HHS Public Access
Author manuscript
Am J Ind Med. Author manuscript; available in PMC 2022 December 01.

Published in final edited form as:
Am J Ind Med. 2021 December ; 64(12): 1018–1027. doi:10.1002/ajim.23292.

Determining occupation for National Violent Death Reporting System records: An evaluation of autocoding programs

Jonathan Davis, PhD[1], **Corinne Peek-Asa, PhD**[1], **Ann Marie Dale, PhD**[2], **Ling Zhang, MS**[3], **Carri Casteel, PhD**[1], **Cara Hamann, PhD**[4], **Bradley A Evanoff, MD**[2]

[1]Department of Occupational and Environmental Health, College of Public Health, University of Iowa, Iowa City, IA, USA

[2]Division of General Medical Sciences, Washington University School of Medicine, St. Louis, Missouri, USA.

[3]Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA, USA

[4]Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA, USA

Abstract

Background: Suicide is a leading cause of death for working-age adults. Suicide risk varies across occupations. The National Violent Death Reporting System (NVDRS) collects information about violent deaths occurring in the United States. Occupation can be determined using autocoding programs with NVDRS data. The objective of this analysis is to determine the accuracy of autocoding programs for assigning occupations in the National Violent Death Reporting System (NVDRS).

Methods: Deaths from suicide were identified in NVDRS for individuals age 16 and older from 2010–2017. Occupations were assigned after processing job description free text with autocoding programs. Job assigned by autocoding program were compared with the occupation code recorded on the death certificate.

Results: Assignment of major occupation group had substantial agreement (Cohen's kappa > 0.7) for the two autocoding programs evaluated. Agreement of assigned code varied across race/ethnicity and occupation type.

Jonathan Davis, Injury Prevention Research Center, University of Iowa, 2195 Westlawn, Iowa City, IA 52242-5000, Phone: (319) 384-4226, jonathan-a-davis@uiowa.edu.

Contributions: All authors designed the project, drafted and revised the manuscript, approved final version of the work to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to accuracy and integrity are appropriately investigated. Jonathan Davis performed analysis for the project. Corinne Peek-Asa acquired the data for this work. Institution at which the work was performed: University of Iowa

Disclosure (Authors): The authors declare no conflicts of interest.

Institution and Ethics approval and informed consent: This study was approved as an expedited study through the University of Iowa Human Subject Office as a secondary analysis of deidentified data. Data were obtained through a Data Use Agreement with the CDC/National Center for Injury Prevention and Control, National Violent Death Reporting System

Disclaimer: The National Violent Death Reporting System (NVDRS) is administered by the Centers for Disease Control and Prevention (CDC) by participating NVDRS states. The findings and conclusions of this study are those of the authors alone and do not necessarily represent the official position of the CDC or of participating NVDRS states.

Conclusions: Autocoding programs provide an efficient method for identifying the occupation for decedents in NVDRS data. By identifying occupation circumstances of suicide and rates of suicide can be studied across occupations.

Keywords

O*NET; occupation coding; job exposure matrix; suicide; surveillance

INTRODUCTION

In the US, suicide disproportionately affects working-age populations. Specifically, suicide is the second leading cause of death for people age 15 to 34 and the fourth leading cause of death for adults age 35 to 54.[1] Rates of suicide vary by occupation, but relatively little is known about occupation-specific risk factors for suicide.[2] While income,[3] education level,[4] and job stress[5] have been found to increase the risk of suicide, these characteristics do not fully explain the increased risk of suicide associated with specific occupations. Knowing the occupation of a decedent of suicide is essential to studying the determinants of suicide risk.

The Centers for Disease Control and Prevention established the National Violent Death Reporting System (NVDRS) to understand the burden of violent deaths in the US and to provide information to reduce these deaths, including from suicide.[6] The NVDRS collects information about the victim of the violent death and the circumstances that may have led to that death. Additionally, occupational information found on the death certificate is available in NVDRS. This information is reported as the usual occupation in the form of a census occupation code, as well as a free text description of the occupation. Unfortunately, the census occupation codes listed on death certificates have a high degree of missingness – 88% of records had usual occupation codes marked as unknown or not available for the years 2010 to 2017. A large degree of this missingness is due to the NVDRS system only allowing for the input of 3-digit occupation codes based on older occupation coding schemes. However, the free text occupation description is available for most decedents but requires manual review or automated processing to code occupation.

Autocoding software can assign occupation codes to decedents using free-text descriptions of occupations, reducing the need for manual coding of records. Autocoding of occupation with NVDRS data has been completed previouslypreviously[2] but has not been validated. Previous validation studies have examined agreement between occupation codes assigned by autocoding programs and occupation codes assigned by expert manual coders.[7] However, there is no published validation for using autocoding programs with information collected in NVDRS. In this study, we used two autocoding programs to assign US Bureau of Labor Statistics Standard Occupational Classification system (SOC) codes to decedents of suicide using the free text usual occupation information abstracted from the death certificate. To validate the two autocoding systems, we identified all the deaths that had a manually coded 3-digit occupational code recorded on the death certificate. We then compared the manually assigned SOC codes with codes assigned by the autocoding of the free text occupation recorded on the death certificate.

