Health and Retirement Study Cross-Wave Harmonized Job Characteristics File for 1992-2018 HRS
Wave Years

V2023-05-05

Data Description

Joelle Abramowitz
Bo Klauth
Nada Wasi

May 2023

Questions about this dataset and its use can be directed to the Census-Enhanced Health and
Retirement Study (CenHRS) project at CenHRSinfo@umich.edu.

1   Overview

This document describes the construction of the variables from the Health and Retirement Study (HRS) cross-wave harmonized job characteristics file (JCF) for waves 1992-2018. The JCF contains harmonized information on HRS respondents' current main jobs at the time of the interview at the person-wave level. The purpose of this dataset is to link main jobs across waves and impute relevant variables that are missing or inconsistent. We produce this harmonized file because the characteristics of main jobs are complicated to track since information about these jobs is collected at different interview waves. Some characteristics are collected only at the first time the job is reported, some are collected every wave, and some are collected after the job ended.

The JCF dataset contains both harmonized HRS variables and constructed variables. The harmonized HRS variables represent original HRS variables renamed so that the same variables have the same names in all waves and revised with consistent codes for missing values. The constructed variables recode these harmonized HRS variables so that they are comparable across waves and correct or impute missing or conflicting values. For example, the HRS variables of whether the respondent is working for pay are V2717 for the 1992 wave year; V1174 for 1993; W3316 for 1994; D2651 for 1995; and so on. In the JCF, these variables are renamed as "*WORK4PAY*". In the 1992-1995 waves, HRS used two or more variables to record respondents' answers to the same question asked in different branches of the survey. For example, HRS used three variables to record the year the respondent stopped working for the previous wave employer (W3504, W4201, and W4898) for the 1994 wave year. In the JCF, the single variable "*YRSTOPEMP*" compiles the information from these three variables. Associated flags are constructed to indicate whether values are imputed and how they are imputed. The JCF does not contain all time-variant job characteristics, but users can easily merge those variables by respondents' household and person identification numbers (HHIDPN) and wave to obtain such characteristics.

For users who are familiar with the raw HRS datasets (or RANDFAT) and the RANDHRS dataset, the JCF can be viewed as an intermediate version between the two. While working with the raw HRS variables provides users flexibility in coding and imputing missing or inconsistent information, it is very tedious and time-consuming. The RANDHRS, on the other hand, includes RAND's constructed variables and associated imputation codes. The JCF contains both harmonized original HRS variables and constructed variables to enable users to ascertain how these variables are constructed and create alternative imputation rules if needed.

This document proceeds by describing the harmonized original HRS variables and constructed variables in the JCF. More detail on the variable values can be found in the JCF codebook. More information on the source of the original HRS variables is provided in the appendix.

2 Harmonized Original HRS Variables

The original HRS variables included in the JCF are listed in Table 1. These variables come from three sources: the HRS Core Interview, the HRS Exit Interview, and the HRS Tracker file. The variables are linked by respondent and wave using the identifier variables, which can be used to merge the JCF to other HRS datasets. The JCF codebook provides details on values for these variables.

Table 1: Harmonized Original HRS Variables

Source	Variables	Description
Identifiers	HHIDPN	HHID + PN (Numeric)
	HHID	Household Identification Number
	PN	Person Identifier
	OVHHID	Overlap Case: Old HHID
	OVPN	Overlap Case: Old PN
	WAVEYR	Wave Year R is in the Study
Core interview	WORK4PAY	Whether R is Working for Pay
	SELFEMP	Whether R is Self-Employed
	SAMEEMP	Whether R Works for the Same Employer
	YRSTARTNEWJ	Which Year R Started Current Job
	YRSTARTNEWJ_YRSAGO	How Many Years Ago Current Job Started
	YRSTARTNEWJ_AGE	At Which Age R Started Current Job
	MOSTARTNEWJ	Which Month R Started Current Job
	YRSTOPEMP	Which Year R Stopped Working for Previous Employer
	MOSTOPEMP	Which Month R Stopped Working for Previous Employer
	YRSTOPSELF	Which Year R Stopped Working for Self
	MOSTOPSELF	Which Month R Stopped Working for Self
	REV_YRSTART	Revised Year Started Current Job
	REV_YRSTOP	Revised Year Stopped Previous Job
	YR_RET	Which Year R Retired
	YR_DIS	Which Year R Became Disabled
	YR_UNEMP	Which Year R Became Unemployed
Exit interview	XWAVEYR	Main Interview Year of Wave (Exit Interview)
	XYRSTOPEMP	Which Year R Stopped Working for Previous Employer (Exit Interview)
	XMOSTOPEMP	Which Month R Stopped Working for Previous Employer (Exit Interview)

