USING THE CENSUS BUREAU'S OCCUPATION AND INDUSTRY CODING SYSTEM FOR CODING DEATH CERTIFICATES

William Crouse and Lois Schuster, National Institute for Occupational Safety and Health
Harry Rosenberg, National Center for Health Statistics, Delana Kametani, Bureau of the Census
John Sestito, National Institute for Occupational Safety and Health

A. INTRODUCTION AND BACKGROUND

The Occupational Safety and Health Act of 1970 requires the Secretary of Health and Human Services to conduct research, experiments, and demonstrations that call for innovative methods, techniques, and approaches for dealing with occupational safety and health problems. This mission is conducted by the National Institute for Occupational Safety and Health (NIOSH), an institute within the Centers for Disease Control (CDC).

Congress has recognized that the effective administration of the Occupational Safety and Health Act is dependent on the availability of information relating to the incidence and prevalence of occupational health hazards, diseases, and injuries. This is reflected in Section 20(a) of the Act which authorizes NIOSH to establish programs to develop information on potentially toxic substances or harmful physical agents and to determine the incidence of occupational illnesses.

The growing awareness of the associations between occupational factors and elevated risk of mortality has stimulated a high level of interest among states and at the federal level in the use of occupation and industry information routinely reported on the death certificates of all states. As early as 1975, the National Center for Health Statistics (NCHS) and NIOSH discussed the need for developing uniform procedures to code the information, and thus facilitate its use at both state and federal levels in disease investigation and control.

Form OMB No. 68R1901 is the U.S. Standard Death Certificate which is recommended by NCHS as the model death certificate. Five portions of the certificate are used for the surveillance of job-related mortality: age, sex, race, cause of death, usual occupation, and kind of business or industry. The portion dealing with cause of death contains both immediate and contributing causes of death which are customarily coded using the International Classification of Diseases. The portion dealing with kind of business (industry) contains the kind of activity that was performed at the decedent's place of work; construction and paint manufacturing are typical entries contained in this section. The usual occupation portion contains the kind of work the decedent performed during most of his working life; carpenter and policeman are two examples of this entry.

Each year about two million death certificates are filed in the United States. All show the underlying cause of death. Most record the industry and occupation (I & O) of the decedent. Although this information was originally placed on the certificate as a socio-economic indicator, it was not long before health scientists used it for leads in the search of occupationally related mortality. One of the first salient uses of this information was made by Dr. Samuel Milham[1] who, in 1967, used death certificates to link woodworkers and Hodgkin's Disease. Since then, interest in using death certificates for possible leads to occupationally related mortality has been steadily increasing. Presently NIOSH, NCHS, and fifteen states are working together to code death certificates for the surveillance of occupationally related mortality. This paper presents the reasons the Census Bureau's classification coding system was chosen to code I & O information for analyses.

There was an infinite number of possible coding systems which could have been constructed, and it was impossible to give each one full consideration. Therefore, the strategy for selecting a coding system was to limit the scope of consideration to three major coding systems in the U.S. Eleven criteria for an acceptable coding system were developed before deciding which, if any, of the three U.S. systems would be used. As a result, the Census system was chosen as the most acceptable. (If none of the three had been deemed acceptable, other coding systems would have had to be considered.)

B. CRITERIA FOR AN ACCEPTABLE I & O CODING SYSTEM

Twenty-five years ago, before modern computers, it would have been a monumental task to analyze I & O death certificate data for just one year in one state. Today such analyses can be performed in seconds. However, in order to analyze the data using modern computers, the data must be coded. The criteria for a coding system acceptable for the surveillance of occupationally related mortality are listed below roughly ordered by their importance.

1. Coverage of United States workers must be adequate.

The coding system must include as a minimum all industries and occupations where workers are covered by the 1970 Occupation Safety and Health Act.

2. Detail must be adequate.

The coding system must classify the occupations into groups small enough (i.e., have enough detail) that routine analyses can be made to sift through the data to look for leads in the search of occupationally related mortality.