The use of SOC codes allows for identifying occupation, allows estimation of specific job exposures using the Occupational Information Network (O*NET) databases, which describe multiple occupational exposures and tasks linked to SOC codes. In addition to evaluation of occupation assignment, we compare how job exposures differ between the autocoded occupation and the manually coded occupation, since the autocoding program may assign an occupation with similar exposures even if the assigned SOC code is not an exact match. Correct assignment of workplace exposures would support the use of autocoding programs in the analysis of job exposures even when some misspecification of occupation may occur. Finally, we evaluate the performance of the autocoding programs across race/ethnicity to identify any disparities that may occur in the correct assignment of occupation. The results of this investigation inform the use of autocoding programs to assign occupation and associated job exposures in the analysis of NVDRS or death certificate data.

METHODS

Study Population

The NVDRS is a national reporting system established by the National Center for Injury Prevention and Control that collects information on violent deaths from homicide, suicide, law-enforcement involvement, and unintentional discharge of firearms. The NVDRS contains information on violent death abstracted from the death certificate, medical examiner report, and law-enforcement reports. For this analysis, deaths from suicide were included for the years 2010 to 2017. These NVDRS years include information from 35 states, as well as Puerto Rico and the District of Columbia. To be included in the analysis, decedents had to be 16 years or older at the time of death. Earlier age groups were excluded because many US states have restrictions on employment or require a work permit for workers under the age of 16.

Manually Coded Occupation from Death Certificate

The NVDRS captures information about an individual’s occupation and industry using free text job titles, and manually coded occupation codes using the census occupation coding system. Free text and manually coded occupation are both collected from the death certificate. Of the 137,719 included records for deaths from suicide in NVDRS between 2010 and 2017, only 9,136 (6.6%) had a manually coded occupation by the state’s registry for vital records on the death certificate. These 9,136 records were the source population for the autocoding evaluation with the following exclusions: records that were missing any occupation free text (14.3%), had a non-valid census occupation code 2002 Census Occupation Code (0.8%), and records with occupation free text the indicated the decedent was non-paid or non-working (10.4%). Non-paid workers and non-workers included the following groups: homemakers, volunteers, students, and never worked based on keywords listed in the Census Occupation Coding Manual.[8] In addition to the keywords identified in the coding manual, coding instructions were expanded to include male and female versions of the keyword (e.g., housewife and househusband). If the free text occupation description stated “retired” without an occupation from which the individual retired, these records were also excluded from the analysis. For example, individuals with “Retired” as their occupation were excluded while “Retired Truck Driver” were included in the analysis. Figure 1 provides a flowchart of the inclusion criteria for records used in this study. After exclusions, 6,811 records (74.6%) were used in this analysis.

Crosswalk of Death Certificate Census Occupation Codes to SOC Code

Multiple occupation classification systems exist to group individuals of similar occupations and type of work performed. SOC codes are developed by the Bureau of Labor Statistics and use a hierarchal system for grouping occupations. There are 23 major SOC groups and 867 detailed occupations nested within those groups allowing for more specific identification of an occupation. We evaluated these groupings at the major, the first 2-digits of the code, and detailed occupation level, first 2-digits and 4 more digits to identify a more specific occupation. SOC codes allow for linking occupation codes to exposures associated with that occupation as captured in the O*NET databases. To link the O*NET exposure databases that use SOC codes with the NVDRS dataset that uses census occupation codes, a crosswalk between the two coding systems was required. For this study, we assumed the form of the codes recorded on the death certificate were the 2002 Census Occupation Codes with the zero removed. The removing of the zero is necessary to make the code compatible with the restriction of only uploading a 3-digit code to NVDRS. This allowed for a one-to-one determination of SOC code that would not be possible with later census coding systems that had multiple codes with the same first 3 digits and a zero or five as the last digit. Because of the uncertainty in the form, this evaluation provides a reduced estimate of autocoding programs' accuracy compared to the accuracy that could be identified with a more accurate standard occupation to compare with the autocoded occupation. The manually coded census occupation from the death certificate was crosswalked to 2010 SOC codes (hereafter called *manually coded occupation*) using available crosswalks from the United States Census Bureau to determine reference SOC codes at the 2-digit (major occupation) and 6-digit (detailed occupation) levels.[9] This required linking a one-to-one crosswalk of 2002 death certificate Census Occupation Codes to 2002 SOC codes followed by an update of 2002 SOC codes to 2010 SOC codes. During this crosswalk, 98 records (1.4%) could not be crosswalked from 2002 codes to 2010 codes because a change in the coding system made the previous census code obsolete and a corresponding new code was not created. These records were not included in the analysis (Figure 1). There were 426 codes (6.3%) that had multiple matches between the 2002 and 2010 coding systems. When multiple matches occurred one code was selected based on similarity in naming conventions between the two versions of the SOC coding system. For example, the 2002 SOC code "11–3040 Human Resource Managers" has three corresponding 2010 SOC codes: "11–3111 Compensation and Benefits Managers", "11–3121 Human Resources Managers", and "11–3131 Training and Development Managers". "11–3121 Human Resources Managers" would be selected as the corresponding 2010 SOC code for the 2002 SOC code "11–3040 Human Resource Managers". Sensitivity analysis excluding the 426 records with multiple possible crosswalked codes did not meaningfully change the interpretation of the results. Some census occupation codes do not correspond to a 6-digit SOC code. This resulted in about a third of the records only having a major occupation (2-digit occupation) code (n=2,437; 35.8%), leaving 4,276 records in the analysis for the detailed occupation (6-digit SOC code) agreement (Figure 1). The process for crosswalking death certificate Census Occupation

