Creating a Public Resource: O*NET Job Characteristics Data Set for Use with the Health and Retirement Study and Other Surveys

*Brooke Helppie-McFall, Dawn Carr, and Amanda Sonnega**

Background

A growing body of research examines the relationship between occupational experiences and later-life outcomes. Many studies that have examined the impact of job characteristics on retirement behavior have used the rich data resources available in the Health and Retirement Study (HRS). Over the last decade, some studies have sought to expand their characterizations of work environments and experiences by adding information available in the Occupational Information Network (O*NET) database. Sponsored by the U.S. Department of Labor's Employment and Training Administration, the O*NET database provides detailed occupational information based on a combination of surveys of workers and expert assessments. The

resulting measures can be thought of as population average ratings of the characteristics of particular occupations. With a growing need to understand how individuals within similar work environments experience different outcomes, particularly among those from under-represented groups (e.g., racial and ethnic minorities), pairing O*NET data with HRS data has important implications for scientific advances that can promote more equitable work environments, health and well-being.

This work arose in response to the need for a publicly-available, validated, high-quality HRS-O*NET linked data set. As a first step in this direction, we held a workshop in July 2020, sponsored by the MRDRC, that brought together a multidisciplinary group of researchers (in both academic

* **Brooke Helppie-McFall** is an associate research scientist at the University of Michigan's Survey Research Center. **Dawn Carr** is an associate professor at Florida State University in the Department of Sociology and faculty associate at Pepper Institute for Aging and Public Policy. **Amanda Sonnega** is associate research scientist in the Health and Retirement Study program at the University of Michigan's Survey Research Center. This brief is based on project UM22-Q1.

and government settings) who were currently engaged in research involving O*NET data linked to the HRS and other data sets, such as Mid-life in the United States (MIDUS). Practical issues discussed included current practices for linkage, limitations of these methods, best practices, and how to make linked data available to the wider research community.

Following the workshop, we formed a working group to create a linkage between the 2019 O*NET measures and the 2010 Census occupational code frame, which has been used by the HRS since 2010. In the first stage of our work, we created a crosswalk between the 2019 O*NET-SOC occupational code frame and the 2010 Census occupational code frame (used in HRS from 2010 forward) and created summary measures of the O*NET measures at the 2010 Census occupation level. In spring 2021, we held a meeting to obtain feedback on our proposed linkage from the researchers who were part of the working group. We met with Kathleen Mullen (RAND then, University of Oregon now), Jim Grosch (National Institute for Occupational Safety and Health), Ross Andel (University of South Florida), and Pam Frugoli (DOL, O*NET project lead). Feedback from these discussions led to a small number of modifications to the O*Net 26-Census 2010 data set that we implemented before linking to the detailed occupation data in the HRS. We completed the O*NET 26-HRS data linkage and documentation report. The data were released as a restricted HRS data product in summer 2022, with the documentation report and codebooks available on the HRS website. The O*NET 26-Census 2010 data set and a separate documentation report were also posted on the Florida State University Pepper Center website in summer 2022.

In the second stage of this work, we linked the 2000 Census occupation codes to SOC 2000 and from there to the O*NET SOC 2000 and O*NET SOC 2006 taxonomies. To do this, we used existing crosswalks posted on the BLS and O*NET websites. The 2000 Census occupation to SOC crosswalk is available on the BLS website. To make the slightly more-detailed O*NET SOC 2000 and O*NET SOC 2006 from the SOC 2000 code frame, two additional digits were added to the SOC codes. The resulting codes clearly map to SOC 2000 codes. At this stage we obtained feedback from Pam Frugoli, and again, made a few modifications.

We were then in a position to create historical O*NET-Census 2000 data sets for the following O*NET releases: 5.0 (released April 2003)[**] and 10.0 (released June 2006). For cases in which there is more than one O*NET SOC code for a census code, as in the O*NET SOC 2019-Census 2010 linkage, measures in these data include medians or weighted means, as well as statistics about the distribution of these measures among SOC codes comprising a census code. Finally, we linked the historical O*NET-Census 2000 data sets to HRS occupation data. We linked the O*NET 5.0-Census 2000 data set and the O*NET 10.0-Census 2000 data to occupations in the HRS data from 2004 through 2010.

Through an NIA supplement to the HRS grant, work is currently underway to recode the HRS historical occupational information from 1992 to 2002 to the 2000 and 2010 Census occupational code frames. Once complete, expected in late 2023, it will be straightforward to link the HRS data from 1992 to 2002 to the O*NET-Census data sets created by this project.

Validation analyses

We conducted quality control checks throughout the process of creating the linkages to check for coding errors.

** O*NET 5.0 was the first release in which all occupations had data from the methodology that is used to this day, and is thus more comparable to future releases. The O*NET website states that "[p]rior to the transition to the current multi-method data collection methodology featuring Job Incumbent, Occupational Expert, Big Data, and other sources, the O*NET databases from O*NET 98 to O*NET 4.0 were populated using data supplied by occupational analysts. Using earlier occupational taxonomies, analysts evaluated and refined existing Dictionary of Occupational Titles (DOT) data, then extrapolated these data to the O*NET Content Model." (see https://www.onetcenter.org/db_transitional.html)

To establish the validity of the linked measures, however, we conducted a series of analyses to establish that our "historical" O*NET measures correlate with responses to related questions about job characteristics in the HRS and are sensibly correlated with important outcomes. Indeed overall, O*NET measures correlated as expected with similar HRS job characteristic measures. For example, the O*NET measure "Performing General Physical Activities" are moderately positively correlated with the HRS measure for "Job requires a lot of physical effort," and measures for "stooping" are highly correlated between the two data sets.

We next conducted analyses to test whether the O*NET measures in our new linkage were associated with three different outcomes. Turning first to early retirement, we found that HRS respondents who worked in jobs that O*NET characterized as physically demanding, requiring static strength, or having to deal with aggressive people had an elevated risk of retiring before age 62. In terms of health outcomes, we found that those who worked in jobs characterized in O*NET as physically demanding or physically dangerous had a higher risk of reporting difficulties with activities of daily living and instrumental activities of daily living after age 65. Lastly, we examined the connection between O*NET measures likely to be related to dementia risk at age 65 or later. In particular, we examined work characteristics related to physical demands of the job and two measures potentially indicative of cognitive engagement on the job: interacting and/or communicating with people and cognitive complexity other than interaction with people ("ideas"). As expected, we found that those with jobs in the top tercile for the O*NET characteristics related to people and ideas had a statistically significantly lower risk of dementia.

Accessing the data

The Health and Retirement Study (HRS) Linkage to Occupational Information Network Data (HRS-O*NET) are available to registered HRS data users as an HRS restricted data product housed in a secure data enclave at the University of Michigan Institute for Social Research. The data description[***] provides extensive documentation of the O*NET measures included in the linked data set. The enclave can be accessed through a virtual desktop infrastructure (VDI). Visit the HRS website to apply for access. The HRS-O*NET data are listed with the Industry and Occupation data under the Survey Data heading. All O*NET-Census occupation data sets and documentation reports are available on the Claude Pepper Center website. ❖