If the occupation categories are too large, workers with the disease of interest are grouped with unaffected workers. This grouping can cause the analyst to miss leads. On the other hand, the coding system should not have too much detail. Too much detail could cause two undesirable conditions. One condition would exist when the large number of groupings causes the number of workers in the groups to be too small to yield meaningful information. Another would exist when there are so many groups that chance, alone, causes a large number of the groups to appear to have excessive mortality.

3. Must be compatible with the U.S. Census Bureau coding system.

To have the most effective work-related mortality surveillance, yearly death rates should be calculated for occupations within industries. (A death rate is the number of workers who died divided by the number of workers.) The number of workers who died during a specific year can be obtained from the death certificate information. The number of workers who worked during a specific year can be obtained from a census for the years a census was taken and estimated for the years in between. The Bureau of Labor Statistics' Current Population Survey would provide a framework for estimates between decennial censuses since it provides current population estimates and the Census Bureau's classification systems are used in this survey. Since the Census Bureau uses its own I & O coding system for their employment data, their I & O coding system would need to be translated to a coding system used for death certificate data or vice versa. If the two coding systems are not compatible, many groups in one coding system do not have a similar group in the other system. As a result, ambiguities arise and information is lost during recoding.

4. Trained standardized coders must be available.

Unfortunately, a useable coding system is so complicated that coder training is necessary. It is best to have only one training manuscript so that all coders are taught exactly the same thing, standardizing the coders. If coders are standardized, differences among coders will have minimum effect on the analyses of the data. It is possible to have two sets of data be incompatible merely because the coders interpreted directions differently.

5. Necessary quality must be obtainable with a reasonable effort.

Ambiguities or difficult-to-understand portions in a coding system can cause entries to be coded incorrectly causing bias which, in turn, can cause spurious results. The chosen coding system does not have to be perfect. A perfect coding system would be so sophisticated it would incur an unreasonable amount of time, effort, and costs. However, there must be some evidence that the coding system will give the quality needed for surveillance of mortality.

6. Must be compatible with major coding systems used in the United States.

Rarely will the analytical results from death certificate data be so conclusive that other substantiating data are not desired. The other data, however, may have been coded using a different coding system than the one used for death certificates. The table summarizes the major industry and occupation coding systems that are widely used in the United States.

7. A standardized translation for revisions must be available.

There is always a need for changes in I & O coding systems, either to correct mistakes or to make the data more useable. When these changes are made the revised coding system must be translatable to the previous one. Otherwise, problems similar to those of the incompatible coding systems will appear. Someone must be responsible for implementation and dissemination of the coding revisions.

8. Health scientists must be able to use the coding system with a reasonable amount of effort.

If many days are needed to train scientists before they can understand which workers are in each group, the coding system is probably too esoteric. An acceptable system is one that researchers can use with minimal knowledge of the coding procedures.

9. The coding system should appeal to the states.

Since the coding and analysis will be done in the state in which the data are collected, the states must be convinced that it is the best system to use. If two states use different coding systems the results would be difficult to compare.

10. Costs must be acceptable to all potential users (states).

If costs do not stay reasonable, many potential users will be discouraged from using the system. There are fewer compatibility problems as the number of users increases. Many costs can be inadvertently caused by developing a new coding system. For example, a very sophisticated but complicated system would be slow to code and would increase coding costs. The cost of publishing and disseminating the coding procedures must also be considered.

11. The coding system must have a useable hierarchical structure.

It is necessary to group workers easily by what is known or suspected about their exposures. If the coding system facilitates selecting codes by these exposures, the job of the programmer is easier. For example, there may be a need to look at all managers as a group because they may have the same exposure. In the

Census system, the manager data can be recognized easily from the codes. Manager codes range from 003 to 019 and codes for management related jobs start at 023 and end at 037. Specific managers can also be analyzed such as Personnel and Labor Relations Managers, code 008.

C. THREE MAJOR INDUSTRY AND OCCUPATION CLASSIFICATION SYSTEMS

The table serves as a description of the three major coding systems. The three systems are the SIC combined with the SOC, the Census I & O, and the DOT (often used with the SIC). Note that all I & O codes in one system can be converted to I & O codes in another system. However, ambiguities arise in the conversion, and there is some loss of information.