Codes to 2010 SOC codes, and ultimate comparison with autocoded 2010 SOC codes, is depicted in Figure 2.

Autocoding Programs

Autocoding programs are used to assign SOC codes based on free text, which reduces the need to manually review free-text records to determine an occupation. The National Institute for Occupational Safety and Health (NIOSH) has developed the autocoding program NIOSH Industry & Occupation Computerized Coding System (NIOCCS), which is a web-based tool that uses industry and occupation text to assign SOC codes. NIOCCS uses the Census Alphabetical Index of Industries and Occupations, which assigns SOC codes to job titles. NIOCCS assigns a likelihood of a corresponding occupation based on the words in the free text job title using a series of matching approaches (including fuzzy and partial n-gram) to the Census Alphabetical Index of Industries and Occupations. A description of the program is available from NIOSH.[10] Occupation and industry text are inputs for the NIOCCS program; the program outputs a corresponding 2010 SOC code that was used in this analysis. Version 3.0 of NIOCCS was used for this investigation.

SOCcer 2.0 is an autocoding program and is available from the National Institutes of Health. The development of SOCcer has been described previously.[11] Briefly, SOCcer uses free text classifiers associated with SOC codes. The classifiers are then combined using a stacked ensemble approach.[12] A logistic regression model is used to give the probability of a given SOC code.[13] Job title and job tasks are used as inputs for the SOCcer autocoding program. For this analysis, the job title text field from the NVDRS data was used as an input to the SOCcer program; the program assigned a 2010 SOC code to the records that were used in this analysis. Earlier versions of SOCcer (SOCcer 1.0) allowed for an industry code as an input, similar to NIOCCS, but industry code is not used in SOCcer 2.0 as it did not meaningfully improve the assignment of SOC codes to records.[14] SOCcer also provides a confidence score that can be used to judge the reliability of the assigned occupation code. SOCcer guidance suggests manual review for records with confidence below 0.3.

O*NET Exposures

The Occupational Information Network, commonly referred to as “O*NET”, is a set of descriptors of occupational exposures, tasks, abilities, and skills for nearly 1,000 occupations by SOC codes. Data for each occupation is collected from workers in the occupation and from experts in the field and stored as a series of databases that contain occupation-level skills, exposures, and knowledge needed to work in each job.[15] Several scales are used to describe the relative level of each skill or ability (Lowest level 1- Highest level 7), the importance of each skill or ability (Not Important 1 – Very Important 5), and the frequency of exposure or task (Never 1 – Every Day 5) as defined by three sources: job incumbents, occupational experts, and occupational analysts. Measures are normalized from responses across the 5 and 7 levels. The variables used for this study were selected from those used in similar studies and included multiple O*NET databases: static strength and dynamic strength variables from the Abilities database, handling objects and general physical activities variables from the Work Activities database, and repetitive motion from the Work Context database. Job strain was derived from O*NET variables to compute a ratio of psychological job demands relative to decision latitude.[16,17] All variables were extracted from the 23.3 version of O*NET, the most recent release of the databases using the 2010 SOC coding taxonomy.

Analysis

Percent agreement was calculated for manually coded SOC codes and SOC codes determined by autocoding programs. The agreement was reported for NIOCCS and SOCcer autocoding programs at the major group (2-digit SOC code) and detailed group (6-digit SOC code) levels. The assignment confidence output from the SOCcer program was used to evaluate how the restriction of confidence impacted percent agreement. Percent agreement was determined separately in the following three ways using the SOCcer program: for all records, records with 0.1 or above assignment confidence, and records with 0.3 or above assignment confidence. Cohen's kappa statistic was calculated to further describe the reliability between SOC codes from the autocoding programs and the manually coded SOCs. Standard cutoffs for Cohen's kappa were used to describe the precision adjusted for chance of the autocoding programs: 0.41– 0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 almost perfect agreement.[18,19] Agreement was calculated for specific major occupation groups when at least 10 crosswalked SOC codes were identified for a major occupation group. The agreement for each major occupation category across race was calculated to determine if the agreement between the autocoding programs and manually coded SOC codes differed across race. Exposure variables derived from O*NET were treated as continuous variables. Pearson correlation coefficients were calculated for O*NET occupational exposures based on occupations assigned by the autocoding programs and manually coded occupation codes. Correlations were calculated across Non-Hispanic White (White), Non-Hispanic Black or African American (Black), and Hispanic decedents to detect any differences in the agreement of assignment autocoded and crosswalked SOC code across race.