	XYRSTOPSELF	Which Year R Stopped Working for Self (Exit Interview)
	XMOSTOPSELF	Which Month R Stopped Working for Self (Exit Interview)
Tracker	BIRTHYR	R's Birth Year
	BIRTHMO	R's Birth Month
	DTHYEAR	R's Coalesced Death Year
	DTHMONTH	R's Coalesced Month Year
	IWYEAR	Actual Interview Year
	IWMONTH	Interview Month
	STUDY	Study Membership
	WGTR	Sampling Weight

3   Constructed Variables

The JCF contains a total of 15 variables constructed from the original HRS variables, including five flag variables. Table 2 lists the constructed variables. The JCF codebook provides details on values for these variables. We next describe how these variables are constructed.

Table 2: Constructed Variables

Constructed Variables	Description
WORKSTAT	Work Status
FLAG_WORKSTAT	Flag of WORKSTAT
SAMEJOB	Whether Current Job is the Same Job as in Previous Wave
FLAG_SAMEJOB	Flag of SAMEJOB
MAINJOBID	Identifier of Main Jobs
YR1STHIRE	Year R was First Hired
FLAG_YR1STHIRE	Flag of YR1STHIRE
MO1STHIRE	Month R was First Hired
YRSTART_JBSP	Year This Job Spell Started
FLAG_YRSTARTJBSP	Flag of YRSTART_JBSP
YRSTOP_JBSP	Year This Job Spell Ended
MOSTOP_JBSP	Month This Job Spell Ended
FLAG_YRSTOPJBSP	Flag of YRSTOP_JBSP
IWSEQ	Respondent Interview Sequence Number
JOBSEQ	Job Sequence Number

3.1   Work Status

The constructed variable *WORKSTAT* takes four values: (1) employee or wage and salary sector;

4

(2) self-employment; (3) working, self-employment status unknown; and (4) not working. The original variables used to construct work status (*WORKSTAT*) were *WORK4PAY* and *SELFEMP*. The *WORK4PAY* variable contains the answer to the question, *"Are you doing any work for pay at the present time?"* If the answer to this question was affirmative, the respondent was then asked about self-employment, *"Next are some questions about your current, main job. Do you work for someone else, are you self-employed, or what?"* and the answer was recorded in the variable *SELFEMP*. These questions were asked every wave and were not conditional on past information. The variable *WORKSTAT* was imputed from these two variables when their values were not missing.

When either *WORK4PAY* or *SELFEMP* were missing, work status was imputed from other variables when possible. For example, if these two variables were missing, but the variable *MOSTOPEMP* indicated that the respondent still worked either for the same employer or themselves at the time of the interview, then the respondent was imputed to be working for the employer or themselves, as reported in the previous wave. As another example, if there was no information about the respondent's current job, but there was information about the year they retired or stopped working for their previous employer, *WORKSTAT* was imputed as not working. *FLAG_WORKSTAT* indicates whether the value for *WORKSTAT* was imputed.

3.2   Same Job

The constructed variable *SAMEJOB* is a dummy variable indicating whether the respondent currently works in the same job as the previous wave (1 for the same job and 0 otherwise). *SAMEJOB* was constructed from the original variable *SAMEEMP*, which connects jobs across waves. For respondents who reported working for someone else in both the previous and current survey waves, *SAMEEMP* contains responses to the question, *"According to our records, in [previous wave month/year] you were working for [previous wave employer name]. Are you still working there?"* The response categories for this question have changed slightly in recent waves. Prior to 2008, the response categories included same employer, different employer, and other special codes for don't know or refused. Since 2008, a new response category, whether employer ownership has changed, was added. Self-employed respondents were not asked this question.

To construct *SAMEJOB*, respondents who said they have the same job but ownership has changed were treated as holding the same job. Respondents who were self-employed in the previous interviewand continued to work for themselves were coded as working at the same job. Respondents who switched between self-employment and wage and salaried employment were coded as having a different job. For jobs at the initial interview, *SAMEJOB* was coded as ".b" (baseline job). When a respondent's work status in the previous wave was "not working" or missing, *SAMEJOB* was coded ".r" (reentry to the labor market).