Also note the large number of categories for the DOT. Although numerous categories are sometimes considered a blessing when detailed information is required, researchers find it does not merge as a blessing when the criteria are applied.

D. CRITERIA APPLIED TO THE THREE MAJOR CLASSIFICATION SYSTEMS

The criteria are applied below. Numbers listed below refer to criteria numbers listed in Section B above.

1. Coverage.

All three coding systems will allow surveillance of the industries and occupations under the 1970 OSH Act.

2. Detail.

The number of categories in the table indicate that the most acceptable system under this criteria would be the Census system, the least acceptable, the DOT. Milham[2] used a modified Census system to analyze death certificates from the State of Washington. His analysis confirmed occupational mortality associations which were already generally accepted. For example, his analysis showed elevated incidence of mortality among the following: deaths among electricians by electrocution, deaths among pilots by airplane crashes, and deaths among loggers caused by falling objects such as tree limbs. Since these associations are so widely accepted, they serve to empirically confirm that there is adequate detail in the Census system.

If a small, specific group of workers needs to be examined, and there are no I & O codes for the group in any of the systems, then special I & O codes can be created to augment the Census system. However, any special code should be such that it can be converted back to the appropriate three-digit Census code.

3. Compatibility with the Census system.

Of course the Census system appears best for the criterion. However, all three systems are compatible with the census system. Ambiguities do arise and information is lost when converting from one system to another.

4. Availability of trained coders.

Only the Census Bureau has training courses and training material readily available. If one of the other two systems had been chosen, these materials would have to be developed for the chosen system at considerable cost. The more complicated the coding system, the more involved is the training.

A relatively unexplored area of concern is coder burn-out. One-half of the attrition of NIOSH coder losses relate to burn-out. Coding work is very grueling. Experience to date has shown that a good coder using the Census system to code death certificates can complete 35 to 40 pairs (industry and occupation) per hour if only 4 hours of coding per day is required. The production coding rate falls to 20-25 pairs per hour if the coder codes for 8 hours. Error rates also rise after 4 or 5 hours per day and the burn-out rate of coders can reach as high as 80 percent after three months if the coding work is not combined with non-coding work such as typing, filing, or copying. Although we do not feel we have the complete solution, we have noticed that using sound managerial techniques is critical. Note that the burn-out problem with the Census system is probably less severe than with other systems because both the system and the training are less complicated.

5. Quality.

Because the Census system has training available, and it is simpler to use, the Census system appears most desirable under the quality criterion. The DOT with its intricate detail appears least desirable: In general, the I & O entries on death certificates do not have enough information to code to a specific DOT detailed category. This means the coders often would have to make difficult decisions to place the entry in the correct level of DOT detail. The SIC/SOC system, with its amount of detail between the DOT and Census system, is less desirable than the Census system but more desirable than the DOT.

NIOSH, NCHS, and the Census Bureau have collaborated to use the coding procedures employed for the 1980 decennial census to write a coding procedures manual for state production coders to code the I & O entries on state death certificates. The production coder (one who codes all entries for which routine procedures can be used) first checks the special cases for industry entries in the procedures manual. If a code cannot be obtained from the procedures manual, the "Alphabetical Index of Industries and Occupations" is checked to obtain the code. The same procedure is followed for coding the occupation entry. If either code cannot be determined from the Index the certificate is then given, or referred, to a referral clerk who has received more extensive training and has access to more reference materials.

The procedures, training materials, and accuracy of the Census coding system were pretested and evaluated by NIOSH, NCHS and the Census Bureau before the choice of systems was finalized. The coders from eight states were trained by the Census Bureau in a one-week course. They then coded a sample of about 23,000 certificates from their respective states. From this sample, 15,349 cases were selected and independently verified by experienced Census I & O coders using the same procedural materials provided to states. When the codes assigned to these cases did not agree, another independent group of Census I & O coding specialists reviewed all the assigned codes and the certificate entry to determine the best or "preferred" code. A Census Bureau report[3] presented an analysis of the results of the pretest coding project.