RESULTS

Demographics

There were 6,811 suicide decedents in NVDRS between 2010 and 2017 that met the inclusion criteria for the analysis. The demographic differences between the NVDRS records in the study sample and all other NVDRS records that met inclusion criteria but did not have a manually coded occupation are presented in Table 1. Decedents in the study sample had similar age and sex to all other NVDRS records. There was an underrepresentation of Black or African American decedents with Black or African American decedents making up only 2.4% of the study sample suicides compared to 5.9% of all other NVDRS records. Most of the individuals in the study and excluded datasets were white (84.7% and 85.0% respectively).

Agreement of Autocoding Programs with Crosswalked SOC Codes

The percent agreement between the autocoding programs and manually coded occupation is presented in Table 2. Agreement for the SOCcer autocoded records increased when only records with a confidence score for the autocoded occupation of 0.1 (SOCcer-0.1) were included. Percent agreement continued to rise when only records with a confidence score of 0.3 (SOCcer-0.3) were included in the analysis. The output from NIOCCS does not have a similar confidence score, so additional stratification based on confidence score was not used in the evaluation of the NIOCCS program. If NIOCCS was unable to identify a SOC code for an individual, the autocoded SOC was left blank and the records were excluded. NIOCCS successfully assigned a job code for 5740 (85.5%) decedents of suicide. The number of NIOCCS autocoded records was similar to the number of decedents with SOC codes autocoded using the SOCcer program and a confidence score of 0.1 (n=5,789; 86.2%). Overall, both SOCcer and NIOCCS programs were able to identify the major SOC job categories with substantial agreement (SOCcer-0.1 agreement = 73%, kappa = 0.71 and NIOCCS agreement = 76%, kappa=0.74).

For the detailed (6-digit) SOC codes, the NIOCCS and SOCcer autocoding programs had reduced accuracy in assigning the detailed job codes relative to the assignment of major occupation codes, but both autocoding programs still identified more than half of the detailed jobs (Table 2). After restriction to 0.1 confidence and 0.3 confidence, the SOCcer program could correctly assign detailed job codes for 64% (95%CI: 62%−65%) and 73% (95%CI: 71%−75%) of the suicides, respectively. The NIOCCS program had a percent agreement of 69% (95%CI: 68%−71%) for detailed occupation codes.

Difference Across Major Occupation Categories

Figure 3 displays the differences in the agreement of occupation assignment across selected major occupation categories by race/ethnicity using the NIOCCS program relative to manually coded occupation. Occupation categories where all races had at least 10 decedents in the occupation group are displayed. The percent agreement across each occupation category is presented in the Appendix 1. Occupation varied significantly in the three race/ethnic groups (Chi-square p<0.001), as did the percent agreement across occupational categories. Overall percent agreement was highest for White decedents (76% 95%CI: 75% −77%) followed by Black decedents (63% 95%CI: 54%−71%) and then Hispanic decedents (49% 95%CI: 43%−54%, Appendix 1). Among the lowest sensitivity for the prediction of major job codes was Business and Financial Operations (percent agreement 23% 95%CI: 16%−29%). Business and Financial Operations was more frequently an occupation among Black decedents (12.5% vs. 8.1% White and 3.0% Hispanic). The highest agreement was found for Construction and Extraction (94% 95%CI: 92%−95%) and Legal (94% 95%CI: 87%−100%) – these are not presented in Figure 3 because of less than 10 decedents in a race/ethnicity group and are presented by race/ethnicity in Appendix 1. Several job categories had overall percent agreement between autocoded and crosswalked major occupation code over 80%, including Healthcare Practitioners and Technical; Protective Service; Food Preparation and Serving Related; Building and Grounds Cleaning and Maintenance; Sales and Related; Farming, Fishing, and Forestry; Installation, Maintenance, and Repair; and Transportation and Material Moving. However, within each occupation category, the agreement varied based on the race of the decedent. Overall, this sensitivity was lower for Hispanic decedents for most of the occupation categories analyzed when compared to White decedents (Appendix 1). The use of the SOCcer program with restriction to a 0.1 confidence score resulted in similar trends regarding occupation and race/ethnicity (data not shown).

O*NET Exposure Variables

The O*NET database was used to assign physical and psychological exposure levels to detailed occupations. Correlation between the exposure associated with manually coded occupation and the NIOCCS autocoded occupation category are presented in Table 3 for White, Black or African American, and Hispanic decedents. For the population overall, the exposures associated with the autocoded occupations were highly correlated with the exposures associated with the manually coded occupation of the decedent. The NIOCCS autocoding program was able to consistently assign similarly exposed occupations for White decedents across types of exposure. For Black decedents, the correlation of occupation exposures between autocoded and manually coded occupations was lower than the correlation found for White decedents, except for repetitive motion (0.80 95% CI: 0.78–0.80 for Black decedents and 0.80 95%CI: 0.79–81 for White decedents). For all O*NET exposures analyzed, the correlation was above 0.5 for Black decedents. Conversely, the exposures among occupations assigned to Hispanic decedents had low correlation with exposures of the manually coded occupation for most of the exposures evaluated. Only dynamic strength (0.55 95%CI: 0.45–0.64) and static strength (0.52 95%CI: 0.41–0.62) prediction had correlations above 0.50 between the crosswalked and autocoded jobs for Hispanic suicide decedents. The use of the SOCcer autocoding program with a confidence score of 0.1 had similar performance across decedent race and job exposures.

