Health conditions among male workers in mining and other industries reliant on manual labor occupations: National Health Interview Survey, 2007–2018

Detailed data modification methods

The Integrated Public Use Microdata Series (IPUMS) tool, funded by the National Institutes of Health (NIH), was used to aggregate the National Health Interview Survey (NHIS) data and harmonize variables to create the 12-year dataset.[1] Unconditional subpopulation analysis methods were used to ensure that the full complex sample design of the NHIS was accounted for in all variance estimation procedures. Guidelines for pooling data across multiple years as recommended by the National Center for Health Statistics (NCHS) were followed, and adjustments for variance calculations and sample weights were made.[2,3] Weights were adjusted for the 12-year pooling of observations by multiplying individual Sample Adult weights by the fraction of the sample year's observations to obtain the total number of observations in the pooled sample (number of observations in sample year/number of observations in pooled sample). The IPUMS tool automatically accounts for strata and cluster adjustment in the dataset output.

- **Industry subpopulations of interest**

The NHIS asked sample adults if they worked in the week before the interview. Current workers, ages 18–64 years, were asked about the job they last held; those age 65 years or older were only asked what job they had held the longest. Verbatim responses to the industry and occupation questions were obtained from each eligible sample adult. The industry and occupation text data were reviewed by Census Bureau computer-assisted coding specialists, who assigned the appropriate 4-digit Census Bureau codes. Census industry and occupation codes were based on the North American Industry Classification System (NAICS) and U.S. Bureau of Labor Statistics Standard Occupational Classification (SOC) System, respectively.[4] For the study, the NHIS 2-digit recodes of the 4-digit Census codes were used.[5,6]

These data include only nonmilitary, currently working sample adults with coded industry information. A sample adult was considered a currently working adult if they held a job within the last week or if they were unemployed but were looking for work and had worked within the last 12 months.

Workers in mining, except in oil and gas extraction (OGE), were the primary subpopulation of interest for the study. Therefore, it was necessary to split the mining sector into its two subsectors. "Miners" were those in NHIS detailed industry code 07 (equivalent to NAICS 212 [“mining, except oil and gas]) and OGE workers were those in either industry code 06 (equivalent to NAICS 211 [“oil and gas extraction]) or 08 (equivalent to NAICS 213 [“support activities for mining”]). Workers in the NHIS detailed industry recode 08 are almost entirely OGE workers.[7]

Manual labor occupations were defined as those with NHIS simple occupation recodes 18–22, comprising the following occupation groups: farming, fishing, and forestry; construction and extraction; installation, maintenance, and repair; production; and transportation and material moving.[8] Four industry groups with over 50% of workers in manual labor occupations were selected for comparison: construction; manufacturing; transportation and warehousing; and

agriculture, forestry, fishing, and hunting (NHIS industry simple recodes 04, 05, 08, and 02, respectively).

For regression analyses, a comparison group was created containing working adults across all other industries (i.e., workers whose NHIS simple industry recodes were not in 04, 05, 08, 02 or detailed recodes were not in 06, 07, 08).

- **Selected health conditions**

All health conditions were self-reported. Other than pain frequency and hearing quality, reported conditions were based on health professional diagnoses (i.e., "ever told by a doctor or health professional you had…"). From the Sample Adult core module, these included: cancer (any kind); cardiovascular disease (one or more: heart attack, coronary heart disease, heart condition, or angina); diabetes or prediabetes; hypertension (on 2+ separate visits); hearing quality without hearing aid; any lung condition (one or more: emphysema, current asthma, or chronic bronchitis [in past 12 months]); current asthma; vision problems (even with corrective lenses); lower back pain (within past 3 months); leg pain (spreading from lower back pain, within the past 3 months); neck pain (within past 3 months); and joint pain (within past 30 days). Hearing quality was reported in five categories, and the answers "a lot of trouble" and "deaf" were lumped together. From the Family core module, data on self-reported functional or activity limitation from a lung/breathing problem or hypertension were used.

- **Covariates**

Demographic variables included age, race/ethnicity, geographic region of residence (based on Census regions: Northeast, North Central/Midwest, South, and West), educational attainment, body mass index (BMI), heavy alcohol use (defined for males as consuming at least five drinks on at least 5 days a month over the past year[9]), current or former smoking status, health insurance status, and interval since last healthcare visit. Age was categorized into 18–34, 35–44, 45–54, and 65+ years based on the age distribution of current workers. Race and ethnicity were separate variables that were collapsed into a four-category variable: "white, non-Hispanic," "black, non-Hispanic," "other race, non-Hispanic," and "Hispanic." Collapsing categories was necessary due to the relatively low racial/ethnic diversity and small sample sizes in some of the industry groups of interest. Education was categorized by "less than high school," "high school graduate or GED," "some college or technical school," and "college graduate or more." Smoking status was categorized as "current or former" and "never." BMI was collapsed into the three standard categories <25 (normal), 25–29.9 (overweight), and 30+ (obese). Time since last healthcare visit was dichotomized as "less than 12 months" or "more than 12 months."

Due to the small sample size for the mining industry and to ensure a better model fit, all nonbinary modeling covariates (except BMI) described above were dichotomized. Where appropriate, the modeling covariates were dichotomized as: age <55 or 55+ years; race/ethnicity (white, non-Hispanic versus all others); education (less or more than high school/GED); and region (West versus all others). The age binary of 55 years was chosen due to the onset of some chronic conditions (as has been done in previous NHIS analyses[10]) and to the

distribution of the age of miners (who are generally older). In instances where the model was still not a good fit, age was recategorized and the modeling methodology reiterated; this occurred with neck pain (age modeled continuously) and functional limitation from hypertension (age modeled as a binary).

References

1. Blewett L, Rivera Drew J, King M, Williams K [2019]. IPUMS Health Surveys: National Health Interview Survey, Version 6.4 [dataset]. doi:https://doi.org/10.18128/D070.V6.4.

2. National Center for Health Statistics [2018]. Variance Estimation Guidance, NHIS 2016-2017. https://www.cdc.gov/nchs/data/nhis/2016var.pdf.

3. National Center for Health Statistics [2016]. Variance Estimation Guidance, NHIS 2006-2015. https://www.cdc.gov/nchs/data/nhis/2006var.pdf.

4. United States Census Bureau [2022]. Industry and Occupation Code Lists & Crosswalks. https://www.census.gov/topics/employment/industry-occupation/guidance/code-lists.html

5. IPUMS [2022]. Detailed industry classification, 2004 forward. https://nhis.ipums.org/nhis-action/variables/INDSTRN104#codes_section.

6. IPUMS [2022]. Simple occupational classification, 2004 forward. https://nhis.ipums.org/nhis-action/variables/OCCUPN204#codes_section.

7. United States Census Bureau [2016]. Statistics of U.S. Businesses. https://www.census.gov/programs-surveys/susb/data.html.

8. Krieger N, Barbeau E, Soobader MJ [2005]. Class matters: U.S. vs. U.K. measures of occupational disparities in access to health services and health status in the 2000 U.S. National Health Interview Survey. Int J Health Serv 35(2):213-236. https://doi.org/10.2190/JKRE-AH92-EDV8-VHYC.

9. National Institute on Alcohol Abuse and Alcoholism. Drinking levels defined. Date accessed: February 11, 2018. https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking.

10. Luckhaupt SE, Calvert GM [2014]. Prevalence of coronary heart disease or stroke among workers aged <55 years--United States, 2008-2012. MMWR Morb Mortal Wkly Rep 63(30):645-649.