The pretest evaluation indicates that the Census Bureau I & O coding procedures can be used by state coders to code occupation and industry at levels of accuracy and production that are consistent with state experience in coding other vital statistics items on death certificates. For all of the states combined, the state-assigned codes agreed with the preferred Census three-digit level codes (most detailed) for 87 percent of the occupations and 88 percent of the industry entries. These percentages increased markedly (to 93 and 94 percent, respectively) when referral differences were eliminated. In other words, a large number of differences did not come from assigning a wrong code, but were differences between coding an entry and referring an entry. Because the pretest coding was done by coders with limited I & O coding experience, it is expected that the agreement percentage will improve as the coders gain experience.

The production rate of 40 cases per hour for all eight states combined is within the expected range and varies for individual coders from 69 cases per hour to 21 cases per hour. This result reflects both differences in the difficulty among the states (i.e., some states have a greater diversity of industries) and in variations of production rates among individual coders. Both production rates and accuracy would be expected to increase as coders gained experience.

The referral rate among the states ranged from 2 to 15 percent. It appears that the larger, more populated states have a larger variety of industries which increases the referral rate because the coders are less familiar with the industries and hence are less able to code proper names and/or vague entries.

For all states combined the overall error rate was 12.9 percent. By states, the range was 7.7 to 22.4 percent. The Current Population Survey (Bureau of Labor Statistics' monthly publication of total employment estimates) has an overall error rate of 5.7 percent with generally very useable data; the non-referral error rate is 4.4 percent. For the pretest, the non-referral error rate is 6.7 percent. The errors are not evenly distributed. Many are found in non-informative (i.e., leads are unlikely) occupation categories, such as "machine operators, not specified." Many of the errors could have been avoided by more experience. For example, interchanging of digits and erroneous codes for problem referrals accounted for some errors.

In addition to the I & O misclassification of workers (placing workers in I & O categories other than their true ones) by coders the official completing the death certificate can further complicate the situation by making an entry so vague that an accurate code cannot be obtained. Misclassification can cause serious bias, and hence, produce spurious results. Unfortunately, a general algorithm that states when the results are valid and when they are not, is not available. However, using the misclassification rates for occupation and industry[4] a rough estimate of the bias can be obtained.

6. Compatibility with major coding systems.

All but a few conversions from the SOC and SIC systems to the Census are straightforward. Automobile Mechanic is one of the few, SOC 6111 (Automobile Mechanic) can convert to Census 505 (Automobile Mechanic) or 507 (Automobile Mechanic Apprentice). From Census to SOC and SIC there are many overlapping codes because the SOC and SIC have more digits and hence more detail. Algorithms are not available for changing DOT codes to Census. However, the algorithm for going from the SOC to DOT can be reversed and then the codes can be converted to Census codes. As previously mentioned, for non-specific surveillance, the lack of detail is considered a blessing because it is simple to use and the groups tend to have large numbers of workers. If the mortality surveillance uncovers some possible elevated work-related mortality, studies are usually recommended to pinpoint the group of workers experiencing the elevated mortality.

7. Revision accommodation.

The Census Bureau publishes tables which aid in converting from one of their decennial coding systems to that of the next or prior decennial system decade. The conversion is not without some loss of information. However, it is not considered a major loss in the opinion of the authors. This is the only coding system of the three which has anything to accomodate recoding from revision to another. For all three systems revisions are made and disseminated under existing programs.

8. Easily used.

Users of the Census system generally considered it the easiest to understand of all the major classification systems.

9. Appeal.

None of the fifteen states presently using the Census system to code death certificates have expressed dissatisfaction with the coding system. Since the SIC/SOC system is closely related to the Census system it would probably

appeal to states, on the other hand, the DOT, with its complexities and lack of available training, probably would not prove as appealing to the states.