DISCUSSION

Overall, we found good agreement for both the NIOCCS and SOCcer autocoding programs for assigning occupation codes that match manual coding. A previous study validating autocoding systems in two occupational cohort studies with self-reported text-based occupations found similar agreement with 64.0% to 67.4% for major occupation codes for NIOCCS and 62.4% to 72.3% for SOCcer.[7] For detailed occupation codes (6-digit), we found a higher percent agreement than the former study and a lower percent agreement for major occupation (2-digit) occupation codes. The previous study found percent agreement for detailed occupation codes less than 50% for both autocoding programs.[7] The underlying populations used to validate the autocoding programs were different in these two studies. The two cohorts included in the previous validation used data from a cohort evaluating the incidence of carpal tunnel syndrome and a cohort evaluating the availability of health promotion to workers. The current study encompassed a greater variety of occupations by pulling from a national surveillance system, the NVDRS.

For both NIOCCS and SOCcer, agreement of the autocoded occupation with crosswalked occupation varied across different occupations and racial/ethnic groups. There was an apparent disparity in the effectiveness of the autocoding programs, with the poorest performance occurring for Hispanic decedents. The difference in the relative portion of Hispanic decedents employed within specific occupations that had lower percent agreement for autocoded and manual coded occupation does not alone explain this disparity. We found consistently lower percent agreement within several major occupation codes with White decedents having a higher percent agreement than Hispanic decedents. Death certificate records with no recorded job or text descriptions that indicated a non-working category were excluded, so differences in rates of unemployment would not be an explanation for the discrepancy either. Future use of autocoding programs of death certificates within Hispanic populations should be interpreted with caution, or analysis should be restricted to the assignment of occupations with a higher level of certainty from the autocoded program.

The NIOCCS autocoding program has been used to assign occupation codes and describe the rates of suicide across various occupations.[2] We found that NIOCCS had varying effectiveness in assigning an occupation code that matched the death certificate across different occupation categories: Business and Financial Operations; Life, Physical, and Social Science; and Personal Care and Service all had percent agreement below 50%. The variety of occupations within these fields may partially explain the low percent agreement. Rates of suicide for these types of occupations should be interpreted with caution. For all other major occupation codes (2-digit) the percent agreement was above 60% with several occupations having agreement above 80%. For these major occupation groups, NIOCCS or SOCcer can provide an efficient method for identifying a decedent’s occupation. For studies including a greater proportion of Hispanic workers than that in our study population, the lower percent agreement in the assignment of occupations within Hispanic decedents could result in an accuracy lower than that reported in this study of the NVDRS population. Additionally, the study of the performance of autocoding programs in other racial/ethnic groups is pertinent to fully understand the adequate assignment of occupation codes within specific groups.

Identifying occupations that have an increased risk of suicide is important in understanding how to prevent suicides that may have an occupation-related component. Jobs related to veterinary medicine are the only specific occupation category that has been investigated for increased suicide risk using NVDRS data. Veterinarians and veterinarian technicians were found to have higher rates of suicide compared to the general population.[20] The study required searching NVDRS occupation and industry-related variables using key terms developed by experts who had previously worked as a veterinarian and then manual review to assign a final occupation code to the decedent. The use of autocoding programs provides an alternative to developing specific search terms and manual review of job titles. We identified several occupations with a high agreement in assigning the same occupation as the death certificate. Future research can use autocoding programs to investigate differences in the contributing circumstances preceding deaths from suicide and differences in rates of death from suicide across occupations.

The incorporation of exposure information from the O*NET databases allows for the identification of common work-related risk factors for suicide that are present across multiple occupations, enabling a focus on physical and psychosocial work exposures rather than specific occupational groups. We identified a drop in the accuracy of the autocoding programs when the programs were used to identify detailed occupational groups compared to major occupations. In contrast, the O*NET assigned exposures associated with the autocoded detailed occupations were similar to the exposures associated with the manually coded occupations, suggesting that exposures are often similar within broad occupational groups. The accuracy of exposure information in O*NET and other similar job exposure matrices can vary by exposure and occupation, and misclassification may occur as assigned exposures do not account for variation among workers performing the same job. [17,21,22] The limitations of O*NET and other job exposure matrices should be considered when applying this exposure information.

