated in the 2016 Core interview. The LHMS was conducted in three parts which, combined, include all of the 2016 Core interview respondents. The LHMS questionnaire was a pencil-paper survey completed and returned by respondents and included a grid asking

respondents to list and provide information for "all the places you have worked for one year or more after you finished full-time education." Respondents were asked about the start and end date of each job, type of employer, job title, the year the job started and ended, and whether the job was full- or part-time. If the total number of jobs was more than 10, respondents were asked to report on the first 10.

To use data from the LHMS employment histories, it was essential to solve two problems. First, the LHMS retrospective employment history needed to be merged and harmonized with the employment data reported contemporaneously in each wave of the HRS Core in order to construct a complete work history. Second, the data required significant cleaning in order to address data issues such as data omission, duplicated jobs, and inconsistent reports of jobs. To do so, we developed a systematic procedure for constructing employment histories which is summarized in this document.

Data and Sample

The work history file created for this study contains a record for each job reported by each respondent starting from age 15. Each record contains the job tenure, start age, end age, full-time/part-time (FT/PT) status, occupation code, and O*NET score measuring average physical demands associated with that occupation code. The construction of the work history file involved using the following HRS-related data sources: the 2016 RAND cross-wave longitudinal file and three self-administered life history mail survey files which we refer collectively as LHMS: the 2017 Spring Life History Mail Survey, the 2017 Fall Life History Mail Survey (full), and the 2017 Fall Life History Mail Survey (supplement). Because the 2017 LHMS were the source for much of the retrospective work history data in the analytic sample¹, all individuals in the work history file are HRS respondents who were reinterviewed in the 2016 core and still living in 2017. The final respondent-job file consists of N=49,462 respondent-job records from N=10,216 respondents in the mid-Baby Boomer cohort or older cohorts, which includes N=276 individuals who never report any paid work in the LHMS or Core data.

Data Aggregation and Harmonization

We applied the following logic in constructing the final work history file. We constructed a preliminary *HRS Core work history file* using the employment variables in the 2016 RAND cross-wave longitudinal file: FT/PT status (based on RAND-derived HRS measures on hours worked, where full-time is defined as working 35 or more hours per week) and the detailed 3- or 4-digit occupational code for each job. These data were augmented with data on jobs reported in the LHMS, resulting in a new *HRS Core-LHMS work history file*².

The combination of data on jobs from the Core and the LHMS hinged on an age we refer to as "point A," which was the start age (calculated based on the month and year of employment dates) for each respondent of the first job reported in the Core (based on the RAND-derived "current job tenure" variables). At the time of the first HRS Core interview for a given respondent, if the respondent was working for pay, the start age of the job held at that time was designated as their work history "point A". Alternatively, if the person did not work at the time of their first Core interview, we used the RAND-derived variables of the start month and year of the most recent job to calculate the start age and identify "point A". The two rules

¹ The combining of the three LHMS subsamples from 2017 covers the majority of HRS participants who were reinterviewed in 2016 and were still living in 2017.

² In cases where Rs completed a LHMS but did not report any work in this survey, we relied entirely on jobs reported in the Core interview.

that we followed in the cases described below are: 1. data on jobs reported prior to point A were obtained from the LHMS, and 2. data on jobs reported at or after a respondent's first Core interview were obtained from the Core. We illustrate our strategy with four examples that cover most of the reported job histories.

In instances where the first reported Core job started after the last reported LHMS job, we used data from the LHMS for earlier years, and then relied on Core data for the remaining years. An illustrative example is shown as Case 1 in the accompanying figure and table below.

In Case 2, when the respondent's first HRS job began before the first LHMS job, we relied on job data only from the Core.

In Case 3, where a respondent's point A fell between the end age of one job in the LHMS and the start age of the next job in the LHMS, we kept the job data from the LHMS prior to point A, and used data from the Core after point A. Data from the LHMS after point A that overlapped with jobs reported in the Core were not used because we expect higher accuracy with Core data.³

In scenarios where a respondent's point A fell within the period of a job reported in the LHMS, we had to determine if the LHMS job was actually the same job as the job in the Core. We examined the occupational codes and the FT/PT status of the overlapping jobs from the two data sources. If the occupational code or FT/PT status differed, we treated the LHMS job and the Core jobs as distinct jobs, truncated the duration of the LHMS job up to point A, and did not use any other LHMS data on jobs after point A (Case 4A). If the occupational code and the FT/PT status were the same, we used the start age of the job from the LHMS as an updated point A (Case 4B) and relied on data from both sources to identify the end age for that job (taking the later age). Then for subsequent jobs, we relied only on Core data from that point onward in constructing the work history file. For example, in Case 4B, the second job (2-L/1-C) starts at age 45 and ends at age 49 (the end age from the LHMS).

Imputation

We used the *HRS Core-LHMS work history file* to impute missing values⁴ for job tenure, FT/PT status, and occupation code variables using IVWare, a statistical program developed by the Survey Research Center at the University of Michigan (<https://www.src.isr.umich.edu/software/>). The program estimated separate linear regression (tenure), logistic regression (FT/PT status), and multinomial logistic regression (occupation code) models that include demographic variables; childhood variables⁵; and family, health, and wealth measures at first interview, outputting ten multiply imputed datasets.

³ The Core data, with a shorter period of recall, were collected by an interviewer, in contrast to the LHMS which was a self-administered pencil-paper questionnaire.

⁴ 14.6% of Rs were missing values for job tenure, 13.7% of Rs were missing values for FT/PT status, and 12.4% of Rs were missing values for occupational code. In the cases where the detailed occupation code for a job was missing, we imputed the two-digit occupation codes standardized to the Census 2010 occupational codes.

⁵ These variables included: parental education, whether a parent died before 16, number of people in household at age 10, number during childhood, speaking only English before age 18, poor health during childhood, whether respondent's father was unemployed during childhood, and whether lived in rural area during childhood.

Scenarios of Combining Information From the Life History (LHMS) and Core (C) Data

Scenario	LHMS Data			Core Data		
	Job #	Start Age	End Age	Job #	Start Age (Point "A")	End Age
Case 1: First job in Core follows last job in LHMS	Job 1-L	27	31	Job 1-C	40	53
	Job 2-L	31	34	Job 2-C	53	57
Case 2: First job in Core precedes first job in LHMS	Job 1-L	19	21	Job 1-C	18	51
	Job 2-L	59	67	Job 2-C	54	56
Case 3: First job in Core starts between end date of one LHMS job and start date of another LHMS job	Job 1-L	30	35	Job 1-C	36	46
	Job 2-L	37	45	Job 2-C	46	48
Case 4A: First job in core starts in the middle of a job in LHMS (not same job based on occ. code, FT/PT status)	Job 1-L	28	29	Job 1-C	31	45
	Job 2-L	30	46	Job 2-C	48	50
	Job 3-L	47	54			
Case 4B: First job in core starts in the middle of a job in LHMS (same job based on occ. code, FT/PT status)	Job 1-L	18	20	Job 1-C	46	48
	Job 2-L	45	49	Job 2-C	55	60
	Job 3-L	50	54			

Notes: 1-L: 1st (earliest) job in LHMS; 2-L: 2nd job in LHMS; 3-L: 3rd job in LHMS; 1-C: 1st (earliest) job in Core; 2-C: 2nd job in Core