10. Costs.

Generally speaking, coders using the coding system are presently making about $4.00 an hour (not including any benefits they might receive). At 40 pairs of codes/hour, this means $.10 a code without overhead. The production rate of 40 cases/hour for all states combined is within the expected range which varies for individual coders from 69 cases/hour to 21 cases/hour. However, the report Coding Performance in the 1970 Census[4] estimates the production rate could double after 17 weeks.

The cost of publishing and disseminating the coding procedures of all three systems are already absorbed in other programs. The cost of the formal Census Bureau schooling is paid by the U.S. government through NIOSH. With the high burn-out rate, this is a concern because it costs an average of $1500 to train a coder. The DOT system, with its complexities and lack of available training, the most expensive of the three. Cost for the SIC/SOC system are somewhere in between the DOT and Census costs.

11. Hierarchical system.

All three coding systems appear to have an adequate hierarchical system when it becomes necessary to collapse codes in order to focus on specific exposures.

E. SUMMARY AND CONCLUSIONS

The Census system was chosen to code the I & O death certificate entries because it was deemed the most desirable of the three systems examined. None of the systems was more desirable than the Census system under the eleven criteria with the possible exception of the second when identifying a small, specific group of workers which can be more easily identified using the greater detail of the other systems. However, for the second criterion the Census system can be made adequate by augmenting the coding system with additional codes to give adequate detail.

Note that the use of the Census coding system need not be restricted to state vital statistics offices. The same characteristics that makes the Census coding system desirable for coding death certificates, makes it desirable for coding other research data. Presently at NIOSH, the Census coding system is used to code other I & O health data for analyses.

REFERENCES

1. Milham, S. and Hesser, J.E. "Hodgkin's Disease in Woodworkers," The Lancet: 136-137, July 1967.

2. Milham, S. "Occupational Mortality in Washington State," NIOSH Research Report, Volume I, HEW Publication No. (NIOSH) 76-175A.

3. U.S. Bureau of the Census. "An Evaluation on the Accuracy of Coding the Industry and Occupation Entries on Death Certificates by State Coders," September 9, 1982, (unpublished).

4. U.S. Bureau of the Census, Census of Population and Housing: 1970 Evaluation and Research Program PHC(E)-8. Coding Performance in the 1970 Census, April 1974.

SUMMARY OF MAJOR INDUSTRY AND OCCUPATION CODING SYSTEMS

Item	Standard Occupational Classification[1] (SOC)	Dictionary of Occupational Titles[2] (DOT)	Census Bureau[3]		Standard Industrial Classification[4] (SIC)
			Occupational	Industrial	
Developers	Office of Federal Statistical Policy & Standards	U.S. Employment Service	U.S. Bureau of Census	U.S. Bureau of Census	Office of Management and Budget
Structures	4-level; 4-digit	4-level; 9-digit	13 major categories; 3-digit	13 major categories; 3-digit	4-levels; 4-digit
Categories	654	13,000	503	231	1,031
Compatibility	Census & DOT	SOC	SOC	SIC	Census
Major Users	Census, CETA, Administrators[5], & NOICC[10]	SSA[6] & DOL[7]	Census & NCHS[8]	Census & NCHS	Census, BLS[9], & Private business
Major Uses	Coordinate information, focus training needs, & place graduating students into workforce.	Classify disabled & retired workers, job placement, employment counseling, & career guidance services.	Decennial census, CPS [11], & vital health statistics.	Decennial census, CPS, & vital health statistics	Unemployment & labor statistics, & business statistics.
Training materials	No	No	Yes	Yes	No

1. Standard Occupational Classification Manual, Office of Management and Budget, 1980
2. Dictionary of Occupational Titles, U.S. Department of Labor, 1977
3. Alphabetical Index of Industries and Occupations, U.S. Bureau of the Census, 1980
4. Standard Industrial Classification Manual, Office of Management and Budget, 1972
5. Comprehensive Employment and Training Act
6. Social Security Administration
7. Department of Labor
8. National Centers for Health Statistics
9. Bureau of Labor Statistics
10. National Occupational Information Coordinating Committee
11. Current Population Survey