A limitation of this evaluation is the absence of a true gold standard occupation for comparison with the autocoding programs. For analysis, we treated the manually coded occupation on the death certificate as if it were a gold standard, though we acknowledge that coding procedures may vary from state to state. In a prior comparison of manually coded versus auto-coded job titles, SOC codes were assigned independently by two coders with differences resolved by consensus.[7] Our study observed similar agreement to this previous validation study, supporting the use of the death certificate code as a comparison standard. Future work within the NVDRS system will rely on the information from the death certificate to code occupation, so it is important to describe the accuracy of autocoding using the information available from the death certificate. We had to assume the census occupation code in NVDRS was the 4-digit 2002 census occupation code with the zero removed for compatibility with the system. The actual coding system of the manually coded occupation on the death certificate may not be the abbreviated code we assumed. The coding system for the manually coded occupation can vary by the states abstracting the information to NVDRS and the individuals assigning the code to the death certificate itself. The lack of a true gold standard for comparison with the autocoded occupation results in a conservative estimate of the accuracy of the autocoding programs. Some autocoded occupations may match the manually coded occupation but use a different version of the census occupation codes than the assumed 2002 coding system for the manually coded occupation. Despite this limitation, we found strong agreement between manually coded and autocoded occupations. We were also limited in our ability to report validity for each racial/ethnic group within some occupations because some occupations had fewer than 10 decedents.

Conclusions

We have evaluated the performance of the SOCcer and NIOCCS autocoding programs for use with NVDRS data. We found that both autocoding programs performed well at assigning a major occupation code, but this performance varied across the occupation and race/ethnicity of the decedent. We also found a high agreement between occupational exposures assigned by manually coded and autocoded occupations. Despite the potential for misclassification, previous studies have suggested a correlation between actual and predicted exposures using job exposure matrices and O*NET.[17,22] This will allow for the application of O*NET derived exposures to decedents using their occupational information recorded on a death certificate. The use of O*NET exposure data and autocoded occupation in large study populations or surveillance systems has the potential to identify physical and psychosocial work exposures associated with increased suicide risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

This study was supported by the Healthier Workforce Center of the Midwest (grant number 5 U19 OH008868 from CDC/NIOSH) and the University of Iowa Injury Prevention Research Center (CDC/NCIPC R49CE003095)

Data Availability Statement:

Data used in this study are available from the CDC/NCIPC National Violent Death Reporting System through a Data Use Agreement.

REFERENCES

1. CDC. 10 Leading Causes of Death by Age Groung, United States - 2018: National Center for Injury Prevention and Control. CDC.; 2018 [acce1ssed August 8th 2020].
2. Peterson C, Sussell A, Li J, et al. Suicide Rates by Industry and Occupation - National Violent Death Reporting System, 32 States, 2016. MMWR Morb Mortal Wkly Rep 2020;69(3):57–62. doi: 10.15585/mmwr.mm6903a1 [published Online First: 2020/01/24] [PubMed: 31971929]
3. Agerbo E, Mortensen PB, Eriksson T, et al. Risk of suicide in relation to income level in people admitted to hospital with mental illness: nested case-control study. BMJ 2001;322(7282):334–5. doi: 10.1136/bmj.322.7282.334 [published Online First: 2001/02/13] [PubMed: 11159656]
4. Phillips JA, Hempstead K. Differences in U.S. Suicide Rates by Educational Attainment, 2000–2014. Am J Prev Med 2017;53(4):e123–e30. doi: 10.1016/j.amepre.2017.04.010 [published Online First: 2017/08/02] [PubMed: 28756896]
5. Milner A, Witt K, LaMontagne AD, et al. Psychosocial job stressors and suicidality: a meta-analysis and systematic review. Occup Environ Med 2018;75(4):245–53. doi: 10.1136/oemed-2017-104531 [published Online First: 2017/08/31] [PubMed: 28851757]
6. Stone DM, Holland KM, Bartholow BN, et al. Preventing suicide: a technical package of policies, programs, and practice. 2017
7. Buckner-Petty S, Dale AM, Evanoff BA. Efficiency of autocoding programs for converting job descriptors into standard occupational classification (SOC) codes. Am J Ind Med 2019;62(1):59–68. doi: 10.1002/ajim.22928 [published Online First: 2018/12/07] [PubMed: 30520070]
8. Instruction Manual for Industry and Occupation Coding 2010: National Institute for Occupational Safety and Health. CDC.; 2010 [Available from: https://www.cdc.gov/niosh/topics/coding/pdfs/Census2010CodingInstruction.pdf accessed February 10th, 2020.
9. Industry and Occupation Code Lists & Crosswalks: US Census Bureau; 2020 [Available from: https://www.census.gov/topics/employment/industry-occupation/guidance/code-lists.html accessed February 10th, 2020.
10. NIOSH. Industry and Occupation Coding: About the NIOSH Industry & Occupation Computerized Coding System.: National Institute for Occupational Safety and Health. CDC.; 2019 [Available from: https://www.cdc.gov/niosh/topics/coding/overview.html accessed February 10th, 2020.
11. Russ DE, Ho KY, Colt JS, et al. Computer-based coding of free-text job descriptions to efficiently identify occupations in epidemiological studies. Occup Environ Med 2016;73(6):417–24. doi: 10.1136/oemed-2015-103152 [published Online First: 2016/04/23] [PubMed: 27102331]
12. Wolpert DH. Stacked generalization. Neural Networks 1992;5(2):241–59. doi: 10.1016/S0893-6080(05)80023-1
13. Wang H, Xu Q, Zhou L. Large unbalanced credit scoring using Lasso-logistic regression ensemble. PLoS One 2015;10(2):e0117844. doi: 10.1371/journal.pone.0117844 [published Online First: 2015/02/24] [PubMed: 25706988]

14. NIH. Standardized Occupation Coding for Computer-assisted Epidemiological Research: National Cancer Institute.; 2020 [Available from: https://soccer.nci.nih.gov/soccer/ accessed February 10th, 2020.