If *SAMEEMP* was missing, but valid records about when the respondent left their last job or started their new job existed, then *SAMEJOB* was imputed to be 0 (new job). This imputation reflects that the respondent reported that they started a new job. When *MOSTOPEMP* indicated that the respondent was still working for the same employer or working for themselves, *SAMEJOB* was imputed to be 1 (same job). This imputation reflects that the respondent likely initially reported that they switched jobs, but then corrected that they still work on the same job. When there is no other information available, those who continue to work in the wage and salary sector without information about job changes were assumed to hold the same job. *FLAG_SAMEJOB* indicates whether the value for *SAMEJOB* was imputed.

3.3 Interview Sequence Number, Main Job ID, and Job Sequence Number

The Interview Sequence Number (*IWSEQ*), Main Job ID (*MAINJOBID*), and Job Sequence Number (*JOBSEQ*) variables indicate the sequence of interviews, main jobs, and job reports for a respondent across waves. For each respondent, the Interview Sequence Number (*IWSEQ*) variable assigned a sequence number for each interview beginning with their first interview and incrementing by 1 for each subsequent interview. For each respondent, the *MAINJOBID* variable assigned a job identification number for current main jobs, beginning with their first reported current main job and incrementing by 1 for each report of a new current main job using the job transition information contained in the variable *SAMEJOB*. In cases where the interview in the previous wave was incomplete and all variables in the job section were missing, *MAINJOBID* considered the job in the subsequent wave to be a new job. For each *MAINJOBID*, the *JOBSEQ* variable assigns a sequence number for each report of that job, beginning with the first report of a given current main job and incrementing by 1 for each subsequent report that the respondent continues to hold that job.

3.4 Month and Year of First Hire

The constructed variables *YR1STHIRE* and *MO1STHIRE* indicate the year and month the respondent was first hired at a given job. *YR1STHIRE* and *MO1STHIRE* were constructed from the harmonized HRS variables for the respondent's job start year and month, *YRSTARTNEWJ* and *MOSTARTNEWJ*, as well as the information contained in *MAINJOBID*.

For respondents who worked in the wage and salary sector, information about the job start date was collected the first time they reported a job and then again every fourth wave (e.g., 1992, 2000, and 2008). The question asked, *"When did you start working for this employer?"* The interviewer was instructed, "*If the respondent worked more than once for this employer, ask about the first hire."* In each wave, self-employed respondents were asked about the month and year they started their businesses regardless of whether it was the same business they had in the previous wave.

Both month and year were asked from 1994 onward. In 1992 and 1993, the survey asked only for the start year. In 1993, the survey allowed the respondent to answer in three different formats: calendar year (*YRSTARTNEWJ)*, years elapsed since the job started (*YRSTARTNEWJ_YRSAGO*), and the age when the respondent started this job (*YRSTARTNEWJ_AGE*). In the 1992-1995 waves, the HRS collected this information in multiple variables (see the Appendix). *YRSTARTNEWJ* and *MOSTARTNEWJ* collect information from these different variables into a single variable. From 2002 onward, in addition to the questions about the job start month and year, respondents who switched jobs were asked to confirm the year they started their current job again in a later part of the survey. The additional question asked, *"You told us that you started your current job on … and left your previous job on … Is that correct?"* If the respondents said the information was not correct, they were asked the job start year again. The variable *REV_YRSTART* contains this information.

As described previously, not all harmonized HRS variables have values in every wave. The constructed variables *YR1STHIRE* and *MO1STHIRE* compile information across waves to include the job start year and month in every wave for a given *MAINJOBID*. When *REV_YRSTART* exists, it was treated as a truthful answer, and *YR1STHIRE* was constructed based on that information. When multiple start years were recorded for a single job (always the case for continuing self-employed jobs), the most frequently reported (modal) valid value was used. If there were multiple modal values, the minimum modal value was used. Values were considered to be valid if the calculated starting age at the job was 15 or older. Unlike the original HRS variables, which contain numerical codes for missing values, Y*R1STHIRE* and *MO1STHIRE* only contain actual year and month values. When respondents reported starting the job before age 15 and there was no valid alternative value available, *YR1STHIRE* was set to equal the respondent's year of birth plus 15. The flag variable FLAG_YR1STHIRE indicates whether the value of YR1STHIRE was imputed.