15. O*NET® Data Collection Overview: U.S. Department of Labor/Employment and Training Administration; 2020 [Available from: https://www.onetcenter.org/dataCollection.html accessed February 10th, 2020.

16. Karasek R, Brisson C, Kawakami N, et al. The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. J Occup Health Psychol 1998;3(4):322–55. doi: 10.1037//1076-8998.3.4.322 [published Online First: 1998/11/07] [PubMed: 9805280]

17. Cifuentes M, Boyer J, Gore R, et al. Inter-method agreement between O*NET and survey measures of psychosocial exposure among healthcare industry employees. Am J Ind Med 2007;50(7):545–53. doi: 10.1002/ajim.20480 [published Online First: 2007/06/09] [PubMed: 17557294]

18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74. [published Online First: 1977/03/01] [PubMed: 843571]

19. Cohen J A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960;20(1):37–46. doi: 10.1177/001316446002000104

20. Witte TK, Spitzer EG, Edwards N, et al. Suicides and deaths of undetermined intent among veterinary professionals from 2003 through 2014. J Am Vet Med Assoc 2019;255(5):595–608. doi: 10.2460/javma.255.5.595 [published Online First: 2019/08/21] [PubMed: 31429646]

21. Solovieva S, Pensola T, Kausto J, et al. Evaluation of the validity of job exposure matrix for psychosocial factors at work. PLoS One 2014;9(9):e108987. doi: 10.1371/journal.pone.0108987 [published Online First: 2014/10/01] [PubMed: 25268276]

22. Fujishiro K, Koessler F. Comparing self-reported and O*NET-based assessments of job control as predictors of self-rated health for non-Hispanic whites and racial/ethnic minorities. PLoS One 2020;15(8):e0237026. doi: 10.1371/journal.pone.0237026 [published Online First: 2020/08/08] [PubMed: 32760109]

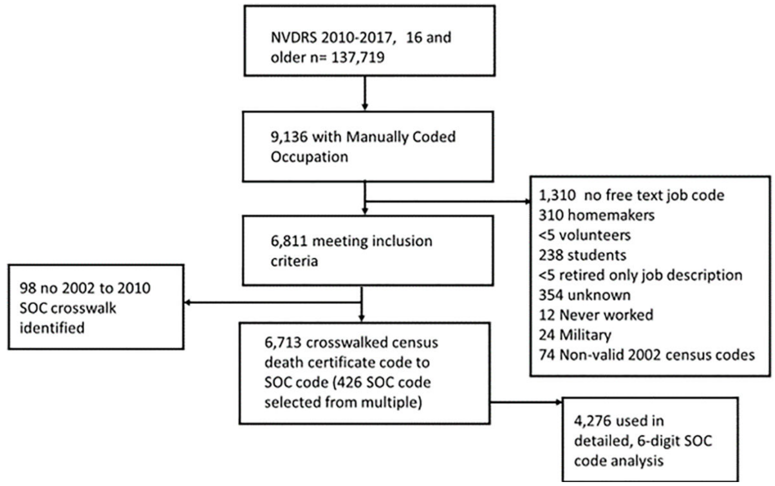

Figure 1:

Description of NVDRS records' occupation codes

Abbreviations: National Violent Death Reporting System, NVDRS; Standard Occupation Classification, SOC

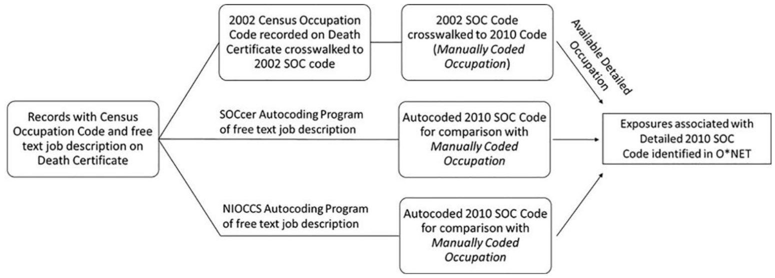

Figure 2:

Process for identifying manual and autocoded occupation codes. SOC, Standard Occupation Classification

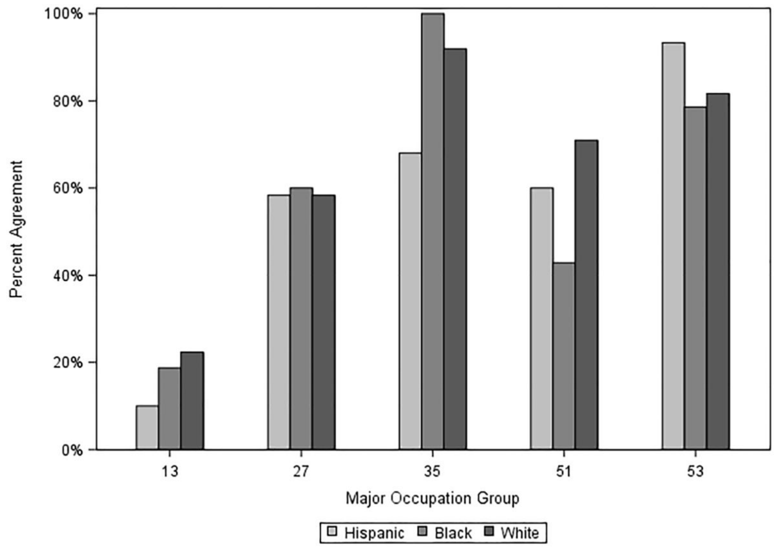

Figure 3:

Percent Agreement of NIOCCS autocoding program by occupation and race

Major Occupation Categories with 10 or more decedents in the category across the groups Hispanic, Black, and White: (13) Business and Financial Operations, (27) Arts, Design, Entertainment, Sports, and Media, (35) Food Preparation and Serving Related, (51) Production, (53) Transportation and Material Moving

Table 1:

Study and Excluded Population Demographics, Age 16 and older, 2010–2017

Variable	Level	Study Sample N = 6,811	%	All Other NVDRS N = 108,783	%
Biological sex of the victim *	Male	5540	81.4	88112	81.0
	Female	1270	18.6	20667	19.0
Age of victim (years)	<25	580	8.5	9009	8.3
	25–44	2317	34.0	36858	33.9
	45–64	2738	40.2	43443	39.9
	>64	1176	17.3	19473	17.9
Race and ethnicity of victim (combined)	White, non-Hispanic	5770	84.7	92510	85.0
	Black or African American, non-Hispanic	166	2.4	6401	5.9
	American Indian/Alaska Native, non-Hispanic	277	4.1	1089	1.0
	Asian/Pacific Islander, non-Hispanic	94	1.4	1844	1.7
	Other/Unspecified, non-Hispanic	16	0.2	232	0.2
	Two or more races, non-Hispanic	94	1.4	1168	1.1
	Hispanic	389	5.7	5451	5.0

Abbreviations: National Violent Death Reporting System, NVDRS

*Missing data results in numbers of records less than the total

Table 2:

Percent agreement between autocoded and manually coded SOC code

Autocoding Model	N	Percent Agreement % (95% CI)	Cohen's Kappa
Major Group (2-Digit)			
SOCcer	6713	67 (66–68)	0.64
SOCcer -0.1	5789	73 (72–74)	0.71
SOCcer -0.3	3780	80 (79–81)	0.78
NIOCCS	5740	76 (75–77)	0.74
Detailed Occupation (6-Digit)			
SOCcer	4276	56 (55–58)	0.55
SOCcer -0.1	3670	64 (62–65)	0.63
SOCcer -0.3	2502	73 (71–75)	0.72
NIOCCS	3643	69 (68–71)	0.68

Abbreviations: National Institute for Occupational Safety and Health Industry & Occupation Computerized Coding System, NIOCCS; Standard Occupational Classification, SOC

Table 3:

O*NET exposure correlation for NIOCCS autocoded and death certificate identified occupations (n=3,643)

Exposure	All, Suicide (95% CI)	Hispanic, Suicide (95% CI)	Black, Suicide (95% CI)	White, Suicide (95% CI)
Physiological Strain	0.76 (0.74–0.77)	0.41 (0.28–0.52)	0.61 (0.43–0.74)	0.78 (0.76–0.79)
Dynamic Strength	0.83 (0.82–0.84)	0.55 (0.45–0.64)	0.60 (0.41–0.73)	0.84 (0.83–0.85)
Static Strength	0.82 (0.81–0.83)	0.52 (0.41–0.62)	0.52 (0.31–0.68)	0.84 (0.83–0.85)
Handling Objects	0.80 (0.79–0.81)	0.41 (0.29–0.52)	0.51 (0.30–0.67)	0.82 (0.81–0.84)
Repetitive Motion	0.79 (0.78–0.80)	0.43 (0.31–0.54)	0.80 (0.70–0.88)	0.80 (0.79–0.81)
Physical Activity	0.80 (0.79–0.82)	0.44 (0.32–0.55)	0.63 (0.46–0.76)	0.82 (0.81–0.84)
Working with Computers	0.80 (0.79–0.81)	0.26 (0.13–0.39)	0.72 (0.58–0.82)	0.84 (0.83–0.85)

Abbreviations: CI, confidence interval; NIOCCS, National Institute for Occupational Safety and Health Industry & Occupation Computerized Coding System.