3.5   Job Spell Start Year

YRSTART_JBSP (year started this job spell) is a constructed variable that compiles information on the year the job spell began. For most respondents, *YRSTART_JBSP* was equal to *YR1STHIRE*. However, for some respondents, their *YR1STHIRE* was before the end year of their previous job, including respondents whose current jobs were not the first job spell for this employer. For these respondents, we imputed *YRSTART_JBSP*. Since *YR1STHIRE* is imputed to be the minimum modal year value when the respondent reported multiple job start years, to address this issue, we instead imputed *YRSTART_JBSP* as the maximum modal year value if that year was the same as or after the previous job ended. If no alternative start year was available or *YR1STHIRE* was missing, *YRSTART_JBSP* was imputed as the current interview year (*IWYEAR)*. The flag variable *FLAG_YRSTARTJBSP* indicates whether *YRSTART_JBSP* was imputed and the source used for imputation.

3.6   Job Spell Stop Month and Year

The constructed variables *MOSTOP_JBSP* (Month This Job Spell Ended) and *YRSTOP_JBSP* (Year This Job Spell Ended) collect information across all waves on when a respondent's job spell ended from the original HRS variables *MOSTOPEMP, YRSTOPEMP, MOSTOPSELF, and YRSTOPSELF* from the core interviews and *XYRSTOPEMP, XMOSTOPEMP, XYRSTOPSELF, XYRSTOPSELF,* and *XMOSTOPSELF* from the exit interviews to include this information in the waves the respondent held the job. It is important to note that these original HRS variables variables collected information on the job stop month and year for a job that the respondent held during the prior wave's interview but left before the current wave's interview. Accordingly, the waves in which this information is collected were not the waves in which the respondent held the job. *MOSTOP_JBSP* and *YRSTOP_JBSP* contain information on when the job spell ended for all waves the respondent held the job.

If a respondent reported they had left their previous wave employer, they were asked about the month and year they stopped working. The question prompt for collecting this information varied according to the respondent's previous wave work status. Respondents who had worked for someone else were asked, *"According to our records, you were working for [previous wave employer name] in [previous wave IW month and IW year]. In what month and year did you stop working for that employer?"* The responses to this question were recorded in the variables *MOSTOPEMP* and *YRSTOPEMP* for month and year, respectively. Respondents who had been self-employed were asked, *"According to our records, in [previous wave IW month and IW year] you were self-employed. In what month and year did you stop working for yourself?"* The answers for the month and year were recorded in the variables *MOSTOPSELF and YRSTOPSELF*, respectively.

For deceased respondents, an exit interview was conducted with a "proxy informant," generally a widow, widower, or some other family member. The proxy was asked questions about the health, family, employment, and financial situation of the deceased. The exit interviews started in 1995, and most questions mirror those in the core interviews. For example, for deceased respondents who had been self-employed, their proxies were asked, *"Now I'd like to ask you some questions about (his/her) employment situation since we last spoke to (him/her) before (his/her) death. According to our records, in [previous IW month and IW year] [R first name] was self-employed. In what month and year did (he/she) stop working for (him/her)self?"* Analogous to the variables of the core interview, the variables from exit interviews that recorded these answers were *XYRSTOPEMP* (year stopped, employee), *XMOSTOPEMP* (month stopped, employee), *XYRSTOPSELF* (year stopped, self-employed), and *XMOSTOPSELF* (month stopped, self-employed). The variable *XWAVEYR* is the wave year that the exit interview was conducted. The JCF merged these variables from the exit interviews to the last core interview wave for each deceased respondent.

Since 2002, respondents who switched jobs were asked to confirm the previously reported year

they stopped working. If they corrected their previous response, the new response was recorded in the variable *REV_YRSTOP*. Similar to the job start year, if the revised answer (*REV_YRSTOP*) existed, it was treated as a truthful answer. Note that the harmonized HRS variables contain both actual months and years and other values representing responses such as "still working for the previous job," "still working for self," or "work until death" for the exit interviews.

To construct *MOSTOP_JBSP* and *YRSTOP_JBSP*, extensive reconciliation of missing or conflicting information from the original HRS variables was necessary, as follows:

- If the year the respondent reported they stopped working was missing and the month the respondent reported they stopped working was not missing and did not indicate that they still worked for the same job, then the value for the month the respondent stopped working was treated as missing.
- If the year the respondent reported they stopped working was not missing and the month the respondent reported they stopped working indicated that they still worked for the same job, then the value for the year was treated as missing.
- For jobs not yet ended in the latest wave year available (i.e., 2018), *YRSTOP_JBSP* was treated as missing.
- If a respondent reported transitioning to non-employment in a subsequent wave but did not report the year they stopped working at their previous job, *YRSTOP_JBSP* was imputed from the variables *YR_RETIRE* (year retired), *YR_UNEMP* (year unemployed), and *YR_DS* (year became disabled) in the subsequent wave if this year was between the last time the respondent was working at this job and the time the respondent left the job.
- If a respondent moved to another job in a subsequent year, and the year the respondent stopped working at the previous job was missing, then *YRSTOP_JBSP* was imputed to be the interview year (*IWYEAR*) the respondent last reported that job.
- If a respondent attritted from the sample and as such had no information on work status in the subsequent wave, *YRSTOP_JBSP* was coded as missing.
- Some deceased respondents had no exit interview conducted with a proxy informant or had inconsistent or missing information reported about when they stopped working. When a proxy informant reported the year the deceased respondent stopped working as after the respondent's death year as recorded in the tracker file, *YRSTOP_JBSP* is imputed to be the tracker file death year. If the proxy informant reported that the respondent stopped working before the last year the respondent themself reported working, *YRSTOP_JBSP* is imputed to be the last interview year. For other deceased respondents whose missing information about when they stopped working, *YRSTOP_JBSP* was imputed to be the last interview year the respondent reported working.
- If a respondent was reported to "work until death" in their exit interview, *MOSTOP_JBSP* and *YRSTOP_JBSP* were coded as the respondent's death month and death year.

The flag variable *FLAG_YRSTOPJBSP* indicates whether *YRSTOP_JBSP* was imputed and the source used for imputation.

4   Conclusion

The JCF contains harmonized information on HRS respondents' current main jobs at the time of the interview at the person-wave level. The purpose of this dataset is to link main jobs across waves and impute relevant variables that are missing or inconsistent. This document described the harmonized original HRS variables and constructed variables in the JCF. More detail on the variable values can be found in the JCF codebook.

Appendix

Appendix Table 1: HRS Core Interview Variables Used to Track Main Jobs

Variable	Description	1992	1993	1994	1995	1996	1998	2000	2002-2018
WORK4PAY						E2627	F3131	G3381	*J020
SELFEMP						E2628	F3132	G3382	*J021
MOSTOPSELF						E2630	F3134	G3384	*J023
	If not working this wave			W4800	D3545				
	If working for someone else this wave			W3318	D2654				
YRSTOPSELF						E2631	F3135	G3385	*J024
	If not working this wave			W4801	D3546				
	If working for someone else this wave			W3319	D2655				
SAMEEMP					D2747	E2655	F3166	G3416	*J045
MOSTOPEMP						E2667	F3188	G3437	*J063
	If not working this wave			W4897	D3639				
	If self-employed this wave			W4200	D3188				
	If working for another employer this wave			W3503	D2759				
YRSTOPEMP						E2668	F3189	G3438	*J064
	If not working this wave			W4898	D3640				
	If self-employed this wave			W4201	D3189				
	If working for another employer this wave			W3504	D3160				
MOSTARTNEWJ									*J248
	If self-employed			W4327	D3273				
	If working for someone else			W3662	D2914				
YRSTARTNEWJ						E2826	F3349	G3608	*J249
	At which age R started this job		V1184						
	Current work start year		V1182						
	How many years ago R started this job		V1183						
	If self-employed	V2834		W4328	D3274				
	If working for someone else	V2816		W3663	D2915				
REV_YRSTOP									*J254
REV_YRSTART									*J256

Note: * represents the wave letter for the respective year.

Appendix Table 2: HRS Exit Interview Variables Used to Track Main Jobs

Variable	1995	1996	1998	2000	2002-2018
XMOSTOPSELF	N3545	P1745	Q2185	R2202	*J023
XYRSTOPSELF	N3546	P1746	Q2186	R2203	*J024
XMOSTOPEMP	N3639	P1814	Q2254	R2271	*J063
XYRSTOPEMP	N3640	P1815	Q2255	R2272	*J064

Note: * represents the wave letter for the respective year.